

## **“Tests and Attitude Scales for the Year Abroad” (TESTATT): Sprachlernmotivation und Einstellungen gegenüber Sprechern der eigenen und der fremden Sprache [1]**

Rüdiger Grotjahn

### **1 Einleitung**

Im Rahmen des von 1993 bis 1996 durchgeführten „European Language Proficiency Survey (ELPS)“ wurden bei über 25000 Studierenden verschiedener Fremdsprachen in mehreren Ländern Europas mit Hilfe von C-Tests die allgemeine Sprachbeherrschung in der jeweiligen Fremdsprache gemessen sowie mit einem informellen Fragebogen Daten zur Sprachlerngeschichte, zur Fremdsprachenverwendungsangst (*foreign language anxiety*), zur Sprachlernmotivation und zu speziellen Einstellungen gegenüber dem eigenen und dem zielsprachlichen Land erfasst und analysiert. [2]

Ein Ergebnis der ELPS-Studie war, dass ein Auslandsaufenthalt der wichtigste Einflussfaktor im Hinblick auf den Lernzuwachs in der Fremdsprache während des Studiums war. Allerdings zeigte eine Analyse der Einzeldaten, dass nicht jeder Studierende in gleicher Weise hiervon profitierte. Mittels des Fragebogens konnten moderierende Faktoren für die Effektivität des Auslandsaufenthalts aufgefunden werden. Dazu gehörten u.a. Einstellungen, die Sprachlernmotivation, Lernstrategien und die Bereitschaft zur Akkulturation im zielsprachlichen Land (L2-Land). Allerdings konnten im Rahmen der ELPS-Studie zu jedem Komplex nur wenige Einzelfragen gestellt werden, so dass die Ergebnisse eher als Hypothesen denn als gesicherte Tatbestände gelten konnten.

Zur Überprüfung dieser und weiterer Hypothesen in Folgeuntersuchungen erschien es notwendig, den in dem Projekt verwendeten Fragebogen zu überarbeiten bzw. gezielt ganz neue Instrumente zu entwickeln. Zu diesem Zweck wurde von 1998-2001 das Anschlussprojekt “Tests and Attitude Scales for the Year Abroad” (TESTATT) durchgeführt. Die Hauptaufgabe dieses Projekts war die Weiterentwicklung und Erprobung der Skalen und Fragebögen, die bei dem Vorgängerprojekt ELPS eingesetzt worden waren. Außerdem sollten für mögliche Folgeuntersuchungen Batterien von C-Tests unterschiedlicher Schwierigkeit für mehrere Fremdsprachen entwickelt werden.

Alle diese Verfahren sollten so konzipiert werden, dass sie einerseits in größeren, internationalen Forschungsprojekten eingesetzt werden konnten, andererseits aber auch dem einzelnen Studierenden bei der Planung eines Auslandsaufenthalts individuelle Informationen als Entscheidungshilfe geben konnten. Fragebögen und Tests sollten sowohl in klassischer Form im Papier-Bleistift-Format als auch computerbasiert über das Internet administriert werden können.

Bei der Entwicklung von computertauglichen und ökonomischen Skalen und Tests für verschiedene Sprachen und einem internationalen Einsatzbereich war es zunächst notwendig, geeignete Skalen und Sprachtests zusammenzustellen, in Papier-Bleistift-Form zu bringen und an größeren Stichproben in verschiedenen Ländern zu erproben. Hierbei waren verschiedene Randbedingungen zu beachten:

---

-2-

1. Da die Messinstrumente computertauglich sein sollten, konnten nur geschlossene Items verwendet werden, wie z.B. Ratingskalen (bei Fragebögen) oder Ergänzungitems (bei C-Tests).
2. Die Anzahl der Verfahren und auch die Anzahl der Items pro Verfahren sollte aus Gründen der Ökonomie möglichst gering sein.
3. Die Fragebögen sollten parallel in mehreren Sprachen, zunächst in Deutsch und Englisch, entwickelt werden. Bei solchen Parallelentwicklungen müssen allerdings neben übersetzungsspezifischen Problemen mögliche populationsbedingte Unterschiede in den Validitäten und Faktorenstrukturen der Instrumente beachtet werden.
4. Die Fragebögen mussten für verschiedene Zielsprachen und verschiedene Gastländer tauglich und damit inhaltlich sehr allgemein sein (vgl. zu der in Punkt 3 und 4 genannten Problematik Hambleton & de Jong 2003).
5. Da die Skalen auch Aussagen über einzelne Studierende ermöglichen sollten, mussten sie hoch reliabel sein. Das bedeutet, dass zu jedem Konstrukt eine ausreichende Zahl von Fragen gestellt und zu Skalen mit

einem Gesamtpunktwert zusammengefasst werden musste, falls die Faktorenstruktur es zuließ.[3]

6. Schließlich sollen die Fragebögen und Tests, die im TESTATT-Projekt entwickelt und erprobt werden sollten, den Instrumenten der ELPS-Studie formal und inhaltlich möglichst entsprechen, damit unmittelbare Vergleiche hergestellt werden konnten.

