

Rüdiger Grotjahn und Burkhardt Allner

Der C-Test in der Sprachlichen Aufnahmeprüfung an Studienkollegs für ausländische Studierende an Universitäten in Nordrhein-Westfalen*

Der Beitrag beginnt mit einer kurzen Darstellung der Aufgaben und des Kursangebots der fünf nordrhein-westfälischen Studienkollegs an wissenschaftlichen Hochschulen in Aachen, Bochum, Bonn, Köln und Münster. Die Kollegs vermitteln nicht direkt zum Hochschulstudium zugelassenen ausländischen Studienbewerbern die für ein Fachstudium erforderlichen Fertigkeiten. Eine Voraussetzung für die Teilnahme am Unterricht der Kollegs ist das Bestehen der Sprachlichen Aufnahmeprüfung (SAP). Die SAP findet zweimal jährlich statt und besteht aus drei der folgenden Prüfungsteile: 'Grammatik', 'Leseverstehen', 'Hörverstehen'/'Textwiedergabe', 'Textproduktion' und 'C-Test'.

Den Hauptteil der vorliegenden Arbeit bildet die statistische Analyse des Testdurchgangs WS 1994/95, der aus den Subtests 'Grammatik', 'Textwiedergabe' und 'C-Test' bestand ($N = 510$). Es werden u.a. folgende Aspekte untersucht: Schwierigkeit, Interkorrelation und Trennschärfe der C-Test-Texte; Reliabilität des C-Tests; Schwierigkeit und Interkorrelation der Subtests; Rangplatzdifferenzen der Teilnehmer in den Subtests; Einfluß der Faktoren 'Kurstyp' und 'Herkunftsland' auf die Testergebnisse; Korrelation zwischen SAP und Semesternote.

Im letzten Teil werden die wichtigsten Ergebnisse zusammengefaßt und praktische Empfehlungen u.a. zum Einsatz des C-Tests in der SAP ausgesprochen. Der C-Test erwies sich nicht nur als ökonomisch, objektiv und reliabel, sondern auch als valide in bezug auf die mit Hilfe der SAP zu treffenden Entscheidungen.

1. Einleitung

Der C-Test gewinnt auch in den Bereichen Deutsch als Fremdsprache und Deutsch als Zweitsprache zunehmend an Bedeutung. In den genannten Bereichen ist dieses Testverfahren bisher u.a. folgendermaßen eingesetzt worden:

- als Mittel der globalen Sprachstandsdiagnose im Deutschen als Zweitsprache bei Kindern (z.B. Süßmilch, 1984, 1985);

* Wir danken Friedrich Denig, Beate Helbig und Ulrich Raatz für die konstruktive Kritik einer früheren Fassung dieses Artikels.

- zur Ermittlung der globalen Sprachfähigkeit bei Schülern mit Deutsch als Zweitsprache und Türkisch, Kroatisch oder Griechisch als Muttersprache mit dem Ziel der Überprüfung der Interdependenzhypothese (Baur & Meder, 1994);
- im Rahmen der Psychologischen Dienste der Arbeitsämter als Entscheidungshilfe in bezug auf Weiterbildungsmaßnahmen (Jakschik, 1992, 1994, 1996);
- als Auswahlhilfe für Weiterbildungsmaßnahmen ausländischer Jugendlicher (Jacobs & Arndt, 1985);
- als Einstufungstest für Deutschkurse mit erwachsenen Ausländern am Eurozentrum in Köln (Grönwoldt, 1985);
- als Meßinstrument alltagssprachlicher und akademischer Fertigkeiten im Deutschen bei aus Deutschland in die Türkei zurückgekehrten türkischen Remigranten (Daller, 1996);
- als standardisierter und geeichter Schultest zur Erfassung der globalen schriftlichen Beherrschung der deutschen Sprache in der zweiten Hälfte der 4. Klasse an Grundschulen (Raatz & Klein-Braley, 1992);
- im Rahmen des "European Language Proficiency Survey" insbesondere bei englischen Studenten zur Erhebung des Sprachstands im Deutschen und anderen Sprachen (vgl. Coleman, 1994, 1996; Coleman, Grotjahn, Klein-Braley & Raatz, 1994);
- als Einstufungstest in studienvorbereitenden und studienbegleitenden Kursen für Deutsch als Fremdsprache an universitären Lehr- und Forschungsgebieten (Bolten, 1992);
- im Rahmen der sprachlichen Aufnahmeprüfung an Studienkollegs für ausländische Studierende in Nordrhein-Westfalen (vgl. Nellessen, 1992).

Im vorliegenden Aufsatz wollen wir ausführlich auf den letztgenannten Einsatzbereich eingehen. Dabei sollen u.a. auf der Grundlage ausführlicher statistischer Analysen begründete Empfehlungen für den Einsatz des C-Tests im Rahmen der Sprachlichen Aufnahmeprüfung (SAP) an universitären Studienkollegs gegeben werden.¹

¹ Der entscheidende Anstoß für den vorliegenden Beitrag kommt von einer Fortbildungsveranstaltung zum Thema 'SAP', die der erste Verfasser (R. Grot-

2. Die nordrhein-westfälischen Studienkollegs

2.1 Aufgabe und Kursangebot

Ausländische Studienbewerber mit einer Hochschulzugangsberechtigung, die mit dem deutschen Sekundarschulabschluß nur bedingt vergleichbar ist, werden erst dann zum Fachstudium an einer deutschen Hochschule zugelassen, wenn sie die "Prüfung zur Feststellung der Eignung ausländischer Studienbewerber für die Aufnahme eines Studiums an Hochschulen in der Bundesrepublik Deutschland" ("Feststellungsprüfung") bestanden haben. Die "Feststellungsprüfung" wird in Deutsch als Fremdsprache und drei weiteren Fächern abgelegt. Für die Vorbereitung auf diese Prüfung und deren Abnahme sind die Studienkollegs verantwortlich.²

In Nordrhein-Westfalen (NW) gibt es vier Studienkollegs für ausländische Studierende an Universitäten, und zwar in Aachen, Bonn, Köln und Münster sowie eines in privater Trägerschaft in Bochum.³ Ihre Aufgabe beschreibt die nordrhein-westfälische "Verordnung über die Ausbildung und Prüfung für ausländische Studierende der Studienkollegs an wissenschaftlichen Hochschulen" (APO) (Kultusminister NW, 1989) folgendermaßen:

"Das Studienkolleg an wissenschaftlichen Hochschulen vermittelt ausländischen Studienbewerbern, die nicht unmittelbar zum Hochschulstudium zugelassen werden können, die Kenntnisse und Fähigkeiten, die für ein Fachstudium an einer Hochschule in der Bundesrepublik Deutschland ... erforderlich sind." (S. 518, §1, Abs. 1)

Die von den Studienkollegs im Rahmen dieser Aufgabenstellung angebotenen Lehrveranstaltungen finden in Schwerpunktkursen statt. Sie orientieren sich an bestimmten Fachrichtungen:

jahn) im Oktober 1994 in Bochum für Lehrer an nordrhein-westfälischen Studienkollegs durchgeführt hat.

² Welcher Studienbewerber das Studienkolleg vor der Aufnahme eines Fachstudiums besuchen muß, ist festgelegt in den Bewertungsvorschlägen der Zentralstelle für ausländisches Bildungswesen (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, 1994ff.). Ausführliche Informationen enthält außerdem Deutscher Akademischer Austauschdienst (1993).

³ Weitere Studienkollegs für ausländische Studierende gibt es in NW an den Fachhochschulen in Köln, Dortmund, Krefeld und Mettingen. Sie erstellen eigene Aufnahmeprüfungen und bleiben deshalb in diesem Artikel unberücksichtigt.

- a) T-Kurse für technische und mathematisch-naturwissenschaftliche Studienfächer;
- b) M-Kurse für medizinische, biologische, agrar- und forstwissenschaftliche Fächer;
- c) W-Kurse für wirtschafts- und sozialwissenschaftliche Fächer;
- d) S/G-Kurse für sprachliche, gesellschaftswissenschaftliche und künstlerische Fächer.

Besucht ein Studienbewerber zum Beispiel einen T-Kurs, erhält er Unterricht in Deutsch, Mathematik, Physik und Chemie; ein Studienbewerber, der an einem G-Kurs teilnimmt, wird in den Fächern Deutsch, Geschichte, Deutsche Literatur (alternativ: Englisch) sowie Sozialkunde/Geographie unterrichtet. Daneben werden freiwillige Arbeitsgemeinschaften angeboten, z.B. im Fach Informatik. Die Ausbildung dauert ein Schuljahr (ca. 10 Monate), eingeteilt in zwei Semester, bei 32 Stunden Unterricht pro Woche.

2.2 Die Sprachliche Aufnahmeprüfung (SAP)

2.2.1 Administrativer Rahmen

Voraussetzung für die Aufnahme in ein Studienkolleg ist – neben der Anerkennung des ausländischen Bildungsabschlusses durch die Bezirksregierung Düsseldorf – das Bestehen der Sprachlichen Aufnahmeprüfung (SAP). Hier muß der Studienbewerber Kenntnisse der deutschen Sprache nachweisen, "die es ihm ermöglichen, mit Aussicht auf Erfolg an den Lehrveranstaltungen des Studienkollegs teilzunehmen" (Kultusminister NW, 1989, S. 519, §2, Abs. 1). Den Rahmen für die SAP in NW setzen die bundesweit geltenden "Empfehlungen über die Prüfung zum Nachweis der Deutschkenntnisse bei der Aufnahme in ein Studienkolleg" (Kultusministerkonferenz, 1983, S. 254) und die bereits oben zitierte nordrhein-westfälische Ausbildungs- und Prüfungsordnung für die Studienkollegs (Kultusminister NW, 1989). Eine weitergehende inhaltliche Beschreibung wird u.a. ein bundesweiter Rahmenplan enthalten, der seit November 1994 von den Deutschlehrern der Studienkollegs erarbeitet wird.

2.2.2 Testerstellung und -durchführung

Die SAP findet in NW zweimal jährlich – Mitte Januar und Ende August – statt. Die Prüfungsaufgaben (mit Korrekturvorschlägen und Instruktionen

und Testdurchführung) werden von Deutschlehrern jeweils eines Studienkollegs erstellt, den anderen Kollegs zur Überprüfung vorgelegt und in einem Fachgremium (Fachausschuß Deutsch) abschließend besprochen.⁴ Seit einigen Jahren werden zudem, wenn dies möglich ist, die einzelnen Testteile an anderen Institutionen in Lerngruppen mit vergleichbarem Leistungsstand überprüft. Alle fünf beteiligten Kollegs führen die Prüfung am gleichen Tag und zur gleichen Uhrzeit in einem Hörsaal der örtlichen Universität durch. Ca. 1000 Studienbewerber, die meisten im Alter von 18-25 Jahren, nehmen pro Jahr an der SAP teil.

2.2.3 Prüfungsteile

Die Prüfung besteht zur Zeit aus drei Subtests. Prinzipiell sind gemäß Absprache zwischen den nordrhein-westfälischen Studienkollegs folgende fünf Testteile möglich: Leseverstehen, Hörverstehen, Grammatik, Textproduktion und C-Test.⁵ Welche drei Teile kombiniert werden, ist nicht festgelegt; die Grammatik war jedoch bisher immer Teil der Gesamtprüfung. Die Reihenfolge für die Bearbeitung der Testteile wird vorher festgelegt. Die Bearbeitungszeit beträgt in der Regel 45 Minuten pro Prüfungsteil. Als Beispiel ist die Sprachliche Aufnahmeprüfung des Wintersemesters 1994/95 im Anhang abgedruckt.

(a) Leseverstehen

In dem Prüfungsteil "Leseverstehen" wird ein Text (einfacher, klar gegliederter Sachtext oder argumentativer Text) mit einem Umfang von ca. 250 Wörtern den Prüflingen vorgelegt. Zu diesem Text werden mehrere Fragen

⁴ Der Fachausschuß Deutsch setzt sich zusammen aus Deutschlehrern der Studienkollegs in NW. In diesem Gremium werden die Prüfungsentwürfe zur SAP abschließend diskutiert sowie durchgeführte Prüfungen evaluiert. Darüber hinaus hat der Ausschuß die Aufgabe, eine Abstimmung der Studienkollegs in didaktischen Fragen zu erreichen, bei der Erstellung von Curricula und Prüfungsordnungen im Bereich DaF die Interessen der Deutschlehrer an den nordrhein-westfälischen Studienkollegs zu formulieren sowie bei der Planung und Durchführung von Fortbildungstagungen mitzuarbeiten.

⁵ C-Tests werden – abweichend von der bestehenden Prüfungsordnung – seit mehreren Prüfungsterminen mit Genehmigung der Bezirksregierung Münster zur Leistungsfeststellung eingesetzt. Dabei soll gleichzeitig herausgefunden werden, ob diese Testform für die SAP geeignet ist.

und Aufgaben gestellt, um das Verständnis der Gesamtaussage und der Detailinformationen zu überprüfen. Die in den letzten Jahren am häufigsten verwendeten Aufgabentypen sind:

- Beantworten von Fragen zum Text;
- Vervollständigen von Sätzen (mit Hilfe der Textinformation);
- Herausfinden von richtigen Aussagen in bezug auf den Text;
- Auffinden von Informationen im Text;
- Bearbeitung von Aufgaben zum Text.

(b) Grammatik

Der Studienbewerber erhält in diesem Prüfungsteil verschiedene Aufgaben zur Grundgrammatik der deutschen Sprache. Häufige Aufgabentypen sind:

- Ausfüllen von Lücken in einem Text;
- Transformationen;
- Bildung von Sätzen.

Die Aufgaben der Grammatik beziehen sich auf das Thema des Prüfungsteils 'Leseverstehen' oder bilden einen eigenen Text.

(c) Textproduktion

Es ist ein kurzer Text über einfache Sachverhalte zu schreiben. Häufige Aufgabentypen sind:

- Beschreiben eines Bildes, einer Bildgeschichte, einer Tabelle, eines Schaubildes o.ä. (evtl. mit einer Stellungnahme zum Thema);
- Formulierung eines Textes nach Stichwörtern;
- Schreiben eines kurzen Berichts oder Briefes.

(d) Textwiedergabe/Hörverstehen

In diesem Prüfungsteil wird ein einfacher Text von einem Lehrer oder mehreren Lehrern (Dialog) zweimal vom Blatt vorgelesen. Mögliche Aufgabentypen sind:

- Wiedergabe der Informationen des gesamten Texts oder von Textteilen mit eigenen Worten (Textwiedergabe);

- Beantwortung von Fragen zum Text;
- Multiple-Choice-Aufgaben.

(e) C-Test

Es werden acht Texte (je 25 Lücken) mit verschiedenen Themen und unterschiedlichem Schwierigkeitsgrad vorgelegt. Die Arbeitszeit beträgt 45 Minuten; den Studenten werden keine Empfehlungen zur Zeiteinteilung gegeben. Der Prüfling erhält die Anweisung, die Wörter so zu vervollständigen, daß ein sinnvoller, korrekter Text entsteht.

Die Vorteile des C-Tests liegen vor allem in seiner hohen Objektivität und Reliabilität sowie in der ökonomischen Erstellung und Korrektur. Hinzu kommt, daß zur Bearbeitung des C-Tests Kenntnisse grammatischer Termini nicht nötig sind und die Aufgabenstellung unmißverständlich ist. Als Nachteil wird gesehen, daß der C-Test lediglich eine globale Aussage über den Leistungsstand erlaubt.

Die hohe Akzeptanz der übrigen Tests bei den Kolleglehrern erklärt sich aus dem direkten Bezug der Tests zum Unterricht des Kollegs: Im Grammatiktest weist der Bewerber das Sprachwissen nach, das ihm einen Einstieg in den Sprachunterricht des ersten Semesters ermöglicht. In den Fertigkeitstests (Hörverstehen, Leseverstehen, Textproduktion) zeigt er vor allem die Fähigkeit zur Teilnahme am Fachunterricht. Hier liegt – neben der Möglichkeit zur detaillierten diagnostischen Auswertung – sicherlich die Stärke der genannten Testformen.

Hervorzuheben ist, daß sich insgesamt gesehen die Tests hinsichtlich des Gütekriteriums 'Objektivität' – trotz detaillierter Anweisungen zur Durchführung und Korrektur – recht unterschiedlich verhalten. Als problematisch ist besonders die Durchführung der Textwiedergabe anzusehen: Durch die unterschiedliche mündliche Darbietung eines Texts durch verschiedene Lehrer kann es zu einer Bevorzugung oder Benachteiligung der Kandidaten eines Kollegs kommen. Das Problem der Auswertungsobjektivität tritt auch bei den Prüfungsteilen 'Textproduktion', 'Hörverstehen' und 'Leseverstehen' auf, da hier die korrigierenden Lehrer auf Grund der freien Beantwortung der Aufgaben einen großen Bewertungsspielraum haben.

2.2.4 Korrektur und Bewertung

Die Korrektur der SAP erfolgt durch die Lehrer des jeweiligen Kollegs entsprechend der Korrektorempfehlung des Lehrerteams, das die Prüfung erstellt hat. Die Korrektorempfehlung enthält eine Musterlösung sowie Angaben zur Bewertung. Im Fall des C-Tests werden lexikalische Abweichungen vom Originaltext nicht als Fehler angesehen, wenn der Sinn des Textes nicht wesentlich verändert wird. Geringfügige, das Verstehen nicht beeinträchtigende orthographische Fehler bleiben ebenfalls unberücksichtigt.

Die SAP ist bestanden, wenn insgesamt 60 von 100 Prozentpunkten erreicht wurden. Dabei bleibt unberücksichtigt, wieviele Prüfungsteile bestanden wurden. Die Prüfung kann zweimal wiederholt werden. Die Ergebnisse der SAP werden in erster Linie zur Entscheidung über die Zulassung zum Studienkolleg verwendet; eine weitergehende diagnostische Auswertung findet nur in Einzelfällen statt.

Hat ein Studienbewerber die SAP nicht bestanden, aber Grundkenntnisse in Form einer bestimmten Punktzahl nachgewiesen, kann er an einem Sprachkurs der Studienkollegs (Grundkurs) teilnehmen, der ihn auf die nächste Prüfung vorbereitet. Der Kurs dauert ein halbes Jahr bei ca. 25 Stunden Deutschunterricht pro Woche.

3. Der C-Test im statistischen Vergleich mit den übrigen Prüfungsformen der SAP

3.1 SS 1992 und WS 1992/93: Die Untersuchung von Nellessen (1992)

Horst Nellessen (Studienkolleg Münster) hat 1992 einen ersten, allerdings nicht sehr weitgehenden statistischen Vergleich zwischen C-Test und den übrigen Prüfungsformen der SAP an mehreren Studienkollegs Nordrhein-Westfalens durchgeführt (Nellessen, 1992).⁶ Wir wollen im folgenden einige zentrale Ergebnisse dieser unveröffentlichten Studie kurz diskutieren.