Nach Sichtung der Literatur und Vorversuchen an kleineren Stichproben in Duisburg fiel die Wahl schließlich auf die im Folgenden beschriebenen Sprachtests und Fragebögen.

## 2 Messinstrumente

### 2.1 Sprachtests

Als ökonomische Methode zur Messung des allgemeinen Sprachstandes in der gelernten Sprache, z.B. vor und nach einem Auslandsaufenthalts im Rahmen eines Fremdsprachenstudiums, haben sich C-Tests vorzüglich bewährt. Da auch bei der ELPS-Studie C-Tests mit großem Erfolg eingesetzt worden waren, wurde entschieden, im Rahmen des TESTATT-Projekts ebenfalls C-Tests in den Sprachen Deutsch, Englisch, Französisch und Spanisch zu entwickeln und zu erproben.

Neben C-Tests wurde in einigen im Rahmen von TESTATT durchgeführten Untersuchungen auch der Gesamtpunktwert des „Bochumer Einstufungstest Französisch“ und des „Duisburg English Language Test for Advanced Students – DELTA“ als Maß allgemeiner Fremdsprachenkompetenz benutzt (vgl. zu diesen Tests Grotjahn 1986; Schwibbe & Schwibbe 1987). Der DELTA-Test enthält als festen Bestandteil auch einen englischen C-Test. Es wurde deshalb auf die Entwicklung von C-Tests für Englisch letztendlich verzichtet. Auf den Zusammenhang der untersuchten Variablen mit dem DELTA und dem „Bochumer Einstufungstest Französisch“ werde ich im Folgenden nicht gesondert eingehen.

---

-3-

Die C-Tests bestanden in der Erprobungsform jeweils aus sechs Texten mit 20 Lücken und waren auf der Basis der klassischen „rule of two“ konstruiert: Beginnend mit dem zweiten Wort des zweiten Satzes wurde bei jedem zweiten Wort die zweite Hälfte getilgt (vgl. zum C-Test z.B. Grotjahn 1995, 2004 sowie <http://www.c-test.de>). Für jede Sprache wurden vier Testformen mit unterschiedlicher Schwierigkeit zusammengestellt, wobei zwei Texte jeweils identisch waren und als Ankeritems zur Kalibrierung der Schwierigkeiten dienten. Im Zuge der Erprobung wurde ein Text aus jeder C-Test-Version eliminiert, so dass die Versionen aus jeweils fünf Texten à 20 Lücken bestehen (maximale Punktzahl = 100). Die Entwicklung der C-Tests ist detailliert in Raatz, Grotjahn & Wockenfuß (2004) beschrieben.

### 2.2 Fragebögen

Im Rahmen des Projekts wurden fünf Fragebögen jeweils in deutscher und englischer Sprache zur Messung der Extraversion, der Fremdsprachenverwendungsangst, der Sprachlernmotivation und der Einstellung zu Personen der eigenen (L1) und der fremden (L2) Nationalität entwickelt und erprobt. Bei der Übersetzung der Fragebögen wurde die Polungsrichtung der Items beibehalten (zur Begründung vgl. z.B. Hinz, Brähler, Geyer & Körner 2003: 162). Diese Fragebögen umfassten in ihrer Erprobungsform im Fall von „Sprachlernmotivation“ 10 vierstufige Items, im Fall der übrigen Skalen jeweils 16 vierstufige Items (Likert-Skalen).

Der Entscheidung für die genannten Variablen lag u.a. die Hypothese zugrunde, dass extravertierte Studierende mit wenig Fremdsprachenverwendungsangst, hoher Sprachlernmotivation und positiver Einstellung zum Gastland sprachlich stärker von einem Auslandsaufenthalt profitieren würden als Lerner mit entgegengesetzten Eigenschaften. Zudem zählt die Variable „Extraversion“ zu den in modernen Persönlichkeitstheorien zumeist unterschiedenen fünf großen Persönlichkeitsfaktoren (so genannte „Big Five“; vgl. [Abschnitt 2.2.1](#)).