⁶ Im WS 94/95 hat Nellessen einige weitere – allerdings auf das Studienkolleg Münster beschränkte – statistische Analysen durchgeführt. In Abschnitt 3.2 haben wir die Daten aus Münster umfassend reanalysiert.

Tabelle 1
SS 92 und WS 92/93
Stichprobenumfänge

	SS 92					WS 92/93	
	Aachen	Bochum	Bonn	Köln*	Münster	Aachen	Münster
C-Test	153	38	187	48	171	152	195
GRAM	153	38	187	48	171	152	195
TW	153	38	187	48	171	–	195
TP	–	–	–	–	–	152	–

GRAM = Grammatik; TW = Textwiedergabe; TP = Textproduktion

* Für Köln lagen nur die Ergebnisse der **bestandenen** Prüfungen vor.

Tabelle 2
SS 92 und WS 92/93
Interkorrelationen der Subtests

	SS 92					WS 92/93	
	Aachen	Bochum	Bonn	Köln*	Münster	Aachen	Münster
C-Test/GRAM	.78	.50	.76	.55	.81	.86	.86
C-Test/TW	.74	.70	.77	.75	.80	–	.85
C-Test/TP	–	–	–	–	–	.69	–
GRAM/TW	.74	.58	.74	.61	.78	–	.80
GRAM/TP	–	–	–	–	–	.71	–

GRAM = Grammatik; TW = Textwiedergabe; TP = Textproduktion

* Für Köln lagen nur die Ergebnisse der **bestandenen** Prüfungen vor.

Die Analysen von Nellessen beruhen auf den in Tabelle 1 angegebenen Stichprobengrößen.

Tabelle 2 zeigt die Interkorrelationen (Pearson-Produkt-Moment-Korrelationen) der einzelnen Testteile.

Berücksichtigt man bei der Interpretation der Resultate in Tabelle 2 nur die wegen der höheren Stichprobenumfänge verlässlicheren Daten aus Aachen, Bonn und Münster, dann liegen die Korrelationen im Bereich .70 oder höher. Insgesamt gesehen scheinen die Subtests der SAP somit in nicht unerheblichem Maße das gleiche zu messen. Zudem dürften die 'wahren' Korrelationen des Testteils 'Textwiedergabe' mit den übrigen Testteilen in Wirklichkeit noch höher sein, da der Subtest 'Textwiedergabe' wegen seiner

geringen Auswertungs- und Durchführungsobjektivität vermutlich wenig reliabel ist (vgl. auch die Diskussion der Tabellen 3 und 4).⁷

Beachtenswert ist auch die relativ hohe Korrelation zwischen C-Test und Textproduktion. Dies kann als Hinweis gedeutet werden, daß der C-Test in nicht unerheblichem Maße schriftlich-produktive Fertigkeiten erfaßt. Diese Folgerung wird u.a. durch die Untersuchungen von Bolten (1992) gestützt, der im Rahmen der "Düsseldorfer Zertifikationsprüfungen Wirtschaftsdeutsch" eine Korrelation von .92 zwischen Textproduktion und C-Test ermittelt hat.

Auch die hohe Korrelation zwischen C-Test und Textwiedergabe ist neben einer gemeinsamen Verstehensdimension möglicherweise in Teilen auf eine gemeinsame produktive Fertigkeitsdimension zurückzuführen. Im Testteil 'Textwiedergabe' wird u.a. gefordert, daß der Prüfling Informationen aus einem vorgelesenen Text in eigenen Worten auf deutsch wiedergibt (vgl. Abschnitt 2.2.3). Dadurch mißt dieser Testteil neben Hörverstehen im beträchtlichen Umfang auch produktive Fertigkeiten.

Tabelle 3 zeigt die prozentualen Schwierigkeiten der einzelnen Subtests für die verschiedenen Studienkollegs.

Tabelle 3
SS 92 und WS 92/93
Schwierigkeiten (%)

	SS 92					WS 92/93	
	Aachen	Bochum	Bonn	Köln*	Münster	Aachen	Münster
C-Test	55.7	63.8	57.0	(77.3)	58.0	54.7	52.4
GRAM	60.6	68.4	59.7	(73.7)	59.3	56.8	53.2
TW	55.6	60.9	40.8	(76.0)	50.2	-	48.8
TP	-	-	-	-	-	66.0	-

GRAM = Grammatik; TW = Textwiedergabe; TP = Textproduktion

* Für Köln lagen nur die Ergebnisse der **bestandenen** Prüfungen vor.

⁷ Nellessen (1992) hat keine Reliabilitäten für die einzelnen Teile der SAP berechnet. Wegen der Datenlage konnten wir leider auch nachträglich keine Reliabilitätsabschätzungen vornehmen.

Wie aus Tabelle 3 ersichtlich ist, lag bei den einzelnen Studienkollegs der C-Test in seiner Schwierigkeit stets relativ nah bei den übrigen Testteilen. Am schwierigsten war in allen Fällen die Textwiedergabe. Betrachtet man erneut lediglich die verlässlicheren Daten aus den Studienkollegs Aachen, Bonn und Münster, dann unterscheiden sich die Kollegs nur sehr wenig hinsichtlich der Ergebnisse im C-Test und im Grammatiktest. Deutlichere Unterschiede gibt es hingegen bei der Textwiedergabe.

Dieser Eindruck wird bestätigt, wenn man für den Testtermin 'Januar 1992' die Schwierigkeiten der einzelnen Subtests der SAP über die Kollegs Aachen, Bochum, Bonn und Münster mittelt und die Standardabweichungen der gemittelten Schwierigkeiten berechnet.⁸ Die Resultate dieser Berechnungen finden sich in Tabelle 4.

Tabelle 4
SS 92 und WS 92/93
Mittlere Schwierigkeiten sowie Standardabweichungen der
mittleren Schwierigkeiten – gemittelt über die Kollegs in
Aachen, Bochum, Bonn und Münster

Subtest	mittlere Schwierigkeiten	Standardabweichung der mittleren Schwierigkeit
C-Test	57.4	1.9
GRAM	60.4	2.2
TW	49.2	6.7

GRAM = Grammatik; TW = Textwiedergabe

Erwartungsgemäß liegen die mittleren Schwierigkeiten von C-Test und Grammatik relativ dicht beieinander, während die Textwiedergabe deutlich schwieriger ist. Wichtig für die Bewertung der Subtests ist vor allem die erheblich höhere Standardabweichung bei der Textwiedergabe. Dieser Wert reflektiert die relativ hohen Unterschiede zwischen den Studienkollegs in diesem Testteil und ist vermutlich vor allem auf die wenig zufriedenstellende Auswertungsobjektivität des Subtests 'Textwiedergabe' zurückzuführen.

⁸ Die in Tabelle 4 aufgeführten Zahlen wurden von uns neu berechnet. Sie stimmen nur annäherungsweise mit den Werten in Nellessen (1992) überein, da dieser nicht die unterschiedlichen Stichprobenumfänge bei den vier Kollegs berücksichtigt hat.

In Tabelle 5 finden sich die von Nellessen (1992) für die einzelnen Studienkollegs angegebenen Standardabweichungen der jeweiligen Subtests.

Tabelle 5
SS 92 und WS 92/93
Standardabweichungen der Schwierigkeiten (%)

	Januar 1992					August 1992	
	Aachen	Bochum	Bonn	Köln*	Münster	Aachen	Münster
C-Test	17	15	17	(13)	20	21	19
GRAM	20	19	20	(15)	23	24	23
TW	27	25	24	(15)	27	-	31
TP	-	-	-	-	-	24	-

GRAM = Grammatik; TW = Textwiedergabe; TP = Textproduktion
 * Für Köln lagen nur die Ergebnisse der bestandenen Prüfungen vor.

Tabelle 5 zeigt, daß durchgängig bei jedem Kolleg der C-Test die geringste und die Textwiedergabe die größte Streuung aufweist. Dies sollte jedoch nicht notwendigerweise als Beleg für eine geringe bzw. hohe Differenzierungsfähigkeit dieser beiden Testteile gewertet werden. Es ist vielmehr zu vermuten, daß im Fall der Textwiedergabe ein großer Teil der Streuung durch Messfehler – und zwar insbesondere durch mangelnde Auswertungsobjektivität verursacht wird. Aufgrund der geringen Objektivität der Auswertung sind nicht nur Unterschiede zwischen den einzelnen Kollegs, sondern auch interindividuelle Variation bei den verschiedenen Auswertern ein und desselben Kollegs und sogar intraindividuelle Variation bei ein und demselben Auswerter zu erwarten. Die geringere Streuung beim C-Test dürfte dagegen vor allem auf die höhere Objektivität und Reliabilität dieses Testteils zurückzuführen sein.⁹

Abschließend möchten wir noch die von Nellessen (1992) angeführten Pearson-Korrelationen zwischen den Ergebnissen in der SAP (Testtermin Januar 92) und der Semesternote SS 92 angeben. Die Semesternote wurde im Juli 92 vergeben und beruht auf den Semesterklausuren sowie auf der

⁹ In der Regel zeichnen sich auch nicht vorher erprobte C-Tests durch eine vergleichsweise hohe Differenzierungsfähigkeit aus. Diese läßt sich häufig noch deutlich erhöhen, indem man in ihren Kennwerten nicht optimale Texte ersetzt.

Arbeit der Studierenden. Die Korrelationen finden sich zusammen mit den jeweiligen Stichprobengrößen in Tabelle 6.

Tabelle 6
SS 92
Korrelationen zwischen SAP (Januar 92) und Note SS 92

	Semester-Note		
	Bochum (N = 15)	Köln (N = 48)	Münster (N = 61)
C-Test	.03	.51	.38
GRAM	.47	.63	.43
TW	.43	.62	.48
gesamte SAP	.50	.67	.57

GRAM = Grammatik; TW = Textwiedergabe

Verglichen mit den in Tabelle 1 angegebenen Zahlen sind die Stichprobengrößen in Tabelle 6 deutlich reduziert. Darin spiegelt sich die Tatsache wider, daß wegen Nichtbestehens der SAP oder auch aus anderen Gründen relativ viele Probanden kein Studium an dem jeweiligen Kolleg aufgenommen haben. Aus diesem Grund sind auch die Streuungen reduziert – ein Sachverhalt, der bei der Beurteilung der Größe der Korrelationen zu berücksichtigen ist.

Aus Tabelle 6 ist ersichtlich, daß der C-Test die geringste Korrelation, das Gesamtergebnis in der SAP dagegen die höchste Korrelation mit der Semesternote aufweist. Allerdings sind wegen der geringen Stichprobengrößen alle Korrelationen mit einer erheblichen statistischen Unsicherheit behaftet. Dies gilt insbesondere für das Studienkolleg Bochum (N = 15). Dort ist nach Nellessen (1992) der fehlende Zusammenhang in erster Linie auf zwei 'Ausreißer' zurückzuführen. Wir wollen deshalb vorerst den Zusammenhang zwischen SAP und Semesternote nicht weiter thematisieren (vgl. jedoch die Ausführungen am Ende des Abschnitts 3.2.3).

3.2 WS 1994/95

Im folgenden werden wir relativ ausführlich auf die Resultate der SAP im WS 94/95 eingehen. Dabei wird das Studienkolleg Münster im Vordergrund stehen, da von diesem Kolleg die detailliertesten Daten zur Verfügung standen. Im Unterschied zu den Kollegs in Aachen, Bochum, Bonn und Köln lagen aus Münster die Ergebnisse in den einzelnen C-Test-Texten vor. Dadurch war zumindest in bezug auf die Münsteraner Probanden eine Analyse der Reliabilität und Dimensionalität des eingesetzten C-Tests möglich.

Im WS 94/95 umfaßte die SAP einen aus acht Texten bestehenden C-Test, einen Grammatiktest und eine Textwiedergabe.¹⁰ Die Zahl der Probanden betrug in Aachen 86, in Bochum 41, in Bonn 99, in Köln 149 und in Münster 145. Für die folgenden Rechnungen haben wir lediglich diejenigen Teilnehmer berücksichtigt, die alle drei Subtests bearbeitet haben. Dadurch reduziert sich der Stichprobenumfang in Aachen auf 85, in Bonn auf 97, in Köln auf 146 und in Münster auf 141.

3.2.1 Herkunftsland und Kurstyp: Unterschiede zwischen den Studienkollegs

Als erstes soll geprüft werden, ob sich die fünf Studienkollegs hinsichtlich der Variablen 'Herkunftsland' und 'Kurstyp' unterscheiden.¹¹ Tabelle 7 zeigt die Häufigkeiten der Herkunftsländer der Prüfungsteilnehmer in alphabetischer Reihenfolge. Insgesamt stammen die Prüfungsteilnehmer aus 64 Ländern. In Tabelle 7 sind jedoch lediglich die Länder aufgeführt, die zumindest bei einem Kolleg eine Häufigkeit von ≥ 5 aufweisen.

31% der 510 Teilnehmer kommen aus Marokko, das damit das weitaus größte Kontingent stellt. Es folgen der Iran mit 10% und Jordanien mit 6% (vgl. die Spalte 'Gesamt' in Tabelle 8). Allerdings gibt es länderspezifische Schwerpunkte bei den einzelnen Kollegs. So finden sich z.B. die Teilnehmer aus Indonesien schwerpunktmäßig in Aachen, die Teilnehmer aus dem Gaza-Streifen dagegen fast ausschließlich in Bonn.

¹⁰ Die Prüfung wurde am 13. August 1994 durchgeführt. Die einzelnen Subtests gingen mit gleicher Gewichtung in das Endergebnis ein.

¹¹ Alle statistischen Analysen wurden mit dem Programmpaket SPSS/PC+ (Version 4.0) durchgeführt.

Tabelle 7
Studienkollegs WS 94/95
absolute und relative (%) Häufigkeiten der Herkunftsländer der
Prüfungsteilnehmer (alphabetische Reihenfolge)

Land	Aachen		Bochum		Bonn		Köln		Münster	
	f_i	%	f_i	%	f_i	%	f_i	%	f_i	%
Brasilien	1	1.2	-	-	5	5.2	6	4.1	3	2.1
Gaza	-	-	-	-	11	11.3	1	0.7	-	-
Griechenland	-	-	1	2.4	1	1.0	5	3.4	4	2.8
Indonesien	13	15.3	1	2.4	3	3.1	4	2.7	4	2.8
Iran	4	4.7	1	2.4	14	14.4	13	8.9	20	14.2
Jordanien/ Westbank	10	11.8	-	-	7	7.2	5	3.4	9	6.4
Jugoslawien	2	2.4	4	9.8	-	-	12	8.2	1	0.7
Marokko	36	42.4	13	31.7	13	13.4	41	28.1	53	37.6
Peru	-	-	-	-	5	5.2	-	-	2	1.4
Rußland	1	1.2	3	7.3	3	3.1	9	6.2	2	1.4
Syrien	3	3.5	-	-	3	3.1	1	0.7	5	3.5
Ukraine	-	-	3	7.3	-	-	5	3.4	2	1.4
übrige Länder	14	16.5	14	34.1	31	32.0	44	30.1	29	20.6
keine Angabe	1	1.2	1	2.4	1	1.0	1	0.7	7	5.0
Gesamt	85	100	41	100	97	100	146	100	141	100

Tabelle 8 zeigt die Häufigkeiten der Kurstypen für die 11 am häufigsten genannten Herkunftsländer ($f_i \geq 10$). Insgesamt gesehen haben die Teilnehmer vor allem T-Kurse gewählt. Es folgen M-Kurse sowie mit Abstand W-Kurse und S/G-Kurse. Allerdings gibt es länderspezifische Präferenzen. Z.B. haben die Teilnehmer aus dem Iran und dem Gazastreifen vor allem M-Kurse gewählt, die Teilnehmer aus Marokko dagegen vor allem T-Kurse. Testet man den Zusammenhang zwischen 'Kurstyp' und 'Herkunftsland' auf Signifikanz, erhält man $\chi^2 = 97.5$ mit $p < .00000$. Für Cramérs V und den Kontingenzkoeffizienten C_c ergeben sich .25 bzw. .40. Es besteht somit ein hochsignifikanter, jedoch nicht sehr starker Zusammenhang zwischen den Variablen 'Kurstyp' und 'Herkunftsland'.

Wie die Tabelle 9 zeigt, gibt es allerdings in bezug auf die Kurswahl deutliche Unterschiede zwischen den Studienkollegs. So werden in Aachen

Tabelle 8
Studienkollegs WS 94/95
absolute und relative (%) Häufigkeiten der Kurstypen bei den
11 häufigsten Herkunftsländern (alphabetische Reihenfolge)

Land	Kurs									
	S/G		M		T		W		Gesamt	
	f_i	%	f_i	%	f_i	%	f_i	%	f_i	%
Brasilien	4	26.7	4	26.7	4	26.7	3	20.0	15	2.9
Gaza	1	8.3	10	83.3	-	-	1	8.3	12	2.4
Griechenland	3	27.3	5	45.5	2	18.2	1	9.1	11	2.2
Indonesien	1	4.0	-	-	15	60.0	9	36.0	25	4.9
Iran	4	7.7	32	61.5	10	19.2	6	11.5	52	10.2
Jordanien	4	12.9	8	25.8	13	41.9	6	19.4	31	6.1
Jugoslawien	3	15.8	7	36.8	7	36.8	2	10.5	19	3.7
Marokko	25	16.0	23	14.7	78	50.0	30	19.2	156	30.6
Rußland	5	27.8	1	5.6	9	50.0	3	16.7	18	3.5
Syrien	3	25.0	5	41.7	2	16.7	2	16.7	12	2.4
Ukraine	1	10.0	4	40.0	3	30.0	2	20.0	10	2.0
übrige Länder	26	17.4	46	30.9	50	33.6	27	18.1	149	29.2
Gesamt	80	15.7	145	28.4	193	37.8	92	18.0	510	100

Tabelle 9
Studienkollegs WS 94/95
absolute und relative (%) Häufigkeiten der Kurstypen

Kolleg	Kurs									
	S/G		M		T		W		N	
	f_i	%	f_i	%	f_i	%	f_i	%		
Aachen	-	-	-	-	66	77.6	19	22.4	85	
Bochum	-	-	-	-	41	100.0	-	-	41	
Bonn	16	16.5	38	39.2	30	30.9	13	13.4	97	
Köln	46	31.5	53	36.3	20	13.7	27	18.5	146	
Münster	18	12.8	54	38.3	36	25.5	33	23.4	141	
Gesamt	80	15.7	145	28.4	193	37.8	92	18.0	510	

T- und W-Kurse und in Bochum ausschließlich T-Kurse angeboten. Weiterhin liegt der Schwerpunkt der S/G-Kurse in Köln. Der Zusammenhang zwischen 'Kurswahl' und 'Studienkolleg' ist erwartungsgemäß signifikant ($\chi^2 = 210.1$; $p < .00000$) und zudem mit $V = .37$ und $\phi = .54$ stärker ausgeprägt als im Fall von 'Kurstyp' und 'Herkunftsland'.