Die Entscheidung für eine vierstufige Antwortskala anstelle der häufig verwendeten fünf- oder auch siebenstufigen Likert-Skalen erfolgte zum einen aus Ökonomiegründen. Vor allem sollte aber beim Antwortverhalten eine mögliche Tendenz zur Mitte ausgeschlossen werden. So haben z.B. Rost, Carstensen & von Davier (1999) mit Hilfe der probabilistischen Testtheorie bei der deutschen Version des NEO-Fünf-Faktoreninventars (NEO-FFI) von Costa und McCrae (vgl. Borkenau & Ostendorf 1993 sowie auch Hinz, Brähler, Geyer & Körner 2003) bei allen fünf Skalen zur Messung der „Big Five“ festgestellt, dass die mittleren Schwellenparameter der jeweils fünfstufigen Items sehr dicht beieinander lagen oder zumeist sogar vertauscht

waren. Die Autoren stellen hierzu fest:

„Ein solcher Befund, daß die mittlere Antwortkategorie nicht wirklich in der Mitte der zu messenden Dimension liegt, ist sehr häufig bei Ratingskalen zu finden und weist darauf hin, daß diese Kategorie eben *nicht* die zu messende Eigenschaft mit mittlerer Intensität anspricht. Vielmehr kann daraus sogar der Schluß gezogen werden, daß mit dieser Kategorie eine *andere Dimension* erfaßt wird als mit den anderen Kategorien, der Test also zweidimensional ist. So kann es sein, daß die Mittelkategorie nur dann gewählt wird ..., wenn die befragte Person das Item als für sich unzutreffend auffaßt oder aus anderen Gründen nicht antworten möchte. In diesem Fall würde die Mittelkategorie also die Skalierbarkeit oder die Offenbarungsbereitschaft der Person erfassen.“ (Rost, Carstensen & von Davier 1999: 122; Hervorhebungen im Original).

---

-4-

Rost, Carstensen & von Davier (1999: 122) kommen deshalb zu dem für das vierstufige TESTATT-Format sprechenden Schluss, dass „für den NEO-FFI unter meßtheoretischen Gesichtspunkten ein vierstufiges Antwortformat mit zwei zustimmenden und zwei ablehnenden Kategorien angemessener ist“ (vgl. auch Brown 2001: 41 sowie Smit, Kelderman & van der Flier 2003).

Auf die Fragebögen sowie auf einige empirische Resultate aus den Erprobungen soll im Folgenden etwas genauer eingegangen werden – und zwar schwerpunktmäßig auf die Variablen „Einstellung zu Personen der eigenen und der fremden Nationalität“ und „Sprachlernmotivation“. Genauere Ausführungen zu den Variablen „Extraversion/ Introversion“ und „Fremdsprachenverwendungsangst“ finden sich in Grotjahn, Raatz & Wockenfuß (2004). Einen Gesamtüberblick über die Skalen und Ergebnisse des TESTATT-Projekts geben Grotjahn, Raatz, Wockenfuß & Coleman (2001).

Bei der Interpretation der empirischen Resultate ist natürlich zu berücksichtigen, dass Sprachlernmotivation, Extraversion/Introversion, Fremdsprachenverwendungsangst, Einstellungen und Sprachstand in einer L2 in komplexer Weise miteinander in Wechselwirkung stehen und zudem mit einer Vielzahl weiterer Variablen korreliert sind. Da der Schwerpunkt des vorliegenden Beitrags im Bereich der Skalenentwicklung liegt, werde ich auf mögliche komplexe multivariate Beziehungen zwischen den genannten Variablen nicht eingehen. [4]

### 2.2.1 Extraversion/Introversion

Wie bereits angedeutet, unterscheiden moderne Persönlichkeitstheorien häufig zwischen fünf voneinander (weitgehend) unabhängigen grundlegenden Persönlichkeitsmerkmalen (*traits*). Extraversion/ Introversion gilt als eines dieser basalen Persönlichkeitsmerkmale, wobei der typische Extravertierte und der typische Introvertierte die Endpole eines Kontinuums darstellen (vgl. Amelang & Bartussek 1997; Borkenau & Ostendorf 1993; Hinz, Brähler, Geyer & Körner 2003; MacIntyre & Charos 1996; Matthews & Deary 1998; sowie auch das „Big Five + One Persönlichkeitsinventar“ (B5PO; <http://www.schuhfried.de>)).

In der Fremdsprachenforschung wird zumeist die Extraversion als wünschenswertes Ende des Kontinuums angesehen (vgl. auch Brown, Robson & Rosenkjar 2001; Dewaele & Furnham 1999; Kiany 1997; Verhoeven & Vermeer 2002). So wird z.B. argumentiert, dass extravertierte Lerner eher mit anderen Sprechern der Zielsprache interagieren und als Folge auch eher erwerbsfördernden Input erhalten. Wie jedoch Dewaele und Furnham (1999: 525ff) anhand der relevanten Literatur zeigen, sind die Leistungen extravertierter Lerner bei **schriftlichen** zielsprachlichen Aufgaben keineswegs notwendigerweise besser als die introvertierter Lerner. Signifikante Korrelationen gab es jedoch zwischen Extraversion und Flüssigkeit bei **mündlichen** zielsprachlichen Aufgaben.