3.2.2 Studienkolleg Münster: Analyse der SAP-Ergebnisse

Bereits erwähnt, liegen für das Studienkolleg Münster die ausführlichen Daten vor. Diese Daten sollen nun einer detaillierten Analyse unterzogen werden.

In Tabelle 10 sind für die acht Texte des C-Tests, für den C-Test insgesamt, für den Grammatikteil und für die Textwiedergabe folgende Kennwerte aufgeführt: mittlerer Punktwert, Standardabweichung, beobachtetes Minimum und beobachtetes Maximum. Bei den Variablen 'C-TEST', 'GRAM' (Grammatik) und 'TW' (Textwiedergabe) handelt es sich jeweils um Prozentwerte. Das theoretische Maximum liegt somit bei 100. Bei den C-Test-Texten liegt das theoretische Maximum bei 25 (Zahl der Lücken pro Text).

Tabelle 10
Münster WS 94/95
Kennwerte der SAP (N = 141)

Variable	Mittel	Standardabweichung	Minimum (beobachtet)	Maximum (beobachtet)
TEXT 1	21.74	2.49	14	25
TEXT 2	17.60	4.03	6	25
TEXT 3	15.84	4.09	1	25
TEXT 4	14.94	5.30	0	25
TEXT 5	12.08	5.60	0	25
TEXT 6	15.36	5.28	0	25
TEXT 7	11.71	6.76	0	25
TEXT 8	9.50	5.86	0	25
C-TEST (%)	59.71	17.21	18	99
GRAM (%)	63.16	23.54	11	97
TW (%)	48.52	26.49	3	100

GRAM = Grammatik; TW = Textwiedergabe

Abgesehen von Text 6, der aufgrund seiner Schwierigkeit an vierter Stelle in der Textreihenfolge stehen müßte, entspricht die Schwierigkeit der Texte ihrer Reihenfolge. Die Fehlplatzierung von Text 6 ist auffallend, da es sich um eine vorerprobte C-Test-Version handelt. Sofern es sich dabei nicht um einen Zufallseffekt handelt, besteht bei Text 6 offensichtlich eine Wechselwirkung zwischen Textmerkmalen und Probandenmerkmalen.

Auffallend ist weiterhin, daß bei den Texten 4 bis 8 eine Reihe von Probanden 0 Punkte erzielt haben. Bei den Texten 4, 5 und 6 beträgt die Zahl der entsprechenden Probanden allerdings lediglich 1, 3 und 1, bei den Texten 7 und 8 dagegen 9 und 12. Die beiden letzten Texte scheinen damit von einer nicht unerheblichen Zahl von Prüflingen aus Zeitgründen erst gar nicht in Angriff genommen worden zu sein. Bei diesen Probanden ist der Gesamtwert im C-Test deswegen möglicherweise ein nur bedingt valider Indikator ihrer sprachlichen Leistungsfähigkeit. Bei zukünftigen Testdurchläufen sollte durch geeignete Instruktionen ausgeschlossen werden, daß Probanden aus Zeitgründen die letzten Texte nicht mehr bearbeiten (vgl. hierzu die Empfehlungen in Abschnitt 4).

Wie bereits in den Daten von Nellessen (1992) ist die Textwiedergabe am schwierigsten, der Grammatikteil am leichtesten und die Streuung bei der Textwiedergabe am höchsten (vgl. Abschnitt 3.1, Tabellen 3 und 5). Hinsichtlich der Streuungen ist bei der Textwiedergabe wiederum davon auszugehen, daß es sich wegen der nicht zufriedenstellenden Auswertungsobjektivität dieses Testteils zu einem erheblichen Teil um Meßfehlervarianz handelt (vgl. die Argumentation in Abschnitt 3.1)

Als nächstes soll die vorliegende C-Test-Version einer genaueren Analyse unterzogen werden. Tabelle 11 zeigt die Interkorrelationen und Trennschärfen der acht C-Test-Texte (Pearson-Produkt-Moment-Korrelationen). Wie aus Tabelle 11 ersichtlich ist, korrelieren die Texte relativ hoch miteinander. Die vergleichsweise geringe Korrelation zwischen Text 1 und Text 8 ist vermutlich auf die stark unterschiedlichen Schwierigkeiten (d.h. Lösungshäufigkeiten) der beiden Texte zurückzuführen.¹²

¹² In Text 1 ist die Verteilung der Punktwerte rechtsgipflig, in Text 8 dagegen linksgipflig. In einem solchen Fall kann die Korrelation zwischen den beiden Verteilungen nicht das theoretische Maximum von 1 erreichen.

Tabelle 11
Münster WS 94/95
Interkorrelationen und Trennschärfen (TS) der acht C-Test-Texte
(N = 141)

	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	TS
Text 1	.60	.66	.70	.59	.67	.64	.52	.70
Text 2		.71	.65	.68	.67	.69	.58	.74
Text 3			.75	.72	.78	.75	.71	.84
Text 4				.79	.80	.76	.72	.86
Text 5					.80	.80	.75	.86
Text 6						.78	.71	.87
Text 7							.81	.87
Text 8								.80

Auch die Trennschärfen (d.h. die korrigierten Korrelationen der Textpunktwerte mit dem Gesamtpunktwert im C-Test) sind hoch. Insgesamt gesehen scheinen die acht Texte weitgehend das gleiche zu messen.

Zur Abschätzung der Reliabilität haben wir Cronbachs Alpha sowie die Lambda-Koeffizienten von Guttman (1945) berechnet. Da die Lambda-Koeffizienten bei allen Reliabilitätsabschätzungen in diesem Abschnitt nicht zu höheren Werten als Cronbachs Alpha führten, haben wir sie nicht aufgeführt. Cronbachs Alpha beträgt .94. Der C-Test mißt somit die zugrundeliegende Fähigkeit hoch reliabel.¹³

Wie allerdings der hochsignifikante Wert für den Tukey-Test auf Nichtadditivität zeigt, sind die Texte deutlich nichtadditiv und messen damit nicht eindimensional im Sinne des klassischen latent-additiven Testmodells (KLA-Modell) von Moosbrugger & Müller (1981).¹⁴ Es liegt vielmehr eine Wechselwirkung zwischen Textschwierigkeit und Fähigkeit der Probanden vor.

Es soll nun untersucht werden, ob der C-Test verkürzt werden kann, ohne daß seine Meßgenauigkeit leidet. Tabelle 12 zeigt, daß eine Verkürzung auf 5 bzw. 4 Texte die Meßgenauigkeit nicht beeinträchtigt. Insbesondere die aus 5 Texten bestehende Version ist zudem weit weniger nichtadditiv

¹³ Eine Abschätzung der Reliabilität der übrigen Testteile der SAP war wegen fehlender Daten leider nicht möglich.

¹⁴ Zum KLA-Modell vgl. auch Tarnai (1989) sowie die Hinweise in Grotjahn (1992, S. 235ff.).

im Sinne des KLA-Modells als die vollständige Version. Da der C-Test vor allem im Bereich des für das Bestehen festgesetzten Schwellenwerts, der im WS 94/95 bei 60% lag, differenzieren soll, sollte der C-Test zudem nicht zu schwer sein. Es spricht deshalb nichts dagegen, den vorliegenden C-Test zwecks Erhöhung der Auswertungseffizienz auf vier Texte (plus einen nicht auszuwertenden sehr leichten 'Eisbrechertext') zu verkürzen.¹⁵

Tabelle 12
Münster WS 94/95
Kennwerte des C-Tests bei zwei verschiedenen Textauswahlen
(N = 141)

Textauswahl	Mittel	Standardabw.	Schwierigkeit	Alpha
Texte 4-8	63.58	26.09	50.86	.94
Texte 4-7	54.08	21.07	54.08	.93

Tabelle 13 zeigt die Interkorrelationen zwischen den drei Subtests der SAP. Neben den Pearson-Produkt-Moment-Korrelationen r sind auch die Werte für den Kendallschen Rangkorrelationskoeffizienten τ_c aufgeführt.¹⁶

Tabelle 13
Münster WS 94/95
Interkorrelationen der Subtests (N = 141)*

	Grammatik	Textwiedergabe
C-Test	.75 (.57)	.81 (.63)
Grammatik		.68 (.51)

* Pearson-Produkt-Moment-Korrelationen und Kendalls τ_c (in Klammern)

¹⁵ Dies bedeutet nicht, daß die getroffene Textauswahl optimal für die vorliegende Zielsetzung ist. Um eine optimale Textauswahl zu treffen, bedarf es weitergehender Analysen – z.B. auf der Basis des KLA-Modells.

¹⁶ Die Werte für das zusätzlich berechnete τ_b sind bei dieser und allen folgenden Rechnungen quasi identisch mit τ_c . Weiterhin haben wir Spearmans Rangkorrelationskoeffizienten r_s berechnet. Da r_s lediglich ein Spezialfall von Pearsons r darstellt und zudem bei allen durchgeführten Berechnungen der Zusammenhang zwischen den Variablen weitgehend linear ist, unterscheiden sich die Werte für r_s und r erwartungsgemäß bei den Rechnungen dieses Kapitels kaum. Auf die Angabe von Werten für r_s ist deshalb verzichtet worden.

Alle Korrelationen sind hochsignifikant ($p < .000$; einseitiger Test). Gibt man von den quadrierten Produkt-Moment-Korrelationen als Maß für den jeweiligen Anteil an gemeinsamer Varianz aus, dann können die Produkt-Moment-Korrelationen als substantiell angesehen werden.¹⁷ Die Subtests scheinen somit in nicht unerheblichem Maße die gleichen Fähigkeiten zu messen. Wie bereits bei Nellessen (1992) korrelieren C-Test und Textwiedergabe am höchsten (vgl. Tabelle 2 in Abschnitt 3.1). Allerdings sind die Kendallschen Tau-Koeffizienten beträchtlich niedriger als die Pearson-Korrelationen. Auf diesen Sachverhalt werden wir weiter unten noch eingehen.

Berechnet man Cronbachs Alpha für die drei Subtests als Superitem, ergibt sich ein Wert von .88. Der Tukey-Test auf Nichtadditivität ist wiederum hochsignifikant. Die SAP mißt somit nicht eindimensional im Sinne des KLA-Modells.

Es soll nun geprüft werden, ob die SAP zumindest im faktoranalytischen Sinne eindimensional ist. Tabelle 14 zeigt die Ergebnisse einer Faktorenanalyse mit Hilfe der Hauptachsenmethode. Es ergab sich lediglich ein einziger Faktor. Dieser hat einen Eigenwert von 6.99 und klärt 70% Varianz auf. Die SAP mißt somit eindimensional im faktoranalytischen Sinne. Wie aufgrund der hohen Interkorrelationen zwischen den C-Test-Texten zu erwarten war (vgl. Tabelle 11), laden auch alle C-Test-Texte hoch auf dem extrahierten Faktor.

Zuweilen wird argumentiert, daß Tests, deren Pearson-Produkt-Moment-Interkorrelation .80 oder höher ist, austauschbar seien (vgl. z.B. Oller & Streiff, 1975, S. 32f.). Wir wollen nun anhand der Daten aus dem Studienkolleg Münster zeigen, daß diese Folgerung äußerst problematisch ist. Dazu haben wir die Rangplätze der Prüflinge in den drei Subtests der SAP ermittelt, wobei für die niedrigste Punktzahl der Rang 1 und für die höchste Punktzahl der Rang 141 vergeben wurde. Die aufsteigend entsprechend dem Rang im C-Test geordneten Rangreihen finden sich in Tabelle 15.

Ein Blick auf Tabelle 15 zeigt, daß es eine Vielzahl von Prüflingen mit z.T. erheblichen Rangplatzdifferenzen in den drei Subtests der SAP gibt. In Anbetracht der nicht sehr hohen Werte für Kendalls τ_c in Tabelle 13 waren entsprechende Rangplatzdifferenzen auch zu erwarten. Im Unterschied zu Pearsons r basiert Kendalls τ nämlich auf der Zahl der Konkordanten und

¹⁷ Der Anteil an gemeinsamer Varianz liegt zwischen 46% und 66%.

Tabelle 14
Münster WS 94/95
Ergebnisse der Faktorenanalyse
(Hauptachsenmethode; N = 141)

Variable	Kommunalität
Text 1	.54
Text 2	.60
Text 3	.74
Text 4	.79
Text 5	.76
Text 6	.80
Text 7	.79
Text 8	.65
GRAM	.62
TW	.70

GRAM = Grammatik; TW = Textwiedergabe; Eigenwert des Faktors: 6.99

diskordanten Rangpaare und ist damit ein adäquateres Maß als Pearsons r , wenn die Rangordnung der Daten von Interesse ist.

Die extremsten Rangdifferenzen in Tabelle 15 finden sich mit +94.5 und -83.5 Plätzen zwischen Grammatik und Textwiedergabe. Es folgen C-Test und Grammatik mit +92.0 und -67.5 Plätzen sowie C-Test und Textwiedergabe mit +89.0 und -59.0 Plätzen. Die Zahlen bedeuten u.a., daß ein bestimmter Prüfling im Grammatikteil 94.5 Plätze höher liegt als in der Textwiedergabe, während zugleich ein anderer Prüfling in der Textwiedergabe 83.5 Plätze höher liegt als in der Grammatik.

Es sollen zumindest drei Fälle herausgegriffen werden. Der Prüfling mit der extremsten Rangdifferenz liegt im C-Test und im Grammatikteil mit den Rängen 120.5 und 126 in den **oberen** 15% der Rangskala, gleichzeitig jedoch bei der Textwiedergabe mit einem Rangplatz von 31.5 in den **unteren** 22% der Skala. Ein weiterer extremer Fall weist folgende Ränge auf: 129 (C-Test), 37 (Grammatik) und 91 (Textwiedergabe). Ein anderer Prüfling hat schließlich eine mittlere Position im C-Test (Rang 82.5), gleichzeitig jedoch einen deutlich niedrigeren Rang in der Textwiedergabe (Rang 43) und einen sehr viel höheren Rang im Grammatikteil (Rang 126).

Tabelle 15
Münster WS 94/95
Rangplätze der Prüflinge in den Subtests (geordnet aufsteigend
nach den C-Test-Ergebnissen) (N = 141)

C-Test	GRAM	TW	C-Test	GRAM	TW	C-Test	GRAM	TW
1.0	33.5	16.0	47.0	46.5	91.0	93.5	95.5	113.5
2.5	23.5	6.5	50.0	68.5	2.5	93.5	86.5	51.5
2.5	2.0	43.0	50.0	19.5	80.5	98.5	74.0	80.5
4.0	14.0	1.0	50.0	63.0	43.0	98.5	95.5	119.5
5.5	4.0	13.5	53.5	14.0	19.5	98.5	37.0	102.0
5.5	6.0	4.5	53.5	95.5	105.0	98.5	126.0	109.0
7.0	16.5	11.0	53.5	40.0	77.5	102.0	95.5	108.0
8.0	22.0	4.5	53.5	52.0	68.5	102.0	110.5	66.0
9.0	6.0	24.5	56.5	18.0	75.5	102.0	27.5	111.0
10.5	16.5	43.0	56.5	44.5	35.0	104.5	140.0	91.0
10.5	2.0	21.5	58.5	126.0	91.0	104.5	74.0	115.0
12.0	25.5	8.5	58.5	27.5	29.5	107.0	133.5	128.0
13.5	8.5	11.0	63.5	80.5	105.0	107.0	106.0	119.5
13.5	14.0	6.5	63.5	126.0	91.0	107.0	102.0	91.0
15.0	44.5	21.5	63.5	52.0	68.5	110.0	113.0	122.5
17.0	29.0	2.5	63.5	68.5	35.0	110.0	118.5	85.0
17.0	55.0	64.0	63.5	77.0	68.5	110.0	140.0	111.0
17.0	6.0	15.0	63.5	63.0	49.0	113.0	95.5	71.0
19.0	48.5	64.0	63.5	58.0	43.0	113.0	84.0	54.0
20.5	2.0	61.0	63.5	118.5	35.0	113.0	137.0	117.0
20.5	37.0	59.0	69.0	84.0	72.5	116.0	106.0	122.5
22.0	19.5	24.5	69.0	37.0	43.0	116.0	122.0	133.0
23.0	25.5	49.0	69.0	95.5	100.0	116.0	113.0	105.0
25.5	21.0	35.0	71.0	113.0	54.0	118.5	78.0	118.0
25.5	63.0	56.0	74.0	74.0	105.0	118.5	106.0	116.0
25.5	88.0	54.0	74.0	37.0	59.0	120.5	126.0	31.5
25.5	56.5	18.0	74.0	115.0	125.0	120.5	118.5	125.0
28.5	33.5	38.5	74.0	68.5	79.0	122.0	68.5	125.0
28.5	30.0	24.5	74.0	52.0	35.0	124.0	118.5	91.0
30.0	59.5	17.0	77.5	95.5	62.0	124.0	59.5	131.0
32.0	84.0	91.0	77.5	118.5	100.0	124.0	95.5	77.5
32.0	23.5	13.5	80.0	106.0	91.0	126.5	80.5	97.5
32.0	11.0	19.5	80.0	86.5	59.0	126.5	106.0	128.0
35.0	31.5	28.0	80.0	95.5	74.0	129.0	130.0	111.0
35.0	48.5	24.5	82.5	74.0	72.5	129.0	37.0	91.0
35.0	11.0	8.5	82.5	126.0	43.0	129.0	89.0	128.0
38.5	52.0	11.0	84.0	56.5	105.0	131.0	123.0	134.5
38.5	42.0	47.0	85.5	68.5	49.0	133.0	110.5	134.5
38.5	8.5	43.0	85.5	68.5	100.0	133.0	133.5	97.5
38.5	63.0	83.5	88.5	130.0	68.5	133.0	137.0	130.0
41.0	11.0	51.5	88.5	74.0	121.0	135.0	106.0	136.5
43.5	46.5	31.5	88.5	130.0	91.0	136.5	95.5	136.5
43.5	42.0	38.5	88.5	133.5	113.5	136.5	118.5	132.0
43.5	52.0	27.0	93.5	95.5	57.0	138.0	133.5	139.5
43.5	95.5	64.0	93.5	80.5	91.0	139.0	106.0	139.5
47.0	31.5	83.5	93.5	42.0	29.5	140.0	140.0	141.0
47.0	80.5	82.0	93.5	63.0	75.5	141.0	137.0	138.0

Tabelle 16 zeigt die Häufigkeiten der in Tabelle 15 zu findenden Rangplatzdifferenzen. Der Tabelle ist zu entnehmen, daß es die größten Rangunterschiede im Fall von Grammatik und Textwiedergabe gibt (Standardabweichung der Rangdifferenzen: 31.5). Es folgen C-Test-GRAM sowie C-Test-TW (Standardabweichung der Rangdifferenzen: 28.1 bzw. 25.3). Diese Reihenfolge spiegelt die Höhe der Interkorrelationen in Tabelle 13 wider. Bemerkenswert ist, daß trotz der hohen Interkorrelation von .81 zwischen C-Test und Textwiedergabe insgesamt 36 von 141 Prüflinge Rangdifferenzen von +30.0 bzw. -30.5 oder mehr aufweisen. Dies impliziert, daß die beiden Testteile zu deutlich unterschiedlichen Entscheidungen hinsichtlich Selektion und Einstufung führen können und damit trotz ihrer hohen Pearson-Produkt-Moment-Korrelation keineswegs prinzipiell austauschbar sind.

Der Tabelle 16 ist weiterhin zu entnehmen, daß die größten Rangplatzdifferenzen positiv sind. Dies bedeutet, daß die extremsten Rangdifferenzen folgender Art sind: (a) weit höherer Rang im C-Test als in der Grammatik; (b) weit höherer Rang im C-Test als in der Textwiedergabe; und (c) weit höherer Rang in der Grammatik als in der Textwiedergabe.

Die Bemerkungen zu den Rangplatzdifferenzen sollten lediglich aufzeigen, daß dann, wenn es um individuelle Entscheidungen auf der Basis von Rangplätzen geht, zwei Tests auch bei einer Pearson-Interkorrelation von .80 keineswegs austauschbar sind. Aufgrund der Definition des Pearson-Korrelationskoeffizienten waren die beobachteten zahlreichen und z.T. hohen Rangdifferenzen zudem theoretisch durchaus zu erwarten. Bei der faktischen Beurteilung der Güte der anhand der vorliegenden Tests vorgenommenen Selektionsentscheidungen müßten weiterhin die Rangdifferenzen speziell am *cutting point* (Punktwert des Bestehens der SAP) analysiert werden.

Es soll nun untersucht werden, ob Art und Höhe des Zusammenhangs zwischen den drei Testteilen vom sprachlichen Leistungsstand der Probanden abhängt. Da kein von den drei Testteilen unabhängiges Kriterium für die Leistung in Deutsch zur Verfügung stand, sind wir zunächst einmal von den C-Test-Ergebnissen als vermutlich zuverlässigstes Maß ausgegangen und haben die Gesamtstichprobe ($N = 141$) am Median der C-Test-Werte in eine obere Leistungsgruppe (≥ 59 ; $N = 71$) und eine untere Leistungsgruppe (< 59 ; $N = 70$) geteilt. Für die beiden Leistungsgruppen wurden dann getrennt die Werte für Pearsons r und Kendalls τ_c ermittelt. Die

Tabelle 16
Münster WS 94/95
Häufigkeiten der Rangplatzdifferenzen in den Subtests ($N = 141$)

Intervall	C-Test-GRAM	C-Test-TW	GRAM-TW
< - 80	-	-	1
-80 bis -70.5	-	-	1
-70 bis -60.5	3	-	2
-60 bis -50.5	3	3	5
-50 bis -40.5	8	7	7
-40 bis -30.5	6	10	8
-30 bis -20.5	11	10	9
-20 bis -10.5	13	14	17
-10 bis -0.5	23	30	23
0 bis 9.5	29	21	14
10 bis 19.5	14	18	19
20 bis 29.5	10	12	12
30 bis 39.5	11	9	13
40 bis 49.5	4	4	4
50 bis 59.5	2	1	1
60 bis 69.5	2	1	2
70 bis 79.5	1	-	-
80 bis 89.5	-	1	2
≥ 90	1	-	1

C-Test-GRAM = Rangplatzdifferenz C-Test und Grammatik

C-Test-TW = Rangplatzdifferenz C-Test und Textwiedergabe

GRAM-TW = Rangplatzdifferenz Grammatik und Textwiedergabe

Ergebnisse finden sich in Tabelle 17 (τ_c ist wiederum durch Klammerung gekennzeichnet).

In Tabelle 17 fällt vor allem die relativ geringe Korrelation zwischen dem Grammatiktest und den übrigen Testteilen in der oberen Leistungsgruppe ins Auge. Allerdings ist zu berücksichtigen, daß die Standardabweichung des Grammatiktests in der oberen Leistungsgruppe deutlich niedriger als in der unteren Leistungsgruppe ist.¹⁸ Berechnet man den Standardschätzfeh-

¹⁸ Die Standardabweichungen für C-Test, Grammatik und Textwiedergabe be-

Tabelle 17
Münster WS 94/95

Interkorrelationen der Subtests getrennt für die obere und untere Leistungsgruppe (geteilt am Median des C-Tests bzw. am Median des Grammatiktests)*

	C-Test \geq 59 ($N = 71$)		C-Test $<$ 59 ($N = 70$)	
	GRAM	TW	GRAM	TW
C-Test	.34 (.25)	.64 (.49)	.60 (.45)	.57 (.41)
GRAM		.31 (.23)		.54 (.37)
	GRAM $>$ 67 ($N = 71$)		GRAM \leq 67 ($N = 70$)	
	GRAM	TW	GRAM	TW
C-Test	.39 (.26)	.63 (.46)	.55 (.42)	.71 (.47)
GRAM		.26 (.19)		.40 (.28)

GRAM = Grammatik; TW = Textwiedergabe

* Pearson-Produkt-Moment-Korrelationen und Kendalls τ_c (in Klammern); alle Korrelationen signifikant mit $p < .01$.

ler als Maß für die prädiktive Kraft eines Testteils in bezug auf die übrigen Testteile, erhält man für die Vorhersage der Grammatikwerte aus den C-Test-Werten für die obere Gruppe einen Wert von 12 Punkten, für die untere Gruppe dagegen einen Wert von 17 Punkten. Für die Vorhersage der Resultate des Grammatiktests anhand der Textwiedergabe betragen die entsprechenden Werte 13 und 18 Punkte. Der mittlere absolute Fehler bei der Vorhersage der Grammatikwerte aus den übrigen Testteilen ist damit in der oberen Leistungsgruppe niedriger als in der unteren Gruppe.

Teilt man die Gesamtstichprobe nicht am Median der C-Test-Werte, sondern am Median der Grammatikwerte ($Med = 67$) in zwei Leistungsgruppen, ergeben sich bis auf zwei abweichende Korrelationen in der unteren Leistungsgruppe weitgehend die gleichen Werte. In der unteren Leistungsgruppe ist jetzt die Korrelation zwischen Textwiedergabe und Grammatik mit $r = .40$ und $\tau_c = .28$ etwas niedriger und die Korrelation zwischen C-Test und Textwiedergabe mit $r = .71$ und $\tau_c = .47$ deutlich höher als in der am Median des C-Tests geteilten Stichprobe. Insgesamt gesehen bestätigen

tragen in der oberen Leistungsgruppe 9.6, 13.2 und 21.6, in der unteren Leistungsgruppe dagegen 9.8, 21.3 und 17.1.

jedoch die bei einer Aufteilung am Median der Grammatikwerte ermittelten Korrelationen die zuvor erhaltenen Resultate.

In folgenden soll der Frage nachgegangen werden, ob es möglicherweise nationalitäten- und kursspezifische Unterschiede bei den Leistungen in den einzelnen Testteilen gibt. Tabelle 18 zeigt Mittelwert, Standardabweichung und Median der Subtestwerte für die vier Kurstypen. In Tabelle 19 sind die gleichen Kennwerte für die am Studienkolleg Münster am häufigsten genannten Herkunftsländer Marokko und Iran sowie die Gruppe der übrigen Länder aufgeführt.¹⁹

Tabelle 18
Münster WS 94/95

Kurstypen: Mittelwert (M), Standardabweichung (s) und Median (Med) der Subtestwerte ($N = 141$)

Kurstyp	N	C-Test			Grammatik			Textwiedergabe		
		M	s	Med	M	s	Med	M	s	Med
S/G	18	61.3	17.9	61.5	55.2	29.4	67.0	50.1	27.6	44.5
M	54	60.3	18.3	60.5	64.2	24.5	69.0	51.8	27.5	49.0
T	36	56.6	15.8	56.5	64.6	22.5	66.0	41.6	25.3	34.5
W	33	61.2	16.8	66.0	64.1	19.5	67.0	49.9	25.2	51.0
Gesamt	141	59.7	17.2	59.0	63.2	23.5	67.0	48.5	26.5	47.0

Tabelle 19
Münster WS 94/95

Iran, Marokko und sonstige Herkunftsländer: Mittelwert (M), Standardabweichung (s) und Median (Med) der Subtestwerte ($N = 141$)

Land	N	C-Test			Grammatik			Textwiedergabe		
		M	s	Med	M	s	Med	M	s	Med
Iran	20	61.5	12.7	63.5	75.2	14.7	79.0	52.4	19.9	49.0
Marokko	53	52.7	15.6	51.0	57.8	22.2	58.0	38.1	21.2	33.0
Sonstige	68	64.7	17.8	66.0	63.8	25.4	73.0	55.5	29.4	60.0
Gesamt	141	59.7	17.2	59.0	63.2	23.5	67.0	48.5	26.5	47.0

¹⁹ Wegen der geringen Fallzahlen haben wir auf eine detailliertere Aufteilung nach Nationalitäten verzichtet.

Wie der Tabelle 18 zu entnehmen ist, liegt möglicherweise ein schwacher Effekt der Variablen 'Kurstyp' vor. Beim C-Test und der Textwiedergabe sind die Mittelwerte für den T-Kurs niedriger als für die übrigen Kurse. Im Grammatiktest weist hingegen der S/G-Kurs einen im Vergleich zu den übrigen Kursen niedrigeren Mittelwert auf.

Bei den Mittelwerten für die Nationalitäten in Tabelle 19 fällt zum einen auf, daß Marokko beim C-Test und der Textwiedergabe deutlich niedrigere Werte als der Iran und die übrigen Länder aufweist. Weiterhin ist der vergleichsweise hohe Wert für den Iran im Grammatikteil bemerkenswert.

Bei der Interpretation dieser Befunde ist allerdings zu beachten, daß die Variablen 'Herkunftsland' und 'Kurstyp' nicht unabhängig voneinander sind. So zeigt Marokko eine überproportionale Frequenz insbesondere bei den T-Kursen, und abgeschwächt auch bei den W-Kursen. Bei den Prüflingen aus dem Iran ist hingegen vor allem der M-Kurs überproportional häufig vertreten.

Testet man die Variablen 'Herkunftsland' (Ausprägungen: 'Marokko', 'Iran', 'Sonstige') und 'Kurstyp' auf Unabhängigkeit, erhält man $\chi^2 = 27.18$ mit $df = 6$ und $p < .0001$. Der Zusammenhang zwischen den beiden Variablen ist damit hochsignifikant. Für Cramérs V und den Kontingenzkoeffizienten C_c ergeben sich jedoch lediglich Werte von .31 und .40. Die Abhängigkeit zwischen 'Herkunftsland' und 'Kurstyp' ist somit nicht sehr stark ausgeprägt.

Es soll nun mit Hilfe von Varianzanalysen getestet werden, ob der beobachtete Effekt der Variablen 'Herkunftsland' und 'Kurstyp' signifikant ist. Um zu überprüfen, ob die Voraussetzungen für die Anwendung parametrischer Verfahren erfüllt sind, haben wir eine Reihe von explorativen Datenanalysen mit Hilfe der SPSS-Prozeduren EXAMINE und MANOVA durchgeführt (vgl. zu den folgenden statistischen Analysen Norušis, 1990a, 1990b, 1990c).

Beim C-Test ist die Normalverteilungsannahme auch bei einer Differenzierung nach Ländergruppen bzw. Kurstypen im Vergleich zu den übrigen Testteilen am ehesten erfüllt. Dies gilt auch für die Homoskedastizitätsannahme (Gleichheit der Streuungen bei einer Differenzierung nach Ländergruppen bzw. Kurstypen).

Die Verteilung der Grammatiktestwerte ist hingegen schief (rechtsgipflig) und zu flach (vgl. das Verhältnis von Mittelwert und Median). Der Lilliefors-Test auf Normalität ist entsprechend hochsignifikant ($p = .0007$).

Auch nach Ländergruppen bzw. Kurstypen differenziert ergeben sich bei der Grammatik zumeist deutliche Abweichungen von der Normalität (Shapiro-Wilk-Test oder Lilliefors-Test). Auch der Levene-Test auf Varianzgleichheit ist sowohl für die Ländergruppen als auch für die Kurstypen im Fall der Grammatik signifikant ($p = .021$ bzw. $p = .048$).

Die Verteilung der Werte für die Textwiedergabe ist in der Gesamtgruppe insbesondere deutlich zu flach (Lilliefors-Test: $p = .002$). Bei den Ländergruppen ist lediglich im Fall des Iran die Verteilung annähernd normal. Bei den Kurstypen ist die Anpassung an die Normalität etwas besser. Bei den Ländergruppen ist der Levene-Test auf Streuungsgleichheit auch im Fall der Textwiedergabe signifikant ($p < .001$), bei den Kurstypen sind die Varianzen jedoch deutlich homogen ($p = .968$).

Führt man nun trotz der nachgewiesenen Verletzungen der Anwendungsvoraussetzungen für die drei Subtests jeweils univariate parametrische Varianzanalysen mit dem Faktor 'Kurstyp' durch, erhält man in allen drei Fällen F -Werte, deren Wahrscheinlichkeit weit unter der 5%-Schwelle liegen. Eine multivariate Varianzanalyse mit den drei Subtests als abhängige Variable (SPSS-Prozedur MANOVA) ergibt jedoch für die Statistiken von Pillai, Hotelling und Wilk jeweils einen Wert von .07. Es liegt somit möglicherweise zumindest ein globaler Effekt des Faktors 'Kurstyp' vor.

Es soll nun eine multivariate parametrische Varianzanalyse (MANOVA) mit dem Faktoren 'Herkunftsland' und den drei Subtests als abhängige Variablen durchgeführt werden.²⁰ Wegen der Verletzungen der Anwendungsvoraussetzungen sind die Resultate der MANOVA allerdings wie bereits zuvor im Fall des Faktors 'Kurstyp' mit Vorsicht zu interpretieren.²¹ Wir

²⁰ Ein Test auf Interaktion der Faktoren 'Herkunftsland' und 'Kurstyp' war wegen zu geringer Gruppengrößen nicht möglich.

²¹ Zur Überprüfung der Voraussetzungen der MANOVA wurden neben den bereits erwähnten Analysen eine Reihe weiterer Überprüfungen durchgeführt (Analyse der Residuen; Berechnung von Box M ; graphische Tests). Die festgestellten Verletzungen der Voraussetzungen waren jedoch nicht so gravierend, daß ein Verzicht auf die Durchführung der MANOVA angeraten gewesen wäre. So ergab sich z.B. für den multivariaten Test der Homogenität der Gruppen-Kovarianzmatrizen (Box M) eine Wahrscheinlichkeit von $p = .09$. Die M -Statistik von Box ist allerdings ihrerseits weit empfindlicher gegenüber Verletzungen der Normalverteilungsannahme als z.B. die weiter unten durchgeführten univariaten F -Tests zur Überprüfung der Gleichheit von Mittelwerten.

haben deshalb zur weiteren Absicherung etwaiger Folgerungen die parametrische multivariate Analyse durch nichtparametrische univariate Analysen ergänzt.

Die Statistiken von Pillai, Hotelling und Wilk belegen einen hochsignifikanten Effekt der Variablen 'Herkunftsland' auf die drei abhängigen Variablen ($p < .000$). In Tabelle 20 sind die Ergebnisse der univariaten Varianzanalysen aufgeführt. Es wurde sowohl parametrisch (F -Test) als auch nichtparametrisch (Kruskal-Wallis) getestet.

Tabelle 20

Iran, Marokko und sonstige Herkunftsländer: parametrische und nichtparametrische Varianzanalysen ($N = 141$)

	F -Test			Kruskal-Wallis	
	F	p	η^2	χ^2	p
C-Test	8.1	.000	.105	14.2	.001
Grammatik	4.2	.017	.057	8.8	.012
Textwiedergabe	7.3	.001	.096	12.8	.002

Wie sowohl der parametrische als auch der nichtparametrische Test zeigen, ist der Effekt des Faktors 'Herkunftsland' bei allen Prüfungsteilen signifikant, im Fall von C-Test und Textwiedergabe sogar hochsignifikant.²² Wie allerdings die Werte für das Effektmaß η zeigen, ist der Effekt bezogen auf den einzelnen Prüfungsteil relativ schwach (es werden maximal 10% Varianz durch den Faktor 'Herkunftsland' aufgeklärt).

Um festzustellen, welche Mittelwerte aus Tabelle 19 signifikant unterschiedlich sind, wurden sowohl multiple t -Tests als auch multiple U -Tests (nach Mann-Whitney) gerechnet, wobei jeweils das Signifikanzniveau mit Hilfe der Methode von Bonferroni und Holm adjustiert wurde (vgl. zu dieser Adjustierung Holm, 1979 sowie auch Grotjahn, Krause & von Unwerth, 1987, S. 151f.; Stevens, 1990, S. 137). Obwohl sich eine Reihe von Unterschieden zwischen den parametrischen t -Tests und den nichtparametrischen U -Tests ergaben,²³ führten beide Verfahren nach Adjustierung des Signifikanzniveaus

²² Diese Aussage gilt selbst dann, wenn man das Signifikanzniveau adjustiert.

²³ Geht man beim t -Test von Varianzungleichheit aus ($F = 2.99$; $p = .010$), ergibt sich z.B. beim Grammatikteil für den Vergleich 'Iran gegen sonstige

Entscheidungen bezüglich der Signifikanz der beobachteten Mittelwertunterschiede. Wir haben deshalb in Tabelle 21 lediglich die Werte des U -Test aufgeführt – und zwar die den U -Werten entsprechenden z -Werte (Bindungskorrigiert; ohne Vorzeichen) sowie die zweiseitigen Überlappungswahrscheinlichkeiten. Die nach dem Bonferroni-Holm-Verfahren auf der 5%-Ebene signifikanten Differenzen sind mit einem Stern gekennzeichnet.

Tabelle 21

Münster WS 94/95

Iran, Marokko und sonstige Herkunftsländer:
 z -Werte und zweiseitige Wahrscheinlichkeiten (p)
für den U -Test von Mann-Whitney ($N = 141$; multiple
Vergleiche mit Korrektur für Bindungen)

	Land	Marokko		Sonstige	
		z	p	z	p
C-Test	Iran	2.15	.032	.88	.381
	Marokko			3.64*	.000
GRAM	Iran	3.09*	.002	1.51	.130
	Marokko			-.70	.089
TW	Iran	2.43	.015	-.51	.611
	Marokko			3.34*	.001

GRAM = Grammatik; TW = Textwiedergabe

* signifikant auf der 5%-Ebene (nach Adjustierung des Signifikanzniveaus entsprechend Bonferroni und Holm)

Tabelle 21 zeigt beim C-Test einen signifikanten Lokationsunterschied zwischen Marokko und dem Rest der Herkunftsländer und einen nahezu signifikanten Unterschied zwischen Marokko und dem Iran. Beim Grammatikteil erreicht der Unterschied zwischen Marokko und dem Iran das 5%-Niveau. Bei der Textwiedergabe ist schließlich der Unterschied zwischen Marokko und den übrigen Ländern signifikant, der Unterschied zwischen

Länder' ein p -Wert von .015, beim U -Test hingegen ein p -Wert von .130. Geht man von Varianzgleichheit aus, beträgt der p -Wert für den t -Test dagegen .062 und liegt damit weit näher beim Resultat des U -Tests.