Insgesamt gesehen lassen sich auf der Basis der vorliegenden empirischen Forschungsergebnisse u.a. folgende Vorhersagen formulieren:

1. Je höher die Extraversion, desto niedriger ist die Angst beim Gebrauch der Fremdsprache und desto höher ist die Fremdsprachenlernmotivation.

2. Es gibt keine Korrelation zwischen Extraversion und einer schriftlichen Aufgabe wie dem C-Test.

Beide Vorhersagen wurden im Rahmen des TESTATT-Projekts bestätigt.

---

-5-

Auf der Basis der vorliegenden Forschungsergebnisse kann weiterhin vorhergesagt werden, dass extravertierte

Studierende eher als introvertierte Studierende bereit sind, an einem Auslandsaufenthaltsprogramm teilzunehmen (vgl. auch die entsprechende Studie von Müßig-Trapp & Schnitzer 1997).

Zur Entwicklung eines ökonomischen Instruments zur Messung von Extraversion/Introversion haben wir eine von Mummendey (1995: 122ff) für Probanden mit Deutsch als Muttersprache entwickelte Skala adaptiert, die wiederum eine Adaptierung der deutschen Version des *Eysenck Personality Inventory* (EPI) von Eggert (1983) darstellt. [5] Bei der Adaptierung der Mummendey-Skala wurden von uns acht Items ausgeschlossen und die siebenstufige Schätzskaala auf vier Stufen mit den Eckpunkten „trifft sicher zu“ und „trifft sicher nicht zu“ reduziert.

Die resultierende Erprobungsskala bestand aus 16 Items in Form kurzer Aussagen wie „Ich gehe viel aus“. Die Skala wurde sodann empirisch im Hinblick auf ihre messtheoretischen Eigenschaften analysiert. Die Ergebnisse sind in Grotjahn, Raatz & Wockenfuß (2004) dargestellt. [6]

### 2.2.2 Fremdsprachenverwendungsangst

Fremdsprachenverwendungsangst (*foreign language anxiety*) ist einer der besten Prädiktoren im Hinblick auf den Erfolg beim Erlernen von Fremd- und Zweitsprachen. Fremdsprachenverwendungsangst kann die Verarbeitung, das Behalten und die Produktion von Sprache auf allen Ebenen negativ beeinflussen und korreliert u.a. mit Problemen beim Hörverstehen, beim Lexikerwerb und bei der Wortproduktion sowie auch mit geringen Leistungen bei standardisierten Tests oder schlechten Noten in Sprachkursen (vgl. z.B. Brown, Robson & Rosenkjar 2001; Gardner, Tremblay & Masgoret 1997; Horwitz 2001).

Angst bei der Verwendung einer Fremdsprache kann als eine spezielle Form situationspezifischer Angst aufgefasst werden. Horwitz, Horwitz & Cope (1991/1986: 31) z.B. verstehen darunter „a distinct complex of self-perceptions, beliefs, feelings, and behaviors ... arising from the uniqueness of the language learning process“. Sie sind der Meinung, dass im Fall von erwachsenen Fremdsprachenlernern Sprachverwendungsangst in erster Linie das Resultat der Selbstwahrnehmung der Lerner hinsichtlich der Inadäquatheit der eigenen kommunikativen Fähigkeiten in der Fremdsprache ist. Heutzutage wird allerdings zunehmend davon ausgegangen, dass der Zusammenhang zwischen Angst und Erfolg beim Erlernen einer L2 bidirektional ist, d.h., dass Fremdsprachenverwendungsangst sowohl eine Ursache als auch eine Folge des Erfolgs beim Erlernen einer L2 sein kann (vgl. auch Ganschow & Sparks 2001; Sparks & Ganschow 2001).

Es sind eine Reihe von Skalen zur Messung von Fremdsprachenverwendungsangst vorgeschlagen worden (vgl. die Hinweise in Grotjahn, Raatz & Wockenfuß 2004 sowie Ganschow & Sparks 2001; Horwitz 2001; MacIntyre 1999; Onwuegbuzie, Bailey & Daley 2000; Rodríguez & Abreu 2003). Im Rahmen des TESTATT-Projekts wurde auf die *Foreign Language Classroom Anxiety Scale* von Horwitz, Horwitz & Cope (1991/1986) zurückgegriffen, und zwar mit folgender Begründung: (1) Die Reliabilität und Validität der Skala ist hinreichend nachgewiesen. (2) Für viele Studierende, die einen Auslandsaufenthalt planen, ist der fremdsprachenunterrichtliche Kontext der entscheidende Faktor im Hinblick auf die individuellen Erfahrungen mit der jeweiligen L2.

Die *Foreign Language Classroom Anxiety Scale* besteht aus 33 Items mit fünf Antwortoptionen. Für die Erprobung im Rahmen von TESTATT wurden 16 Items ausgewählt, wie z.B.: „Ich zittere, wenn ich weiß, dass ich im Fremdsprachenunterricht gleich aufgerufen werde.“ Items zur Messung von Testangst blieben unberücksichtigt. Wie bereits im Fall von Extraversion/Introversion wurde die Skala auf vier Antwortoptionen reduziert.

### 2.2.3 Einstellungen

Die beiden im TESTATT-Projekt verwendeten Skalen zur Messung der Einstellung zu Personen der eigenen (L1) und der fremden (L2) Nationalität wurden aus dem „European Language Proficiency Survey“ übernommen (vgl. Coleman 1996: 100ff; 1997). Ein Teil der Items wurde bereits in mehreren früheren Untersuchungen verwendet (vgl. z.B. Gardner 1985; Gardner & Lambert, 1972; Pickett 1993; Willis, Doble, Sankaraya & Smithers 1977). Sie bestehen jeweils aus 16 Adjektiven, die Eigenschaften wie „gefühlvoll“, „arrogant“ oder „freundlich“ bezeichnen. Die Skalen beruhen auf der Technik des Semantischen Differential von Osgood (vgl. Osgood, Suci & Tannenbaum 1957) und auf dem Stereotypen-Differential von Gardner, Kirby, Gorosbe & Villamin (1972). Zur Messung der Einstellung gegenüber Angehörigen der Zielkultur wurden die Studierenden gefragt: „Welche der folgenden Adjektive beschreiben Ihrer Meinung nach am besten Menschen, die die Sprache sprechen, in der Sie getestet werden?“ Zur Messung der Einstellung gegenüber Angehörigen der eigenen Kultur wurden die Studierenden gefragt: „Welche der folgenden Adjektive beschreiben Ihrer Meinung

nach am besten Menschen ihrer eigenen Nationalität?“ Die Adjektive beider Skalen waren jeweils identisch. Die Bewertung der Adjektive erfolgte auf einer vierstufigen bipolaren Einstufungsskala – im Gegensatz z.B. zu Osgood, Suci & Tannenbaum (1957) oder Gardner & Lambert (1972), die jeweils eine siebenstufige Skala verwendet haben (vgl. zur Technik des Semantischen Differentials auch Finkbeiner 2001; Schuman & Presser 1996). Der Einstellungsfragebogen (L1) ist im Anhang 1 des vorliegenden Beitrags wiedergegeben.

#### 2.2.4 Motivation

Motivation ist ein komplexes Konstrukt, das verschiedene Dimensionen wie Intentionen, Ziele, Orientierungen, Aufmerksamkeit, Anstrengung und Persistenz umfasst. Motivation gilt neben der Fremdsprachenlernneigung (*foreign language aptitude*) als wichtigster Prädiktor des Lernerfolgs (vgl. z.B. Dörnyei 2001, 2003; Dörnyei & Schmidt 2001; Kleppin 2001, 2002; Masgoret & Gardner 2003; McIntosh & Noels 2004 in dieser Nummer; Noels, Pelletier, Clément & Vallerand 2000; Riemer 2001; Stiensmeier-Pelster & Rheinberg 2003).

Im TESTATT-Projekt wurde den Studierenden folgende Frage gestellt: „Warum lernen Sie die Fremdsprache, die Sie gewählt haben?“ Es wurden 10 Gründe vorgegeben, wie z.B. „für Ihren zukünftigen Beruf“ oder „weil Sie diese Sprache mögen“, die jeweils auf einer vierstufigen Skala in ihrer Wichtigkeit bewertet werden mussten. Die 10 Gründe wurden aus dem ELPS aufgrund ihrer Antworthäufigkeit übernommen und sollten eine extrinsische/instrumentelle und eine intrinsische/integrative Orientierung (Dimension) erfassen. Der Motivationsfragebogen ist im Anhang 2 des vorliegenden Beitrags wiedergegeben.

#### 2.3 Überblick über die verwendeten Tests und Skalen

In Tabelle 1 findet sich ein Überblick über die im TESTATT-Projekt verwendeten Skalen und Tests.