Marokko und dem Iran nahezu signifikant. Es liegt damit folgende Situation vor: Marokko zeigt sowohl beim C-Test als auch bei der Textwiedergabe deutlich schlechtere Leistungen als die Restländer und möglicherweise in beiden Testteilen auch schlechtere Leistungen als der Iran. Faßt man die Kategorien 'Marokko' und 'Sonstige' zusammen, dann liegen die Leistungen der Prüflinge aus dem Iran in ihren beobachteten Mittelwerten stets über der Gesamtheit der übrigen Länder – beim C-Test und der Textwiedergabe allerdings nur geringfügig. Deutlich heben sich die Iraner im Bereich 'Grammatik' ab. Dort zeigen sie weit höhere Leistungen als die Teilnehmer aus Marokko und auch noch beträchtlich höhere Leistungen als die Prüflinge aus der Kategorie 'sonstige Länder', wobei allerdings der letztere Unterschied keine Signifikanz erreicht.

In Tabelle 19 fallen neben den Mittelwertunterschieden die nicht unerheblichen Streuungsunterschiede zwischen den Nationalitätengruppen insbesondere beim Grammatiktest und bei der Textwiedergabe ins Auge. Wir hatten bereits mit Hilfe des robusten Levene-Tests festgestellt, daß die beobachteten Streuungsunterschiede im Fall des Grammatiktests und der Textwiedergabe signifikant bzw. hochsignifikant sind ($p = .021$ bzw. $p = .001$) und im Fall des C-Tests nahe der 10%-Schwelle liegen ($p = .131$). Bemerkenswert ist, daß die Teilnehmer aus dem Iran in allen drei Subtests die geringsten Streuungen aufweisen. Besonders deutlich ist der Unterschied im Fall von Grammatik und Textwiedergabe. In beiden Fällen besteht ein sehr großer Unterschied zur Gruppe der sonstigen Länder. Bei der Grammatik gibt es zudem noch einen weiteren, allerdings etwas kleineren Streuungsunterschied zur Gruppe der Marokkaner. Damit zeichnen sich die Iraner nicht nur durch eine ungewöhnlich hohe mittlere Leistung im Grammatiktest aus, sondern gleichzeitig auch durch eine vergleichsweise hohe **Homogenität** der Leistungen insbesondere im Grammatiktest und in der Textwiedergabe.

Es soll nun mit Hilfe einer rudimentären Diskriminanzanalyse geprüft werden, inwieweit die drei Teilstests eine Differenzierung zwischen den drei Ländergruppen erlauben. Ziel dieser Analyse ist in erster Linie eine grobe Abschätzung der Stärke des in Tabelle 20 festgestellten Effekts. Wir haben deshalb auch keine vollständige Diskriminanzanalyse mit Hilfe der SPSS-Prozedur DISCRIMINANT durchgeführt, sondern lediglich einige diskrimi-

nanzanalytische Möglichkeiten der SPSS-Prozedur MANOVA mit den jeweiligen Standardeinstellungen genutzt.²⁴

Es wurden zwei kanonische Diskriminanzfunktionen mit Eigenwerten von .20 und .06 extrahiert. Die beiden Funktionen erklären 76% bzw. 24% Varianz. Tabelle 22 zeigt die Korrelationen zwischen diskriminierenden Variablen und kanonischen Diskriminanzfunktionen.

Tabelle 22
Münster WS 94/95
Standardisierte kanonische Diskriminanzfunktionskoeffizienten
sowie Korrelationen zwischen diskriminierenden Variablen und
kanonischen Diskriminanzfunktionen (N = 141)

diskriminierende Variable	Korrelationen	
	Funktion 1	Funktion 2
C-Test	-.67	.67
Grammatik	-.08	.98
Textwiedergabe	-.62	.71

Die Korrelationen zwischen diskriminierenden Variablen und Diskriminanzfunktionen (sog. Strukturkoeffizienten) weisen für die Variablen 'C-Test' und 'Textwiedergabe' in beiden Funktionen ähnliche Werte auf. Dagegen korreliert der Teilstest 'Grammatik' lediglich mit der 2. Diskriminanzfunktion – und zudem sehr hoch. Der Grammatikteil der SAP scheint sich damit von den beiden übrigen Testteilen in bezug auf seine länderspezifische Differenzierungsfähigkeit zu unterscheiden.

Es soll nun noch untersucht werden, ob es auch länderspezifische Muster bei den Interkorrelationen zwischen den Subtests gibt. Tabelle 23 zeigt die Interkorrelationen der Subtests getrennt für die unterschiedenen Kategorien von Herkunftsländern.

Die Korrelationsmatrizen Marokkos und der Gruppe der sonstigen Herkunftsländer unterscheiden sich nur unwesentlich und weichen auch von der Interkorrelationsmatrix der Gesamtgruppe (vgl. Tabelle 13) kaum ab. Dagegen finden sich in der Gruppe der Iraner bei den Korrelationen mit der Variablen 'Textwiedergabe' auffällige Divergenzen, die allerdings wegen des

²⁴ Für diese Analyse gelten ähnliche Einschränkungen wie für die durchgeführten (multivariaten) Varianzanalysen.

Tabelle 23
Münster WS 94/95
Interkorrelationen der Subtests getrennt für drei Kategorien
von Herkunftsländern*

	Iran (N = 20)		Marokko (N = 53)		Sonstige (N = 68)	
	GRAM	TW	GRAM	TW	GRAM	TW
C-Test	.75 (.52)	.59 (.42)	.79 (.64)	.78 (.62)	.76 (.56)	.83 (.67)
GRAM		.39 (.28)		.70 (.57)		.71 (.53)

GRAM = Grammatik; TW = Textwiedergabe

*Pearson-Produkt-Moment-Korrelationen und Kendalls τ_c (in Klammern)

geringen Stichprobenumfangs mit Vorsicht zu interpretieren sind. Zum einen ist die Korrelation zwischen C-Test und Textwiedergabe vergleichsweise niedrig. Auffallend ist jedoch vor allem die geringe Korrelation zwischen Grammatik und Textwiedergabe. Während alle übrigen Korrelationen in Tabelle 22 auf dem 1%-Niveau signifikant sind, ist die Korrelation zwischen Grammatik und Textwiedergabe in der Gruppe der Iraner nicht einmal auf dem 5%-Niveau signifikant ($p = .09$). Dies deckt sich mit den Befunden in Tabelle 17: Auch dort korrelieren in der oberen Leistungsgruppe, zu der die Iraner tendenziell gehören, Grammatik und Textwiedergabe am niedrigsten.

Abschließend soll noch ein Blick auf die Häufigkeiten der Rangplatzdifferenzen ≥ 35 bzw. ≤ -35 differenziert nach Ländergruppen geworfen werden. Die Daten finden sich in Tabelle 24.

Die Tabelle 24 zeigt einige auffallende Muster. Sowohl der Iran als auch Marokko weisen tendenziell vergleichsweise höhere Rangplätze in der Grammatik als in C-Test und Textwiedergabe auf. Diese Tendenz scheint stärker ausgeprägt im Fall des Iran und des C-Tests zu sein. In der Gruppe der sonstigen Länder ist die Tendenz genau umgekehrt: Die Prüflinge nehmen im Grammatikteil tendenziell niedrigere Ränge als in den übrigen Testteilen ein. Weiterhin ist bemerkenswert, daß von den insgesamt 53 Marokkanern nur 58.5% Rangplatzdifferenzen größer oder gleich $|35|$ aufweisen, während die Prozentsätze für den Iran und die Gruppe der sonstigen Länder 75% bzw. 69% betragen. Damit sind die Leistungen der Marokkaner bezogen auf alle drei Subtests vergleichsweise homogen, wobei jedoch gleichzeitig die

Tabelle 24
Münster WS 94/95
Häufigkeiten der Rangplatzdifferenzen $\geq +35$ und ≤ -35
differenziert nach Herkunftsländern

Rangdifferenz	Iran (N = 20)		Marokko (N = 53)		Sonstige (N = 68)	
	f_i	%	f_i	%	f_i	%
C-Test-GRAM $\geq +35$	–	–	2	3.8	11	16.2
C-Test-GRAM ≤ -35	5	25.0	7	13.2	5	7.4
C-Test-TW $\geq +35$	2	10.0	6	11.3	6	8.8
C-Test-TW ≤ -35	3	15.0	4	7.5	6	8.8
GRAM-TW $\geq +35$	4	20.0	7	13.2	6	8.8
GRAM-TW ≤ -35	1	5.0	5	9.4	13	19.1
Gesamt	15	75.0	31	58.5	47	69.1

C-Test-GRAM = Rangplatzdifferenz C-Test und Grammatik

C-Test-TW = Rangplatzdifferenz C-Test und Textwiedergabe

GRAM-TW = Rangplatzdifferenz Grammatik und Textwiedergabe

Leistungen im Grammatikteil weniger homogen als die des Iran sind. Umgekehrt zeigen die Iraner in jedem einzelnen Testteil die homogensten Leistungen, während ihre Leistungen bezogen auf alle drei Testteile relativ inhomogen sind.

3.2.3 Vergleichende Analyse der SAP-Ergebnisse der 5 Studienkollegs

Nachdem wir relativ ausführlich auf das Studienkolleg Münster eingegangen sind, wollen wir nun untersuchen, inwieweit die Resultate aus Münster mit den übrigen Studienkollegs übereinstimmen. Die entsprechenden Daten finden sich in Tabelle 25.

Wie die Tabelle 25 zeigt, unterscheiden sich die Studienkollegs bei den Ergebnissen im C-Test und in der Grammatik nur unwesentlich. Ein deutlicherer Unterschied besteht allerdings bei der Textwiedergabe. Hier beträgt z.B. der Unterschied zwischen Bochum und Bonn mehr als 20 Punkte. Die Variable 'Textwiedergabe' weist auch die größte Streuung auf, was wiederum als Beleg für die geringe Objektivität dieses Prüfungsteils interpretiert werden kann. Auffallend ist u.a., daß das Studienkolleg Bonn bei dieser Prüfung in allen Prüfungsteilen die niedrigsten Werte zeigt. Sieht man von der relativ kleinen Stichprobe aus Bochum ab, dann ist wie bereits in den

Tabelle 25
Studienkollegs WS 94/95
Mittelwerte (M) und Standardabweichungen (s) der SAP

Kolleg	N	C-Test		Grammatik		Textwiedergabe		SAP-Gesamt	
		M	s	M	s	M	s	M	s
Aachen	85	61.1	15.8	62.7	23.6	58.3	22.8	60.7	19.2
Bochum	41	62.5	19.1	65.5	24.9	66.5	27.0	64.8	22.5
Bonn	97	60.3	16.9	62.1	24.1	45.2	25.5	55.9	20.3
Köln	146	61.6	17.2	64.6	21.8	51.9	27.9	59.4	20.9
Münster	141	59.7	17.2	63.2	23.5	48.5	26.5	57.1	20.4
Gesamt	510	60.8	17.0	63.5	23.2	51.9	26.7	58.7	20.6

Daten von Nellessen (1992) auch im WS 94/95 die Textwiedergabe insgesamt am schwierigsten und der Grammatikteil am leichtesten.

Wir wollen nun prüfen, inwieweit sich die Unterschiede in Tabelle 25 inferenzstatistisch absichern lassen. Dazu haben wir zunächst eine multivariate Varianzanalyse (SPSS-Prozedur MANOVA) mit dem Faktor 'Kolleg' und den 3 Subtests als abhängigen Variablen gerechnet.²⁵ Für den multivariaten Test auf Homogenität der Streuungsmatrix (Box *M*) ergab sich mit $p = .004$ ein hochsignifikantes Resultat. Die univariaten Tests auf Homogenität der Varianzen (Cochrans *C* und Bartlett-Box *F*) blieben hingegen für alle drei Subtests über der Schwelle von $p = .20$. Dies bedeutet, daß die festgestellte Streuungsinhomogenität in erster Linie auf Unterschiede zwischen den Subtests beruht (und nicht auf Unterschiede zwischen den Kollegs). Die multivariaten Tests der Auswirkung des Faktors 'Kolleg' auf die Subtestleistungen ergaben für alle von SPSS berechneten Statistiken einen hochsignifikanten globalen Effekt ($p < .000$).

Tabelle 26 zeigt die Ergebnisse der entsprechenden univariaten *F*-Tests. Erwartungsgemäß ist lediglich bei der Variablen 'Textwiedergabe' der Effekt des Faktors 'Kolleg' signifikant. Allerdings ist der festgestellte Effekt relativ schwach und klärt nur 5% der Varianz auf. Vergleicht man die Mittelwerte

²⁵ Die Gültigkeit der Anwendungsvoraussetzungen der durchgeführten parametrischen (multivariaten) Varianzanalysen wurden wiederum – wie in Abschnitt 3.2.2 beschrieben – überprüft. Die festgestellten Verletzungen der Annahmen waren u.E. nicht so gravierend, daß ein Verzicht auf parametrische Verfahren notwendig erschien.

der Variablen 'Textwiedergabe' mit Hilfe des Scheffé-Tests, ergibt sich auf dem 5%-Niveau ein signifikanter Unterschied zwischen Bochum und den übrigen Kollegs mit Ausnahme von Aachen sowie zwischen Aachen und Bonn. Dies bedeutet u.a., daß Bonn verglichen mit Aachen und Bochum signifikant niedrigere Punktwerte zeigt.

Tabelle 26
Studienkollegs WS 94/95
Effekt des Faktors 'Studienkolleg':
Univariate Varianztests (N = 510)

	<i>F</i>	<i>p</i>	η^2
C-Test	0.350	.844	.003
Grammatik	0.274	.895	.002
Textwiedergabe	6.623	.000	.050
SAP-Gesamt	1.823	.123	.014

In Tabelle 27 sind die Mittelwerte und Standardabweichungen der SAP für die 11 häufigsten Herkunftsländer angegeben. Es fallen sofort die extrem hohen Werte für die Teilnehmer aus Griechenland ins Auge. Es handelt sich dabei vielfach um Lerner, die einen Teil ihrer Jugend in Deutschland verbracht haben. Mit einem SAP-Gesamtwert von 90% erzielen die Griechen Leistungen, die möglicherweise bereits im Bereich des Leistungsspektrums von Muttersprachlern liegen. Auffallend sind auch die hohen Leistungen der Teilnehmer aus Rußland und der Ukraine. Spezielle Sprachprogramme sowie Deutsch als Sprache des Elternhauses dürften hier die Gründe sein.

Die schlechtesten Leistungen weisen die Prüfling aus Brasilien auf. Danach folgen die Jordanier. Wie bereits bei den Daten aus Münster zeichnen sich die Iraner auch in der Gesamtstichprobe durch eine im Vergleich zu C-Test und Textwiedergabe hohe Leistung im Grammatikteil aus. Die Marokkaner schneiden dagegen bezogen auf den C-Test und den Grammatiktest vergleichsweise schlecht in der Textwiedergabe ab.

Ein möglicher Grund für die schlechteren Leistungen der Kandidaten aus Marokko in der Textwiedergabe ist, daß sie mit der offenen Aufgabenstellung dieses Testtyps wenig vertraut sind. Die insgesamt höheren Leistungen der Iraner sind möglicherweise darauf zurückzuführen, daß es sich hier tendenziell um eine Auswahl von Kandidaten handelt, die aufgrund

Tabelle 27

Studienkollegs WS 94/95

Mittelwerte (*M*) und Standardabweichung (*s*) der SAP bei den 11 häufigsten Herkunftsländer (alphabetische Reihenfolge)

Land	N	C-Test		GRAM		TW		SAP-Gesamt	
		M	s	M	s	M	s	M	s
Brasilien	15	53.4	16.4	49.6	26.3	44.0	28.0	49.0	22.2
Gaza	12	65.2	12.6	70.9	18.7	50.9	22.4	62.3	15.2
Griechenland	11	90.3	11.2	87.4	8.4	91.5	12.0	89.7	8.9
Indonesien	25	66.2	15.1	69.0	20.6	57.5	23.9	64.3	18.2
Iran	52	59.5	16.8	67.7	20.5	49.4	25.0	58.9	18.7
Jordanien	31	55.6	16.2	58.8	27.2	51.4	26.8	55.3	22.5
Jugoslawien	19	66.1	17.4	62.9	26.9	65.1	25.6	64.7	22.4
Marokko	156	56.4	15.2	60.8	21.7	45.3	23.3	54.2	18.2
Rußland	18	75.4	13.9	77.1	17.1	76.9	24.1	76.5	17.2
Syrien	12	55.2	20.6	62.5	33.4	45.7	26.2	54.4	25.2
Ukraine	10	75.9	10.0	80.8	12.5	71.0	23.9	75.9	14.5
übrige Länder	149	61.3	16.6	61.3	23.6	51.4	28.0	58.0	20.7
Gesamt	510	60.8	17.0	63.5	23.2	51.9	26.7	58.7	20.6

GRAM = Grammatik; TW = Textwiedergabe

guter Schulleistungen im Heimatland berechnete Chancen haben, zum Medizinstudium in Deutschland zugelassen zu werden.

In Tabelle 28 sind die Ergebnisse in der SAP differenziert nach Kurstypen aufgeführt. Beim C-Test sind die Leistungen bezogen auf die vier Kurstypen nahezu gleich. Bei den übrigen Testteilen sowie bei dem SAP-Gesamtwert liegt dagegen möglicherweise ein schwacher Effekt des Faktors 'Kurstyp' vor, wobei allerdings insbesondere im Fall der Textwiedergabe nicht auszuschließen ist, daß Objektivitätsprobleme die Ursache für die beobachteten Unterschiede sind.

Wir haben nun wiederum mit Hilfe der SPSS-Prozedur MANOVA überprüft, inwieweit sich die in Tabelle 27 und 28 festgestellten Unterschiede inferenzstatistisch absichern lassen. Für die Daten in Tabelle 27 ergab sich eine hochsignifikante multivariate und univariate Varianzhomogenität ($p < .01$). Die Varianzen in Tabelle 28 sind hingegen sowohl multivariat als auch univariat homogen.

Tabelle 28

Studienkollegs WS 94/95

Mittelwerte (*M*) und Standardabweichung (*s*) der SAP differenziert nach Kurstypen

Kurstyp	N	C-Test		GRAM		TW		SAP-Gesamt	
		M	s	M	s	M	s	M	s
S/G	80	60.1	17.1	58.2	24.9	46.1	26.7	54.8	21.3
M	145	61.3	17.0	64.7	23.0	50.4	26.8	58.8	20.6
T	193	61.4	16.4	66.0	22.3	56.3	26.2	61.2	19.7
W	92	59.5	18.3	60.7	23.4	50.3	26.8	56.8	21.1
Gesamt	510	60.8	17.0	63.5	23.2	51.9	26.7	58.7	20.6

GRAM = Grammatik; TW = Textwiedergabe

Wegen zu geringer Zellenbesetzungen haben wir darauf verzichtet, die Variablen 'Herkunftsland' und 'Kurstyp' auf eine mögliche Interaktion zu testen. Eine Inspektion der entsprechenden Mittelwerte ergab allerdings keinen entsprechenden Hinweis.

Die von SPSS durchgeführten multivariaten Tests ergaben für den Faktor 'Herkunftsland' in allen Fällen hochsignifikante Werte ($p < .000$). Auch für den Faktor 'Kurstyp' ergab sich ein hochsignifikanter multivariater Effekt ($p < .001$). Allerdings ist wegen Verletzungen der der MANOVA zugrundeliegenden Annahmen das Ergebnis insbesondere im Fall des Faktors 'Herkunftsland' nur mit Einschränkungen aussagekräftig.

Die Tabelle 29 zeigt die Resultate des univariaten *F*-Tests für den Effekt des Faktors 'Herkunftsland'. Die entsprechenden Daten für den Faktor 'Kurstyp' finden sich in Tabelle 30.

Aus Tabelle 29 ist ersichtlich, daß der Effekt der Variablen 'Herkunftsland' offensichtlich beim C-Test am stärksten ist, dicht gefolgt von der Textwiedergabe. Ein deutlich geringerer Effekt scheint im Fall des Grammatiktests vorzuliegen.

Für den Faktor 'Kurstyp' ist der Effekt im Fall von Grammatik und Textwiedergabe zwar signifikant, jedoch äußerst gering.

Testet man die Mittelwertunterschiede in Tabelle 27 mit Hilfe des konservativen Scheffé-Verfahrens auf Signifikanz, ergeben sich auf der 5%-Ebene für die Variable 'Grammatik' keine signifikanten Unterschiede. Beim C-Test, bei der Textwiedergabe sowie auch bei der gesamten SAP zeichnen

Tabelle 29
Studienkollegs WS 94/95
Effekt des Faktors 'Herkunftsland':
Univariate Varianstests (N = 510)

	<i>F</i>	<i>p</i>	η^2
C-Test	8.079	.000	.151
Grammatik	3.618	.000	.074
Textwiedergabe	6.347	.000	.123
SAP-Gesamt	6.238	.000	.121

Tabelle 30
Studienkollegs WS 94/95
Effekt des Faktors 'Kurstyp': Univariate Varianstests (N = 510)

	<i>F</i>	<i>p</i>	η^2
C-Test	0.345	.793	.002
Grammatik	2.734	.043	.016
Textwiedergabe	3.286	.021	.019
SAP-Gesamt	2.200	.087	.013

sich dagegen insbesondere Griechenland, aber auch Rußland durch signifikant höhere Leistungen aus. Beim C-Test haben 6 Länder, bei der Textwiedergabe und der gesamten SAP jeweils 5 Länder, einen signifikant niedrigeren Mittelwert als Griechenland. Rußland zeigt dagegen bei den genannten Variablen lediglich einen signifikant höheren Mittelwert als Marokko.

In Tabelle 31 sind die Korrelationen der diskriminierenden Variablen 'C-Test', 'Grammatik' und 'Textwiedergabe' mit den kanonischen Diskriminanzfunktionen aufgeführt. Es wurden jeweils drei Diskriminanzfunktionen extrahiert, die im Fall der Variablen 'Herkunftsland' 75%, 18% und 7% Varianz aufklären und im Fall der Variablen 'Kurstyp' 84%, 15% und 1%. In Tabelle 31 sind lediglich die von SPSS als signifikant gewerteten Funktionen angegeben.

Bei der Variablen 'Herkunftsland' wird die 1. Diskriminanzfunktion vor allem durch C-Test und Textwiedergabe konstituiert. Die zweite, allerdings deutlich schwächere Funktion korreliert dagegen am höchsten mit der Variablen 'Grammatik'. Dies weist auf eine gewisse Sonderstellung der Grammatikkenntnisse bei der Unterscheidung zwischen den Herkunftsländern hin.

Tabelle 31
Studienkollegs WS 94/95
Korrelationen zwischen diskriminierenden Variablen und
kanonischen Diskriminanzfunktionen (N = 510)

diskriminierende Variable	Herkunftsland		Kurstyp
	Funktion 1	Funktion 2	Funktion 1
C-Test	.85	-.52	.12
Grammatik	.43	-.81	.49
Textwiedergabe	.76	-.18	.62

In bezug auf die Variable 'Kurstyp' bestätigt die Diskriminanzanalyse dagegen lediglich die Ergebnisse der univariaten *F*-Tests und bringt damit keine weiteren Erkenntnisse.

Um festzustellen, ob es Unterschiede zwischen den Kollegs hinsichtlich der länderspezifischen Diskriminierungsfähigkeit der drei Subtests der SAP gibt, haben wir zusätzlich die Diskriminanzfunktionen für jedes einzelne Kolleg ermittelt. Dabei wurden allerdings lediglich Herkunftsländer mit einer Häufigkeit von ≥ 10 als eigene Gruppe betrachtet. Auch diese Analysen wiesen auf eine gewisse Sonderstellung der Grammatikkenntnisse bei der Diskrimination zwischen den Herkunftsländern hin.

Tabelle 32 zeigt die Interkorrelationen (Pearson-Korrelationen) zwischen den drei Teilen der SAP getrennt nach Studienkollegs. Alle Korrelationen sind signifikant mit einer Irrtumswahrscheinlichkeit von 0.1% und zudem sehr hoch. Sieht man von den wegen des geringen Stichprobenumfangs weniger verlässlichen Daten aus Bochum ab, dann zeigt der C-Test eine etwa gleich starke Beziehung zu Grammatik und Textwiedergabe. Etwas schwächer ausgeprägt scheint dagegen die Beziehung zwischen Grammatik und Textwiedergabe zu sein.

In Tabelle 33 sind wiederum die Pearson-Korrelationen zwischen den Subtests der SAP getrennt für eine obere und eine untere Leistungsgruppe aufgeführt. Die Leistungsgruppen wurden durch Teilung am Median des C-Tests (*Med* = 62) – dem vermutlich zuverlässigsten Testteil – gebildet. Für alle Korrelationen gilt $p < .001$. Insgesamt gesehen scheint der Zusammenhang zwischen den Subtests in der oberen Leistungsgruppe schwächer zu sein. Allerdings ist wiederum zu berücksichtigen, daß die Streuungen der Variablen in der oberen Gruppe reduziert sind (vgl. den Kommentar zu

Tabelle 32
Studienkollegs WS 94/95
Interkorrelationen der Subtests*

Kolleg	Subtest	Grammatik	Textwiedergabe
Aachen (N = 85)	C-Test	.88	.76
	Grammatik		.74
Bochum (N = 41)	C-Test	.83	.89
	Grammatik		.84
Bonn (N = 97)	C-Test	.82	.73
	Grammatik		.73
Köln (N = 146)	C-Test	.80	.84
	Grammatik		.79
Münster (N = 141)	C-Test	.75	.81
	Grammatik		.67
Gesamt (N = 510)	C-Test	.80	.79
	Grammatik		.73

*Pearson-Produkt-Moment-Korrelationen; $p < .001$ in allen Fällen

Tabelle 17). Die Standardabweichungen betragen dort für C-Test, Grammatik und Textwiedergabe lediglich 8.36, 11.85 und 20.26. Für die untere Gruppe gilt hingegen: 11.09, 21.85 und 20.90. Hervorzuheben sind insbesondere die deutlich homogeneren Resultate der oberen Leistungsgruppe im Grammatiktest. Die deutlich niedrigeren Korrelationen mit der Variablen 'Grammatik' in der oberen Leistungsgruppe dürften zumindest teilweise hierauf zurückzuführen sein.

Tabelle 33
Studienkollegs WS 94/95
Interkorrelationen der Subtests getrennt für die obere und untere Leistungsgruppe (geteilt am Median des C-Tests)*

	C-Test > 62 (N = 256)		C-Test ≤ 62 (N = 254)	
	Grammatik	Textwiedergabe	Grammatik	Textwiedergabe
C-Test	.47	.56	.69	.66
Grammatik		.37		.64

*Pearson-Produkt-Moment-Korrelationen; $p < .001$ in allen Fällen

Tabelle 34 zeigt die Pearson-Korrelationen zwischen den Subtests der SAP getrennt für die vier häufigsten Herkunftsländer sowie die übrigen Länder. Es gilt wiederum für alle Korrelationen $p < .001$. Es fällt u.a. auf, daß bei den Teilnehmern aus Jordanien möglicherweise ein etwas stärkerer Zusammenhang zwischen den von den drei Subtests gemessenen Leistungskomponenten besteht als bei den Studierenden aus dem Iran und aus Marokko. Da über mögliche Ursachen wegen unzureichender Informationen nur spekuliert werden kann,²⁶ verzichten wir auf eine weitere Analyse und Kommentierung der Daten in Tabelle 34.

Tabelle 34
Studienkollegs WS 94/95
Interkorrelationen der Subtests getrennt für fünf Kategorien von Herkunftsländern*

Land	Subtest	Grammatik	Textwiedergabe
Indonesien (N = 25)	C-Test	.90	.67
	Grammatik		.73
Iran (N = 52)	C-Test	.77	.74
	Grammatik		.67
Jordanien (N = 31)	C-Test	.91	.88
	Grammatik		.86
Marokko (N = 146)	C-Test	.76	.75
	Grammatik		.71
übrige Länder (N = 246)	C-Test	.83	.81
	Grammatik		.75

*Pearson-Produkt-Moment-Korrelationen; $p < .001$ in allen Fällen

²⁶ So kann z.B. ein fertigkeitssausbalancierter Deutschunterricht im Herkunftsländ die Ursache für eine hohe Interkorrelation zwischen den drei Subtests sein.

Abschließend möchten wir in Fortführung der Analysen von Nellessen (1992) (vgl. Abschnitt 3.1, Tabelle 6) noch die Pearson-Korrelationen zwischen den Ergebnissen in der SAP (Testtermin August 1994) und der Semesternote WS 94/95 angeben. Die Semesternote wurde im Januar 1995 vergeben und beruht wiederum auf den Semesterklausuren und der Mitarbeit der Studierenden. Die Korrelationen finden sich zusammen mit den jeweiligen Stichprobenumfängen und Standardabweichungen in Tabelle 35. Zusätzlich haben wir dieses Mal auch die mit Hilfe der Fisherschen z -Transformation gemittelten Korrelationen \bar{r} für die einzelnen Testteile angegeben. Da wegen Nichtbestehens der SAP oder auch aus anderen Gründen relativ viele Probanden kein Studium an dem jeweiligen Kolleg aufgenommen haben, sind die Stichprobenumfänge im Vergleich zu den vorangehenden Analysen deutlich reduziert.

Tabelle 35
Studienkollegs WS 94/95
Korrelationen* zwischen SAP und Semesternote;
SAP-Standardabweichungen in Klammern

	Semesternote					\bar{r} ($N=220$)
	Aachen ($N=32$)	Bochum ($N=27$)	Bonn ($N=41$)	Köln ($N=56$)	Münster ($N=64$)	
C-Test	.54 (9.5)	.77 (11.3)	.43 (8.5)	.59 (9.5)	.57 (11.8)	.57
Grammatik	.47 (9.7)	.63 (12.0)	.59 (10.9)	.33 (10.4)	.35 (11.6)	.52
Textwiedergabe	.53 (15.7)	.66 (13.8)	.47 (17.8)	.62 (15.5)	.69 (20.4)	.61
gesamte SAP	.64 (9.0)	.82 (10.3)	.73 (8.4)	.66 (9.4)	.73 (11.2)	.71

*Pearson-Produkt-Moment-Korrelationen; $p < .01$ in allen Fällen (einseitig)

Aus Tabelle 35 ist ersichtlich, daß im Gegensatz zu den Befunden von Nellessen (1992) der Grammatiktest diesmal am geringsten mit der Semesternote korreliert. Etwas stärker scheint die Beziehung im Fall von C-Test und Textwiedergabe zu sein. Erwartungsgemäß ist die Korrelation zur Semesternote im Fall der gesamten SAP am höchsten.

Wegen der hohen Interkorrelation der Subtests der SAP sind allerdings die angeführten bivariaten Korrelationen nur sehr bedingt als Maßzahl für die diskriminative Kraft des jeweiligen Testteils in bezug auf die Semesternote zu interpretieren. Wir haben deshalb zusätzlich für jedes Kolleg eine schrittweise multiple Regression mit den drei Subtests als Regressoren gerechnet. Tabelle 36 zeigt die Werte des multiplen Korrelationskoeffizienten R für die einzelnen Schritte der schrittweisen Regression jeweils vom Programm ausgewählte Prädiktorkombination.

Tabelle 36
Studienkollegs WS 94/95
Schrittweise Regression der Semesternote auf die Subtests
der SAP (inkrementelles multiples R)

	Semesternote				
	Aachen ($N = 32$)	Bochum ($N = 27$)	Bonn ($N = 41$)	Köln ($N = 56$)	Münster ($N = 64$)
C-Test	.54	.59	—	.67	—
Grammatik	—	.69	.59	—	.73
Textwiedergabe	.67	—	.75	.62	.69

In Tabelle 36 entspricht bei jedem Subtests der niedrigere Wert der einfachen Korrelation aus Tabelle 35. Dieser Wert signalisiert, daß der jeweilige Subtest als erster in die Regressionsgleichung aufgenommen worden ist. Der größere Wert bezeichnet dagegen die multiple Korrelation bei Hinzunahme des entsprechenden Prädiktors. Ein Strich signalisiert, daß ein Testteil keinen signifikanten weiteren Beitrag zu Erhöhung der multiplen Korrelation zu leisten vermag ($p < .05$). So bedeutet z.B. im Fall von Aachen die für den Subtest 'Textwiedergabe' angeführte Korrelation von .67, daß sich die Korrelation von .54 zwischen Semesternote und C-Test als einzigem Prädiktor auf .67 erhöht, wenn die Textwiedergabe als weiterer Prädiktor hinzugenommen wird. Eine Hinzunahme des Testteils 'Grammatik' als dritter Prädiktor bringt dagegen keine weitere Verbesserung der Vorhersage.

4. Zusammenfassung und praktische Empfehlungen

Wie oben schon beschrieben handelt es sich bei der SAP um eine Kombination von Einzeltests, durch die ein komplexes Persönlichkeitsmerkmal, nämlich die sprachliche Eignung zur Teilnahme am studienvorbereitenden Sprach- und Fachunterricht an Studienkollegs für ausländische Studierende, überprüft werden soll. Dazu wird eine Testbatterie bestehend aus drei Subtests eingesetzt, zu der seit einigen Jahren zumeist auch ein C-Test gehört. Im folgenden sollen nun die wichtigsten Ergebnisse der vorangegangenen statistischen Analysen zusammengefaßt und darauf aufbauend Überlegungen dazu angestellt werden,

- ob der C-Test eine für die SAP geeignete Testform ist;
- inwieweit die Testteile 'Grammatik' und 'Textwiedergabe' den für objektive Tests geforderten Gütekriterien entsprechen;
- ob eine Kombination der Testteile 'C-Test', 'Grammatik' und 'Textwiedergabe' sinnvoll ist.

Bei der Beurteilung der Reichweite der Aussagen ist zu beachten, daß sich die statistischen Analysen primär auf ein Prüfungsbeispiel des Studienkollegs Münster stützen und daß die anderen Studienkollegs trotz eines gemeinsam akzeptierten Rahmens Prüfungen erstellen, die in Inhalt und Form eigene Akzente setzen.

4.1 C-Test

Der aus acht Texten bestehende C-Test erweist sich in der Untersuchung erwartungsgemäß als hoch reliabel (Cronbachs Alpha = .95). Es konnte nachgewiesen werden, daß selbst bei einer Verkürzung auf vier Texte noch eine Reliabilität von .93 erreicht wird. Es empfiehlt sich deshalb aus Gründen der Testökonomie, einen C-Test mit nur fünf Texten einzusetzen, wobei der erste Text lediglich eine 'Eisbrecherfunktion' hat und nicht ausgewertet wird.

Die fast gleichen C-Test-Mittelwerte der einzelnen Kollegs indizieren zudem eine hohe Durchführung- und Auswertungsobjektivität. Weiterhin konnte durch Vorerprobung an anderen Institutionen mit durchschnittlich .61 eine angemessene Schwierigkeit erreicht werden (die Bestehensgrenze der SAP liegt bei 60%).

Allerdings erwies sich die vorliegende C-Test-Version trotz Vorerprobung als nicht eindimensional im Sinne des KLA-Modells, d.h., nicht alle Texte messen das gleiche Merkmal. Ohne den Ausschluß bestimmter Texte ist damit eine Addition der Punktwerte der Einzeltexte nur mit Einschränkungen vertretbar.

Auffallend ist, daß bei den beiden letzten Texten des untersuchten C-Tests eine Reihe von Prüfungskandidaten 0 Punkte erzielt haben. Es ist zu vermuten, daß die entsprechenden Probanden die genannten Texte aus Zeitgründen gar nicht mehr bearbeitet haben. Um diese Situation in Zukunft zu vermeiden, sollte man entweder den Studenten während der Prüfung Hinweise auf die Zeiteinteilung geben oder den Gesamttest bei gleicher Bearbeitungszeit um zwei Texte kürzen.

4.2 Grammatik

Die Grammatik erweist sich in bezug auf das Kriterium der Objektivität erwartungsgemäß als unproblematisch. Im Gegensatz zur Textwiedergabe unterscheiden sich die Mittelwerte der einzelnen Kollegs nur minimal. Auffällig ist, daß die Grammatikwerte in der oberen Leistungsgruppe weit homogener als in der unteren Leistungsgruppe sind (vgl. den Kommentar zu Tabelle 17). Dies bedeutet u.a., daß der Grammatiktest im oberen Leistungsbereich nicht mehr genügend differenziert. Dem könnte u.a. durch ein breiteres Spektrum von hinreichend schwierigen Aufgaben entgegengewirkt werden.

4.3 Textwiedergabe

In Abschnitt 2.2.5 wurde bereits die Vermutung geäußert, daß es sich bei der Textwiedergabe aufgrund der offenen Aufgabenstellung, der Unterschiede in der Testdurchführung sowie des großen Bewertungsspielraums um ein wenig objektives – und damit auch wenig reliables – Testinstrument handelt. Diese Vermutung wurde durch die vorangehenden statistischen Analysen erhärtet. So weist bei allen Kollegs jeweils die Textwiedergabe verglichen mit Grammatik und C-Test die höchste Streuung auf. Weiterhin sind bei diesem Testteil die Unterschiede zwischen den Kollegs am größten. So beträgt die Differenz zwischen den Mittelwerten des Bonner und Bochumer Kollegs mehr als 20 Prozentpunkte. Die angestrebte Vergleichbarkeit der Testdurchführung und -auswertung an den einzelnen Studienkollegs wird also bei diesem Prüfungsteil nicht erreicht.