*Tabelle 1: Skalen und Tests*

<i>C-Tests</i>	
Zahl der Items	6 Texte à 20 Lücken; 4 Versionen unterschiedlicher Schwierigkeit mit jeweils 2 identischen Texten als Ankeritems; Sprachen: Deutsch, Französisch, Spanisch
<i>Extraversion/Introverson</i>	
Quelle:	Mummendey (1995) Müßig-Trapp, & Schmitzer (1997)
Zahl der Items:	16
<i>Fremdsprachenverwendungsangst</i>	
Quelle:	FLCAS (Foreign Language Classroom Anxiety Scale), Horwitz, Horwitz, & Cope (1991/1986)
Zahl der Items:	16
<i>Einstellung (L1)</i>	
Quelle:	Coleman (1996) Pickett (1993)
Zahl der Items:	16
<i>Einstellung (L2)</i>	
Quelle:	Coleman (1996) Pickett (1993)
Zahl der Items:	16
<i>Sprachlernmotivation</i>	
Quelle:	Coleman (1996)
Zahl der Items:	10

**Anmerkung:** Bei allen fünf Skalen wurden für die möglichen Antworten jeweils vier Stufen („trifft sicher zu“ bis „trifft sicher nicht zu“) vorgegeben

### 3 Versuchspersonen

Die Versuchspersonen (Vpn) waren insgesamt 427 Studierende an den Universitäten von Duisburg, Bochum und Portsmouth (England). Die meisten deutschen Vpn studierten im ersten Jahr Englisch bzw. Französisch. Die englischen Vpn waren entweder aus dem zweiten oder vierten Studienjahr und studierten Spanisch, Französisch oder Deutsch (vgl. Tabelle 2).

Tabelle 2: Stichproben

Gruppe	Ort	Zielsprache	Niveau	n	männl.	weibl.
1	Duisburg	Englisch	Studienanfänger	110	20	90
2	Bochum	Französisch	worwiegend 1. Jahr	81	15	66
3	Portsmouth	Deutsch	2. Jahr	24	9	15
4	Portsmouth	Deutsch	4. Jahr	28	5	23
5	Portsmouth	Spanisch	2. Jahr	38	3	35
6	Portsmouth	Spanisch	4. Jahr	50	9	41
7	Portsmouth	Französisch	2. Jahr	32	9	23
8	Portsmouth	Französisch	4. Jahr	64	16	48
Gesamt				427	86	341

### 4 Durchführung der Untersuchung und statistische Verfahren

Leider konnten die C-Tests und Fragebögen nicht in einer festen Reihenfolge und mit einem vorher festgelegtem „Timing“ eingesetzt werden. Reihenfolgeeffekte können deshalb nicht kontrolliert und somit auch nicht ausgeschlossen werden.

Zudem wurden die Fragebögen und Tests von den Studierenden nicht anonym bearbeitet, da dies die Identifizierung unterschiedlicher Datensätze ein und desselben Studierenden erschwert hätte. Ein Einfluss des Faktors „soziale Erwünschtheit“ auf das Antwortverhalten ist deshalb ebenfalls nicht auszuschließen. Auf den englischen Fragebögen fand sich allerdings folgender Hinweis: „Your personal details are confidential and will not be published, but we need them to identify individual respondents.“

Zur Überprüfung der Dimensionalität der einzelnen Skalen wurden bei allen statistischen Analysen Maximum-Likelihood-Faktorenanalysen der Pearson Produkt-Moment-Interkorrelationsmatrix der Items mit Hilfe von SPSS durchgeführt. Im Hinblick auf die Interpretation der Resultate dieser Analysen ist zu berücksichtigen, dass die Verwendung von Pearson Produkt-Moment-Korrelationen und Maximum-Likelihood-Faktorenanalysen im Fall Likert-skalierten Items zwar weithin üblich, aber nichtsdestoweniger aus einer Reihe von Gründen problematisch ist (siehe [Abschnitt 6](#)).

Um die Interpretation zu erleichtern, wurde lediglich nach dem Varimax Kriterium rotiert, d.h. auf der Basis unkorrelierter Faktoren (vgl. Backhaus, Erichson, Plinke & Weiber 2002, Kap. 5 und 6; Onwuegbuzie, Bailey & Daley 2000; Stevens 1996, Kap. 11; Tabachnick & Fidell 1989, Kap. 12; Wood, Tataryn & Gorsuch 1996). In Anbetracht der relativ geringen Stichprobengrößen und des hohen Stichprobenfehlers der Faktorladungen wurden nur Ladungen mit Werten größer als  $|.40|$  bei der Interpretation berücksichtigt. Hinsichtlich der Zahl der extrahierten Faktoren haben wir uns sowohl von den von SPSS berechneten Anpassungsmaßen und Interpretationshilfen als auch von theoretischen Überlegungen leiten lassen.