Hält man an der Textwiedergabe als mögliche Testform fest, sollte zur Erhöhung der Durchführungsobjektivität der Text bei allen Kollegs mittels eines vorher zentral erstellten Tondokuments dargeboten werden. Außerdem kann man zur Vergrößerung der Auswertungsobjektivität die von den Studierenden produzierten Texte von zwei speziell instruierten Lehrern unabhängig voneinander beurteilen lassen und auf dieser Basis dann ein abschließendes Urteil fällen. Die Beurteilung könnte entweder aus Gründen der Testökonomie global anhand einer einzigen Skala oder mit Hilfe mehrerer, bestimmte Teilfertigkeiten und inhaltliche Komponenten berücksichtigender Einzelskalen vorgenommen werden. Insbesondere im letztgenannten Fall ist es nötig, die zu überprüfenden Einzelaspekte unter Rückgriff auf allgemeine Testspezifikationen detailliert zu beschreiben.

Eine weitere Möglichkeit, die Objektivität zu erhöhen, ist die Trennung des Subtests 'Textwiedergabe' in einen Hörverstehens- und einen Textproduktionsteil sowie die stärkere Vorstrukturierung der von den Probanden geforderten Leistung durch geschlossene oder halboffene Aufgaben. So kann z.B. die Fertigkeit 'Hörverstehen' durch ein Lückendiktat, die Fertigkeit 'Textproduktion' wie bereits in anderen Prüfungsdurchgängen durch die sprachliche Verbindung vorgegebener Stichwörter überprüft werden. Es ist in diesem Zusammenhang darauf hinzuweisen, daß darauf geachtet werden sollte, daß sich die eingesetzten Tests im Aufgabenformat hinreichend unterscheiden.

Einen weitergehenden Vorschlag zur Erhöhung der Objektivität und Validität von Tests findet sich bei Klein-Brale (1992). Sie empfiehlt – allerdings mit Blick auf die PNdS (Prüfung zum Nachweis deutscher Sprachkenntnisse) – die Erstellung einer Testentwicklungsmatrix, in der die vier Fertigkeitsbereiche Hören, Sprechen, Lesen und Schreiben genauer spezifiziert, Regeln zur Entwicklung von Aufgaben für die definierten Bereiche angegeben und die von den Testpersonen erwarteten Reaktionen detailliert beschrieben werden. Dieses Verfahren würde dazu beitragen, die Prüfungsentwürfe der einzelnen Kollegs hinsichtlich Aufgabenqualität und -schwierigkeit vergleichbarer zu machen sowie die diagnostische Aussagekraft der Testresultate zu erhöhen.

4.4 Zusammenhang zwischen C-Test, Grammatik und Textwiedergabe

Die Untersuchungen haben gezeigt, daß die einzelnen Testteile hoch miteinander korrelieren. So zeigt der C-Test eine etwa gleich starke Beziehung zu Grammatik und Textwiedergabe. Etwas schwächer scheint dagegen der Zusammenhang zwischen Grammatik und Textwiedergabe zu sein. Insbesondere in der oberen Leistungsgruppe ist die Korrelation zwischen Grammatik und den übrigen Testteilen deutlich niedriger als in der unteren Leistungsgruppe. Allerdings ist in der oberen Leistungsgruppe wegen der reduzierten Streuung der Grammatikwerte der durchschnittliche absolute Fehler bei der Vorhersage der Grammatikwerte durch die Textwiedergabe bzw. durch den C-Test z.B. bei den Daten aus Münster deutlich geringer als in der unteren Leistungsgruppe (vgl. den Kommentar zu Tabelle 17 und 33).

Die recht hohe Korrelation zwischen den Subtests 'Textwiedergabe' und 'C-Test' dürfte, wie schon in Abschnitt 3.1 angedeutet, darauf zurückzuführen sein, daß beide Testteile sowohl Verstehen als auch produktive Fertigkeiten messen. Bei der Auswahl der Subtests für künftige Prüfungsdurchgänge ist deshalb darauf zu achten, daß sich die einzelnen Testteile sowohl hinsichtlich der Testmethode als auch in den Testinhalten deutlich unterscheiden.

Wir möchten an dieser Stelle nochmals betonen, daß die hohe Interkorrelation zwischen den Testteilen keinesfalls bedeutet, daß die Testteile untereinander austauschbar sind. Wie in den Tabellen 15 und 16 gezeigt wurde, gibt es trotz der hohen Interkorrelationen eine Vielzahl von Probanden, die in den untersuchten Subtests stark unterschiedliche Rangplätze einnehmen. So weisen bei den Münsteraner Daten trotz der Korrelation von .81 zwischen C-Test und Textwiedergabe insgesamt 36 von 141 Prüflingen Rangdifferenzen von mehr als 30 Plätzen auf. Die extremsten Rangdifferenzen finden sich dort mit 94.5 Plätzen zwischen Grammatik und Textwiedergabe, mit 92 Plätzen zwischen C-Test und Grammatik sowie mit 89 Plätzen zwischen C-Test und Textwiedergabe (bei einem maximalen Rangplatz von 141). Die einzelnen Subtests der SAP können somit zu höchst unterschiedlichen Selektions- und Einstufungsentscheidungen führen.

4.5 Einfluß der Faktoren 'Kurstyp' und 'Herkunftsland'

Die insgesamt 510 Prüflinge stammen aus 64 Ländern, wobei Marokko mit 31% das weitaus größte Kontingent stellt. Es folgen der Iran mit 10% und Jordanien mit 6% (vgl. Tabelle 7).

Die Prüflinge verteilen sich relativ ungleich auf die vier Kurstypen. 193 Prüflinge beabsichtigen, an einem T-Kurs teilzunehmen, 145 an einem M-Kurs, 92 an einem W-Kurs und 80 an einem S/G-Kurs. Dabei gibt es allerdings deutliche kolleg- und länderspezifische Schwerpunkte. So bietet Bochum lediglich T-Kurse an und Aachen nur T- und W-Kurse. Weiterhin bevorzugen z.B. die Teilnehmer aus Marokko eindeutig T-Kurse, die Teilnehmer aus dem Iran dagegen M-Kurse (vgl. Tabelle 8 und 9).

Beim C-Test ergeben sich in der Gesamtstichprobe aller Kollegs für die vier Kurstypen nahezu gleiche Mittelwerte. Bei den beiden übrigen Testteilen liegt dagegen vermutlich ein schwacher Effekt des Faktors 'Kurstyp' vor (vgl. Tabelle 28).

Deutlich stärker ausgeprägt ist der Einfluß des Faktors 'Herkunftsland'. Auf die extrem hohen Leistungen der Teilnehmer aus Griechenland sowie die hohen Leistungen der Prüflinge aus Rußland und der Ukraine wurde bereits hingewiesen (vgl. den Kommentar zu Tabelle 27).

In bezug auf die Frage, welche inhaltlichen Bereiche abgetestet und wie die Inhalte zu einem Gesamttest kombiniert werden sollen, sind u.a. die differentiellen Leistungen der Hauptteilnehmergruppen aus Marokko und dem Iran in den drei Teilen der SAP von Interesse. So zeigen die Marokkaner sowohl im C-Test als auch in der Textwiedergabe deutlich schlechtere Leistungen als der Durchschnitt der Restländer, wobei die Marokkaner vor allem in der Textwiedergabe vergleichsweise schlecht abschneiden. Dagegen weisen die Iraner sowohl in der Textwiedergabe als auch im C-Test durchschnittliche Leistungen auf. Beim Grammatikteil liegen die Werte jedoch deutlich sowohl über dem Durchschnitt der Restländer als auch über den Leistungen der Marokkaner.²⁷ Zudem zeichnen sich die Iraner durch eine hohe Homogenität der Leistungen insbesondere im Grammatik- und Textwiedergabeteil aus (vgl. Tabelle 19, 20, 21).

²⁷ Vgl. auch Heilmann (1991) der auf der Basis seiner Lehrerfahrungen am Studienkolleg Münster zu der Feststellung kommt, "daß formale Grammatikkenntnisse ... von der aktuellen Sprachfertigkeit des Kandidaten in Einzelfällen extrem abweichen." (S. 3)

30. Insgesamt weisen die Analysen darauf hin, daß der Effekt des Faktors 'Herkunftsland' beim C-Test am stärksten ist, dicht gefolgt von der Textwiedergabe. Dabei verhalten sich diese beiden Testteile in ihrer länderspezifischen Differenzierungsfähigkeit offensichtlich ähnlich. Ein qualitativ ähnlicher und zudem deutlich geringerer Effekt scheint dagegen im Fall des Grammatiktests vorzuliegen.

C-Test und Textwiedergabe unterscheiden sich vom Grammatikteil insbesondere dadurch, daß sie weitaus stärker auch produktive lexikalisch-grammatische Fertigkeiten erfassen. Die Ergebnisse können deshalb dahingehend interpretiert werden, daß sich die Teilnehmer aus unterschiedlichen Herkunftsländern weniger in ihren grammatischen Kenntnissen, sondern eher in ihren produktiven lexikalisch-grammatischen Fertigkeiten unterscheiden. Diese Interpretation schließt natürlich die Möglichkeit eines länderspezifischen Bias z.B. beim C-Test aufgrund einer unzureichenden Vertrautheit mit diesem Testformat oder auch aus anderen Gründen nicht aus.

4.6 Kompensatorische Beziehung zwischen den Prüfungsteilen

Für das Bestehen der SAP ist ein Gesamtpunktwert von 60% die untere Grenze. Dies impliziert, daß sich gute und schlechte Leistungen in den einzelnen Prüfungsteilen – u.a. als Folge eines in bezug auf die verschiedenen Fertigkeiten nicht ausbalancierten Fremdsprachenunterrichts – gegenseitig kompensieren können und die entsprechenden Teilnehmer trotz schlechter Leistungen in bestimmten Subtests die SAP bestehen. Ein Beispiel ist folgender erfolgreicher Prüfling: 87% in der Grammatik, 64% im C-Test, aber nur 30% in der Textwiedergabe.

Wie der Tabelle 34 zu entnehmen ist, gibt es möglicherweise länderspezifische Unterschiede in der Stärke des Zusammenhangs zwischen den drei Testteilen. So verfügen z.B. die Iraner, wie oben gezeigt, über gute Grammatikkenntnisse, wobei diese jedoch nicht immer mit guten Leistungen in den übrigen Subtests einhergehen. Als Folge kommt es bei dieser Gruppe in einer Reihe von Fällen zu einem Bestehen der SAP in erster Linie aufgrund einer hohen Leistung im Grammatikteil.

Dies ist u.E. jedoch eine problematische Konsequenz aus der globalen 60%-Grenze. Wie bereits festgestellt, soll ein Bestehen der SAP gewährleisten, daß ein Lerner über die zur Teilnahme am studienvorbereitenden

Sprach- und Fachunterricht notwendigen sprachlichen Voraussetzungen verfügt. Zu den unabdingbaren Voraussetzungen gehören jedoch die durch den Grammatiktest nur unzureichend erfaßten produktiven Fertigkeiten sowie das durch den Grammatiktest überhaupt nicht erfaßte Hörverständnis. Angesichts dieses Problems ist folgende Regelung denkbar: Die Probanden müssen nicht nur im Gesamtttest 60%, sondern zusätzlich in jedem einzelnen Teil der SAP 50% erreichen.

4.7 Vorhersage der Semesternote

Es war gezeigt worden, daß im WS 1994/95 der Grammatikteil am geringsten mit der Semesternote korreliert (durchschnittlich .47). Dies gilt für alle Kollegs mit Ausnahme von Bonn. Etwas stärker ausgeprägt scheint dagegen die Beziehung zur Semesternote im Fall von C-Test und Textwiedergabe zu sein (durchschnittlich .57 bzw. .61). Die Korrelation ist mit durchschnittlich .71 im Fall der gesamten SAP am höchsten (vgl. Tabelle 35).

Wie allerdings die Ergebnisse der für jedes Kolleg einzeln durchgeführten Regressionsanalysen zeigen (vgl. Tabelle 36), tragen zur Vorhersage der Semesternote stets nur zwei der drei Subtests substantiell bei, wobei der wichtigste Prädiktor mit Ausnahme von Bonn stets die Textwiedergabe oder der C-Test ist. Dies bedeutet, daß in bezug auf die Vorhersage der Semesternote der Grammatiktest vergleichsweise redundant ist.

Es gibt jedoch nicht unerhebliche Unterschiede zwischen den Kollegs. So ist in Aachen der C-Test der wichtigste Prädiktor, gefolgt von der Textwiedergabe, während in Münster die Textwiedergabe der wichtigste Prädiktor ist, gefolgt von der Grammatik. Über Gründe läßt sich nur mutmaßen: Zum einen dürften Unterschiede in der inhaltlichen Ausrichtung des Sprachunterrichts der einzelnen Kollegs die Ursache sein. Da sich die Kollegs in der länderspezifischen Verteilung ihrer Kursteilnehmer unterscheiden, ist weiterhin denkbar, daß auch die Variable 'Herkunftsland' eine Rolle spielt. Schließlich sind auch die festgestellten Mängel in der Durchführungs- und Auswertungsobjektivität des Testteils 'Textwiedergabe' als mögliche Ursache in Betracht zu ziehen.

5. Der C-Test in der SAP: abschließende Bewertung

Angesichts der Analyseergebnisse und der Zielsetzung der SAP ist u.E. der Einsatz des C-Tests ohne Einschränkungen zu empfehlen. Insbesondere im Vergleich zur Textwiedergabe zeichnet sich der C-Test durch eine hohe Durchführungs- und Auswertungsobjektivität bei einer gleichzeitigen hohen Auswertungsökonomie aus.

Unterscheidet man mit Cummins (1984; 1991) zwischen *Cognitive/Academic Language Proficiency* (CALP) und *Basic Interpersonal Communicative Skills* (BICS), dann mißt der C-Test vermutlich in starkem Maße die rezeptive Seite der *Cognitive/Academic Language Proficiency* (vgl. auch Daller, 1996 sowie die Einwände in Singleton & Singleton, 1996; vgl. ferner die kritische Diskussion der Dichotomie von CALP und BICS bei Baker, 1993, S. 138-146). Dies bedeutet, daß der C-Test bei der vorliegenden Population primär die Fähigkeit erfaßt, kontextreduzierte Sprache in kognitiv anspruchsvollen Aufgaben im Zuge des Verstehensprozesses zu verarbeiten. Diese Fähigkeit ist ein wichtiger Bestandteil der für die erfolgreiche Teilnahme an Kursen der Studienkollegs notwendigen sprachlichen Voraussetzungen. Der C-Test erweist sich damit nicht nur als hochreliables, hochobjektives und sehr ökonomisches Testinstrument, sondern u.E. auch als hinreichend valide in bezug auf die Zielsetzungen der Studienkollegs.

Literaturverzeichnis

- Baker, Colin. (1993). *Foundations of bilingual education and bilingualism*. Clevedon, England: Multilingual Matters.
- Baur, Rupprecht S. & Meder, Gregor. (1994). C-Tests zur Ermittlung der globalen Sprachfähigkeit im Deutschen und in der Muttersprache bei ausländischen Schülern in der Bundesrepublik Deutschland. In Rüdiger Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 2, S. 151-178). Bochum: Brockmeyer.
- Bolten, Jürgen. (1992). Wie schwierig ist ein C-Test? Erfahrungen mit dem C-Test als Einstufungstest in Hochschulkursen Deutsch als Fremdsprache. In Rüdiger Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 1, S. 193-203). Bochum: Brockmeyer.
- Coleman, James A. (1994). Profiling the advanced language learner: The C-Test in British further and higher education. In Rüdiger Grotjahn

- (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 2, S. 217-237). Bochum: Brockmeyer.
- Coleman, James A. (1996). A comparative survey of the proficiency and progress of language learners in British universities. In Rüdiger Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 3). Bochum: Brockmeyer.
- Coleman, James A., Grotjahn, Rüdiger, Klein-Braley, Christine & Raatz, Ulrich. (1994). The European Language Proficiency Survey: A comparative investigation of foreign language learners in schools and universities in several European countries. *Language Testing Update*, 14, 49-55.
- Cummins, Jim. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. In Charlene Rivera (Hrsg.), *Language proficiency and academic achievement* (S. 2-19). Clevedon, England: Multilingual Matters.
- Cummins, Jim. (1991). Conversational and academic language proficiency in bilingual contexts. *AILA Review*, 19, 75-89.
- Daller, Helmut. (1996). Der C-Test als Meßinstrument alltagsprachlicher und akademischer Sprachfähigkeiten türkischer Remigranten. In Rüdiger Grotjahn (Hrsg.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen* (Bd. 3). Bochum: Brockmeyer.
- Deutscher Akademischer Austauschdienst (DAAD). (Hrsg.). (1993). *Der Zugang zum Fachstudium an den wissenschaftlichen Hochschulen in der Bundesrepublik Deutschland über die Studienkollegs für ausländische Studierende*. Bonn: Köllen.
- Grönwoldt, Peter. (1985). Erfahrungen mit dem C-Test am Eurozentrum Köln. In Christine Klein-Braley & Ulrich Raatz (Hrsg.), *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis* (S. 91-95). Bochum: AKS.
- Grotjahn, Rüdiger. (1992). Der C-Test im Französischen. Quantitative Analysen. In Rüdiger Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 1, S. 205-255). Bochum: Brockmeyer.
- Grotjahn, Rüdiger, Krause, Jürgen & von Unwerth, Heinz-Jürgen. (1987). The Bochum Diagnostic Test for English. In Rüdiger Grotjahn, Christine Klein-Braley & Douglas K. Stevenson (Hrsg.), *Taking their measure: The validity and validation of language tests* (S. 133-156). Bochum: Brockmeyer.
- Guttman, Louis. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255-282.
- Heilmann, Erhard G. (1991). *Eine Neuerung bei der Aufnahmeprüfung ins Studienkolleg: C-Test* (Ms.).
- Holm, Sture. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65-70.
- Jacobs, Ruth & Arndt, Victoria. (1985). C-Tests für ausländische Schüler: Auswahl zu Weiterbildungsmaßnahmen. In Christine Klein-Braley & Ulrich Raatz (Hrsg.), *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis* (S. 83-90). Bochum: AKS.
- Jakschik, Gerhard. (1992). Zum Einsatz des C-Tests in den Psychologischen Diensten der Arbeitsämter. Ein C-Test für Deutsch als Zweitsprache. In Rüdiger Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 1, S. 297-311). Bochum: Brockmeyer.
- Jakschik, Gerhard. (1994). Der C-Test für erwachsene Zweitsprachler als Einstufungsinstrument bei der Schulausbildung. In Rüdiger Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 2, S. 259-278). Bochum: Brockmeyer.
- Jakschik, Gerhard. (1996). Validierung des C-Tests für erwachsene Zweitsprachler. Eine Längsschnittuntersuchung bei Trägern von beruflichen Bildungsmaßnahmen. In Rüdiger Grotjahn (Hrsg.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 3). Bochum: Brockmeyer.
- Klein-Braley, Christine. (1992). Objektives Erfassen von Hör- und Leseverstehen. Einige Erkenntnisse aus der Theorie der Leistungsmessung und deren Bezug zur Prüfung zum Nachweis deutscher Sprachkenntnisse. *Info DaF*, 19, 649-663.
- Kultusministerkonferenz. (1983). *Empfehlung über die Prüfung zum Nachweis der Deutschkenntnisse bei der Aufnahme in ein Studienkolleg - Beschluß der Kultusministerkonferenz vom 21.10.1983*. abgedruckt in: Jahrbuch Deutsch als Fremdsprache, 10 (1984), 254-257.
- Kultusminister NW. (1989). Verordnung über die Ausbildung und Prüfung für ausländische Studierende der Studienkollegs an wissenschaftlichen