Zur Bestimmung der Reliabilität ( $r_{tt}$ ) der entwickelten Instrumente wurde Cronbachs Alpha als Maß interner

Konsistenz benutzt (vgl. hierzu z.B. Brown 2001: 173ff; Tremblay 2001: 240ff).

Die Übereinstimmungsvalidität der Skalen und Tests wurde anhand der bivariaten Korrelationen der jeweiligen Summenscores ermittelt. Trotz gewisser methodologischer Bedenken wurden in allen Fällen Pearson Produkt-Moment-Korrelationen berechnet (vgl. hierzu [Abschnitt 6](#)). Außer bei den Einstellungsskalen erfolgten die Berechnungen stets auf der Basis der jeweiligen Ausgangsskala, d.h. der noch nicht faktoranalytisch überprüften Skala.

Unterschiede zwischen den Mittelwerten der einzelnen Teilstichproben wurden mit Hilfe von *t*-Tests bzw. Varianzanalysen auf Signifikanz überprüft. Da die Untersuchung in erster Linie ein exploratorisches Ziel verfolgte, wurde im Fall multipler Signifikanztests bivariater Korrelationen auf eine Adjustierung des Signifikanzniveaus z.B. mit Hilfe des Bonferroni-Holm-Verfahrens verzichtet. Bei den angegebenen Wahrscheinlichkeiten handelt es sich stets um zweiseitige Wahrscheinlichkeiten. Da in vielen Fällen aufgrund von früheren Untersuchungen einseitige Hypothesen formuliert werden können, sind die angegebenen Wahrscheinlichkeiten in einer Reihe von Fällen zu halbieren. Dem Vergleich der Teilstichproben lag die interaktionistische Hypothese zugrunde, dass Persönlichkeitstraits nur eingeschränkt situationell konsistent sind (vgl. hierzu bereits Bem & Allen 1974).

## 5 Einige Ergebnisse

### 5.1 Extraversion/Introversion

Bei der Extraversion-/Introversionsskala gab es keine signifikanten Mittelwertunterschiede zwischen den acht Teilstichproben. Die Reliabilitäten ( $r_{tt}$ ) der Skala lagen in den Teilstichproben zwischen .50 und .87 bei einem Median von 83.5. Für die Gesamtstichprobe betrug die Reliabilität .83. Damit erweist sich die Gesamtskala nicht nur für gruppenbezogene, sondern auch für individuenbezogene Entscheidungen als hinreichend reliabel. Die Faktorenanalysen ergaben sowohl in der Gesamtgruppe als auch in der deutschsprachigen und englischsprachigen Teilgruppe drei interpretierbare Faktoren mit jeweils fünf substantiellen Ladungswerten und 38.5% aufgeklärter Varianz. Ein Item (Item 7) wurde aufgrund der Faktorenanalysen ausgeschlossen. Die Faktoren wurden als „Soziabilität“ (17.3%;  $r_{tt} = .77$ ), „Spontaneität“ (10.8%;  $r_{tt} = .65$ ) und „Zurückhaltung/Schüchternheit“ (10.4%;  $r_{tt} = .74$ ) interpretiert. Allerdings war die Interkorrelation zwischen den Items der beiden Teilskalen „Soziabilität“ und „Zurückhaltung/ Schüchternheit“ mit .58 relativ hoch (die beiden übrigen Interkorrelationen betragen jeweils .35), so dass möglicherweise lediglich zwei Teildimensionen von Extraversion/Introversion, nämlich Soziabilität und Spontaneität unterschieden werden sollten. Die Reliabilitäten der Teilskalen sind in Anbetracht der geringen Zahl von jeweils fünf Items sehr zufrieden stellend. Damit zeigt sich auch in den TESTATT-Untersuchungen die sowohl in Faktorenanalysen als auch in Analysen auf der Basis der probabilistischen Testtheorie beobachtete Mehrdimensionalität des Konstrukts „Extraversion/Introversion“ (vgl. Rost, Carstensen & von Davier 1999: 124). Weitere Hinweise zur Skala „Extraversion/Introversion“ finden sich in Grotjahn, Raatz & Wockenfuß (2004).