- Hochschulen. (1989). *Gesetz- und Verordnungsblatt für das Land Nordrhein-Westfalen*, 49(47), 517-528.
- Moosbrugger, Helfried & Müller, Hans. (1981). Ein klassisches latent-additives Testmodell (KLA). In Wolfgang Michaelis (Hrsg.), *Bericht über den 32. Kongreß der Deutschen Gesellschaft für Psychologie in Zürich 1980* (Bd. 2, S. 482-486). Göttingen: Hogrefe.
- Nellessen, Horst. (1992). *Statistische Beurteilung des C-Tests im Vergleich mit anderen Prüfungsformen der SAP: vorläufige Bilanz (Ende '92)*. (Ms.).
- Norušis, Marija J. (1990a). *SPSS/PC+Statistics™ 4.0 for the IBM PC/XT/AT and PS/2*. Chicago, Ill.: SPSS Inc.
- Norušis, Marija J. (1990b). *SPSS/PC+™ 4.0 base manual for the IBM PC/XT/AT and PS/2*. Chicago, Ill.: SPSS Inc.
- Norušis, Marija J. (1990c). *SPSS/PC+ Advanced Statistics 4.0™ for the IBM PC/XT/AT and PS/2*. Chicago, Ill.: SPSS Inc.
- Oller, John W. Jr. & Streiff, Virginia. (1975). Dictation: A test of grammar-based expectancies. *English Language Teaching Journal*, 30, 25-36.
- Raatz, Ulrich & Klein-Braley, Christine. (1992). *CT-D4. Schulleistungstest Deutsch für 4. Klassen. Test und Beiheft mit Anleitung und Normentabellen*. Weinheim: Beltz.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (1994ff.). *Ausländische Bildungsnachweise und ihre Bewertung in der Bundesrepublik Deutschland: Bewertungsvorschläge*. Bonn (nur für den Dienstgebrauch vorgesehen).
- Singleton, David & Singleton, Emer. (1996). The C-Test and L2 acquisition/processing research. In James A. Coleman (Hrsg.), *University language testing and the C-Test. Proceedings of a conference held at the University of Portsmouth in April 1995*. University of Portsmouth Occasional Papers in Linguistics. (im Druck).
- Stevens, James. (1990). *Intermediate statistics: A modern approach*. Hillsdale, NJ: Erlbaum.
- Süßmilch, Edgar. (1984). Sprachleistungsmessung mittels C-Tests. *Finlance*, 3, 55-93.
- Süßmilch, Edgar. (1985). C-Tests für ausländische Schüler: Sprachdiagnose im Unterricht Deutsch als Zweitsprache. In Christine Klein-Braley &

- Ulrich Raatz (Hrsg.), *Fremdsprachen und Hochschule 19/14: Thematischer Teil: C-Tests in der Praxis* (S. 72-82). Bochum: AKS.
- Tarnai, Christian. (1989). Methodenvergleich des "Klassischen latent-additiven Testmodells" und des Rasch-Modells. In Klaus D. Kubinger (Hrsg.), *Moderne Testtheorie - Ein Abriss samt den neuesten Beiträgen* (2. verb. Aufl., S. 281-295). Weinheim: Beltz.

Anhang

SPRACHLICHE AUFNAHMEPRÜFUNG WS 1994/95

PRÜFUNGSTEIL GRAMMATIK

1. Setzen Sie die fehlenden Konjunktionen ein

Beispiel: Herr Krause hat den Wecker auf sieben Uhr gestellt, **damit** er in Ruhe frühstücken kann.

Herr Krause muß heute nach London fliegen. Meistens fährt ihn seine Frau, _____ er zum Flughafen muß. Aber _____ sie heute keine Zeit hat, bestellt Herr Krause ein Taxi, _____ er seinen Kaffee trinkt. _____ der Taxifahrer draußen hupt, nimmt Herr Krause sein Gepäck und geht zum Taxi. _____ es sehr kalt ist, hat er seinen Mantel nicht angezogen, _____ im Auto ist es ja warm.

2. Setzen Sie die Relativpronomen ein, die fehlen.

Beispiel: Der Fahrer stellt die Gepäckstücke, **die** Herr Krause bei sich hat, in den Kofferraum.

Der Taxifahrer, _____ Herrn Krause zum Flughafen bringt, stammt aus Jugoslawien. Das Taxi, _____ er fährt, gehört einem Mann, _____ er schon lange kennt und mit _____ Tochter er verheiratet ist: Es ist der Wagen seines Schwiegervaters, mit _____ er sein Geld verdient.

3. Lassen Sie die fettgedruckten Verben weg! Achten Sie auf die richtige Zeit.

Beispiel: Vor der Abflughalle **solte** das Taxi **anhalten**.
Vor der Abflughalle **hielt** das Taxi **an**.

Herr Krause **mußte** aussteigen.

Herr Krause _____

Er **wollte** in die Halle hineingehen.

Er _____

Am Schalter **mußte** er das Ticket vorzeigen.

Am Schalter _____

~~Der Koffer~~ muß man wiegen lassen.

Der Koffer _____

Die Taschen brauchte Herr Krause nicht abzugeben.

Die Tasche _____

Aber ein Polizist wollte die Taschen kontrollieren.

Aber ein Polizist _____

Herr Krause **mußte** auf einige Fragen antworten.

Herr Krause _____

Um zehn Uhr sollte die Maschine abfliegen.

Um zehn Uhr _____

4. Setzen Sie die fehlenden Präpositionen ein.

Beispiel: Herr Krause steht **an** dem Fenster der Abflughalle.

Zusammen _____ den anderen Passagieren wartet Herr Krause _____ die Öffnung des Ausgangs _____ dem Flugfeld. Dann steigen sie _____ den Bus, der sie _____ ihrem Flugzeug bringt. Eine Stewardess schaut _____ die Bordkarte _____ Herrn Krause und zeigt ihm seinen Platz _____ einem Fenster. Herr Krause unterhält sich _____ seiner Nachbarin, die _____ Schottland stammt, _____ Englisch _____ das Wetter.

5. Formen Sie die Sätze um. Bilden Sie Passivsätze. Achten Sie auf die richtige Zeit.

Beispiel: Die Flugbegleiterinnen verteilten Zeitungen.

Von den Flugbegleiterinnen wurden Zeitungen verteilt.

Die Türen der Maschinen hat man schon geschlossen.

Der Kapitän begrüßt die Fluggäste und informiert sie:

Man darf beim Start nicht rauchen.

Später werden die Flugbegleiter den Passagieren Orangensaft anbieten.

Sie verkaufen auch Zigaretten.

6. Setzen Sie die Adjektive in der richtigen Form ein.

Beispiel: Herr Krause verzichtet auf den _____ Orangensaft. (kalt)
Herr Krause verzichtet auf den kalten Orangensaft.

Er trinkt lieber _____ Tee. (heiß) Wegen des _____ Windes wird es ein _____ Flug. (stark, unruhig) Bei _____ Wetter kann man unten die _____ Küste sehen. (klar, belgisch) Aber heute sieht man nur _____ Wolken. (grau) - Später steht Herr Krause am _____ Flughafen und bekommt kein _____ Taxi. (London, frei)

PRÜFUNGSTEIL TEXTWIEDERGABE

Geben Sie den Text, den Sie zweimal hören, möglichst vollständig und richtig wieder!

Vorlesetext

Ballons

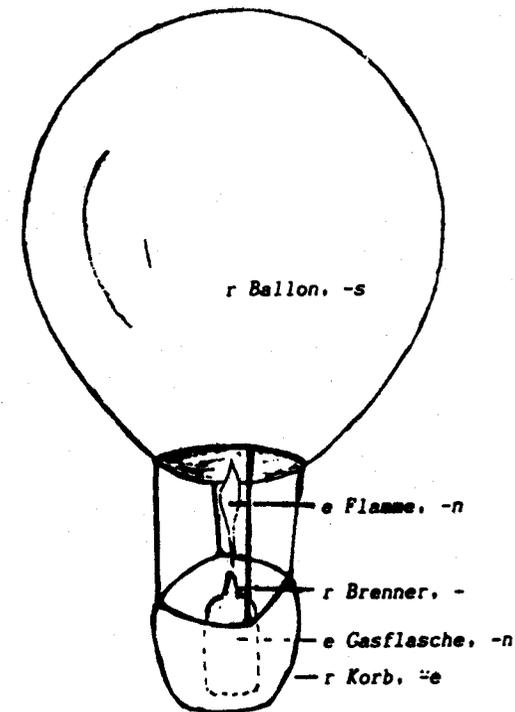
Manchmal sieht man hoch über der Erde einen Ballon. Ein solcher Ballon ist ein großer Sack aus einem leichten, festen Stoff, der unten offen ist. Unter dem Ballon hängt ein Korb für die Ballonfahrer, in dem sich auch eine Gasflasche mit einem Brenner befindet. Von Zeit zu Zeit kommt aus dem Brenner eine Flamme; dann steigt der Ballon ein bißchen höher.

Der Heißluftballon wurde vor mehr als 200 Jahren von den Brüdern Montgolfier in Frankreich erfunden. Diese hatten nämlich über eine Sache nachgedacht, die jeder schon einmal gesehen hat: Der Rauch über einem Feuer steigt nach oben. Die beiden Franzosen fragten sich, warum das so ist, und sie fanden auch die Antwort: Heiße Luft steigt auf, weil sie leichter ist als kalte.

Nun überlegten sich die Brüder Montgolfier folgendes: Wenn man einen Sack mit heißer Luft füllt, muß er ebenfalls nach oben steigen! Im Jahr 1783 bauten sie den ersten Ballon aus Papier, welcher immerhin zweieinhalb Kilometer weit flog. Noch im selben Jahr fand in Paris zum ersten Mal eine Ballonfahrt mit Personen statt.

Durch die Erfindung des Heißluftballons wurde für die Menschen der Traum vom Fliegen wahr. Außer Heißluftballons gibt es auch Gasballons, die vollständig geschlossen und mit einem leichten Gas gefüllt sind, das den Ballon in die Höhe trägt.

Ein Heißluftballon



die Brüder Montgolfier

PRÜFUNGSTEIL C-TEST

In den folgenden Texten gibt es unvollständige Wörter. Bitte vervollständigen Sie die Wörter so, daß sinnvolle, korrekte Sätze entstehen.

Beispiel: Ich m__ die Sprachp__ für me__ Studium mac__.
Ich **muß** die Sprach**prüfung** für mein Studium **machen**.

Jeden Morgen stehe ich um sechs Uhr dreißig auf. Zuerst ge__ ich i__ die Kü__ und ma__ das Früh___. Danach we__ ich sof__ die Kin___. Wir fr__hs__ dann zusa___. Dann bri__ ich s__ mit d__ Auto z__ Schule. Mittags ho__ ich s__ wieder a___. Wir neh__ zusammen i__ dem Eßz__ das Mittage___. ein. Nachmi__ machen w__ in d__ Supermarkt uns__ Einkäufe. Abends sind wir alle müde und gehen früh zu Bett.

Die Deutschen haben in der Regel einen Arbeitstag von etwa acht Stunden. Sie arbe__ an fü__ Tagen i__ der Wo__ und mac__ mehrere Woc__ im Ja__ Urlaub. Jeder Arbeitn__ hat d__ Recht a__ drei Woc__ Urlaub; d__ übliche Urlaubsd__ beträgt vi__ Wochen u__ mehr. D__ Durchschnittszahlen gel__ zwar ni__ für je__ Erwerbstätigen u__ für je__ Beruf, ab__ für d__ meisten Arbeitn__ geben s__ doch den Berufsalltag recht genau wieder.

Ein großes Problem für alle Großstädte nicht nur in Deutschland ist der zunehmende Verkehr. Jeden Mor__ fahren Mas__ von Berufst__ mit ih__ Auto i__ die Zen__ der Stä__ zur Arb__. Jeder Radioh__ kennt Meld__ über kilometerl__ Staus. I__ der Nä__ der Arbeitss__ ist e__ kaum mög__, einen Parkp__ zu fin__. Nur e__ besseres Sys__ von öffent__ Verkehrsmitteln kön__ hier ei__ Lösung bri__; denn Bu__, Straßenbahnen und U-Bahnen können viel mehr Menschen transportieren und brauchen keinen Parkraum.

Was erleichtert die geistige Arbeit? Es gi__ Dinge, d__ das Ler__ leichter mac__. Dazu geh__ ein bequ__ Arbeitsplatz m__ hellem Li__. Die Tem__peratur bet__ 20 b__ 22 Gr__. Wichtig i__ auch fri__ Luft i__ Arbeitszimmer. Da__ soll m__ öfter d__ Fenster öff. Außerdem m__ man si__ manchmal bew__ als Ausg__ für d__ sitzende Hal__. **Übershaupt** braucht der Lernende häufiger Pausen, damit sich Körper und **Gehirn** erholen.

Computer sind heute ein wesentliches Arbeitsmittel in allen Industriestaaten. Während v__ einigen Jah__ Computer n__ von wen__ Spezialisten bed__ werden kon__, erwartet m__ heute v__ fast je__ Studenten Kennt__ in d__ Benutzung sogen__ Personal-Computer. Di__ Geräte erse__ heute i__ den mei__ Büros d__ traditionellen Schreibm__. Computer si__ im Lau__ der Ja__ einerseits im__ schneller, andere__ auch kle__ und bill__ geworden. Und noch ist kein Ende dieser Entwicklung in Sicht.

Es ist schon lange bekannt, daß in Deutschland etwa 75% der Erwachsenen dauernd Medikamente einnehmen; vor einiger Zeit hat man auch den Arzneimittelkonsum von Schülern untersucht. Man fra__ in vie__ Familien, wa__ die Kin__ Medikamente beko__. Das Erge__ war überra__. Von fa__ der Häl__ aller jun__ Menschen zwis__ sechs u__ vierzehn Jah__ wird regelm__ mindestens e__ Medikament einge__. In d__ meisten Fäl__ handelt e__ sich u__ Mittel ge__ nervöse Stör__, Schlaflosigkeit, Kopfs__ und An__ vor Prüf__. Mediziner halten diesen Arzneimittelkonsum zum großen Teil für unnötig und sogar für gefährlich.

Ein Dieb schlich einmal in den Hof eines Mannes, denn er wollte ein Pferd stehlen. Die Bewohner des Hau__ sahen i__ und nah__ ihn gefa__. Da sch__ der Besi__ des Tie__ vor: "We__ du m__ jetzt zei__, wie m__ Pferde sti__, dann wi__ du freige__." Der Di__ war einver__ und wu__ sofort bef__. Er ri__: "Seht al__ her", tr__ zu d__ Pferd, spr__ hinauf u__ verschwand sch__. Obwohl ihn mehrere Männer verfolgten, gelang es keinem, ihn zu fangen.

Das Rote Kreuz ist eine Organisation, die in vielen Notfällen Hilfe leistet. Es ver_____ über Krankenw_____ und Är_____ und bes_____ sogar eig_____ Krankenhäuser. Es sch_____ Leute i_____ alle We_____, die Flüch_____ und Verwu_____ in Kriegsg_____ Hilfe lei_____. Es bet_____ Menschen, d_____ bei Überschw_____ und Erdb_____ obdachlos gewo_____ sind, das he_____, die ih_____ Wohnung verl_____ haben. D_____ Rote Kr_____ besteht i_____ vielen Län_____ der Er_____. Viele Helfer arbeiten freiwillig und ohne Bezahlung in ihrer freien Zeit mit.

Grotjahn, Rüdiger. (Hrsg.). (1996). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen* (Bd. 3, S. 343-366). Bochum: Brockmeyer

Helmut Daller

Der C-Test als Meßinstrument alltagssprachlicher und akademischer Sprachfähigkeiten türkischer Remigranten*

Im vorliegenden Artikel werden die Ergebnisse einer Untersuchung zum Sprachstand türkischer Germanistikstudenten beschrieben. Diese Studenten sind türkische Remigranten aus Deutschland. Zur Messung alltagssprachlicher und akademischer Sprachfähigkeiten im Deutschen werden zwei C-Tests mit unterschiedlicher Textgrundlage eingesetzt. Die Ergebnisse werden mit einer Kontrollgruppe aus türkischen Germanistikstudenten, die Deutsch als Fremdsprache erworben haben, verglichen. Während die Remigranten hohe Ergebnisse im alltagssprachlichen Bereich erzielen, sind ihre akademischen Sprachfähigkeiten vergleichsweise gering. Ein Vergleich der Testergebnisse mehrerer Studienjahre ergibt, daß sich die akademischen Sprachfähigkeiten der Rückkehrer auch nach einem mehrjährigen Germanistikstudium nicht nachweisbar verbessern.

1. Einleitung

Im Dezember 1992 wurde zur vorliegenden Thematik eine erste Pilotstudie unter Germanistikstudenten an mehreren Universitäten der Türkei durchgeführt. Diese Studenten sind zu mehr als 95% sogenannte Rückkehrer aus der Bundesrepublik Deutschland. Die Stichprobe bestand aus 210 Studenten der Germanistikabteilungen folgender Universitäten: Atatürk Universität in Ankara, Cukurova Universität in Adana und Uludağ Universität in Bursa. Die Testpersonen waren Studenten aller vier Jahrgangsstufen. Den Studenten wurde ein Fragebogen zur Spracherwerbsbiographie und ein Sprachtest vorgelegt. Der Sprachtest basierte auf dem C-Test-Prinzip.

Die Pilotstudie stellt einen Teil einer in Arbeit befindlichen Dissertation¹ dar, die u.a. zum Ziel hat, die Unterschiede zwischen "alltagssprachlichen

* Ich danke Herrn Dr. Rüdiger Grotjahn für die zahlreichen Kommentare zu einer früheren Fassung dieses Artikels. Für die Hilfe bei der Datensammlung danke ich meinen Kollegen Jutta Bartolosch (Ankara), Joachim Braun (Istanbul), Siegfried Mühlfelder (Bursa) und Suzan Üner (Adana).

¹ Die Dissertation wird von Herrn Dr. R. Grotjahn, Ruhr-Universität Bochum, und Herrn Prof. Dr. W. Herrlitz, Universität Utrecht, betreut.