### 5.2 Fremdsprachenverwendungsangst

Im Vergleich zu den Gruppen 3-6 aus Portsmouth wies die Gruppe 1 der Studierenden aus Duisburg eine signifikant niedrigere Fremdsprachenverwendungsangst auf. Die Ausgangsskala mit 16 Items war in allen Teilgruppen höchst reliabel (jeweils  $r_{tt} > .87$ ; in der Gesamtgruppe  $r_{tt} = .93$ ). Die Faktorenanalysen führten für die Ausgangsskala in der Gesamtgruppe zu zwei Faktoren, die 29.5% und 19.2% an Varianz aufklärten. Nach Ausschluss von fünf Items ergab sich in der Gesamtgruppe ein einziger Faktor (50.3% aufgeklärte Varianz). Für die Teilstichproben resultierten in zwei Fällen einfaktorielle Lösungen (45.7% und 54.6% aufgeklärte Varianz) und in einem Fall eine zweifaktorielle Lösung (32.9% und 20.9% aufgeklärte Varianz). Auch die verkürzte Skala (11 Items) weist mit Reliabilitäten zwischen .90 und .93 eine äußerst zufrieden stellende Messgenauigkeit auf. Dies gilt auch im Vergleich zu der der TESTATT-Skala zu Grunde liegenden Originalskala, nämlich der aus 33 Items mit fünf Antwortoptionen bestehenden *Foreign Language Classroom Anxiety Scale* von Horwitz, Horwitz und Cope (vgl. Abschnitt [2.2.2](#)).

Die Faktoren der zweifaktoriellen Lösung wurden auf der Basis der beiden jeweils am höchsten ladenden Items (den so genannten Marker-Variablen) als „Emotionen und deren physische Korrelate“ und „fehlendes Selbstbewusstsein“ etikettiert. Da jedoch einige Items auf beiden Faktoren substantielle Ladungen aufwiesen, war eine eindeutige inhaltliche Interpretation der zweifaktoriellen Lösung und damit zugleich eine Unterscheidung von Teilskalen nicht möglich. Weitere Hinweise zur Skala „Fremdsprachenverwendungsangst“



finden sich in Grotjahn, Raatz & Wockenfuß (2004).

Die Korrelation zwischen Fremdsprachenverwendungsangst und Extraversion/Introversion (berechnet anhand der Werte für die ungekürzten Ausgangsskalen) betrug für die deutschen Studierenden  $-.37$  und für die Gesamtgruppe  $-.21$ . Diese Werte sind jeweils hochsignifikant mit  $p < .000$ . Für die englischen Studierenden ist die Korrelation dagegen nicht signifikant. Zumindest für die deutschen Studierenden gilt damit erwartungsgemäß: „Je höher die Extraversion, desto geringer die Fremdsprachenverwendungsangst.“ Allerdings ist der Zusammenhang mit nur 13.5% gemeinsamer Varianz nicht sehr ausgeprägt. Zudem werden beide Variablen mit dem gleichen Verfahren gemessen (vierstufige Likert-Skala). Es ist deshalb nicht auszuschließen, dass die Korrelation aufgrund von methodeninduzierter gemeinsamer Varianz künstlich erhöht ist (vgl. zu diesem Problem auch Tremblay 2001: 244f). Die Möglichkeit eines Methodeneffekts gilt es auch bei der Interpretation der übrigen Korrelationen im vorliegenden Artikel zu beachten.

Die Korrelationen zwischen Fremdsprachenverwendungsangst und C-Test-Leistung sind ebenfalls hochsignifikant ( $p < .002$ ):  $-.36$  (Bochum);  $-.39$  (Duisburg);  $-.33$  (Portsmouth). Es gilt damit erwartungsgemäß: Je höher die Fremdsprachenverwendungsangst, desto geringer die fremdsprachliche Leistung. Wie wir im Weiteren sehen werden, entspricht auch die Beziehung zwischen Fremdsprachenverwendungsangst und den Skalen „Einstellung gegenüber den Sprechern der Zielsprache (L2)“ und „Sprachlernmotivation“ den Erwartungen.

### 5.3 Einstellungen

Da die Skalen „Einstellung gegenüber den Sprechern der eigenen Sprache (L1)“ und „Einstellung gegenüber den Sprechern der Zielsprache (L2)“ in ihrer ursprünglichen, 16 Items umfassenden Form keine akzeptablen Reliabilitäten aufwiesen, wurden im Gegensatz zu der Vorgehensweise bei den übrigen Skalen zunächst Faktorenanalysen durchgeführt – in der Hoffnung, die Messgenauigkeit und Interpretierbarkeit der Skalen zu erhöhen. Alle weiteren Berechnungen wurden dann auf der Basis der revidierten Skalen vorgenommen. Die Tabelle 3 zeigt die Resultate der Faktorenanalysen.

**Tabelle 3: Einstellungsskalen (L1 und L2)**  
**Resultate der Faktorenanalysen (Maximum Likelihood Methode, Varimax Rotation)**

<b>Gesamtgruppe</b>	
<b>a) Einstellung (L1)</b>	
nur 13 Items (Item 7, 13 und 15 wurden ausgeschlossen)	
3 Faktoren (20.4%, 11.1% und 10.5% der Gesamtvarianz); $r_{\text{e}} = .60$	
<b>1. Faktor: prosoziales Verhalten (Item 4, 5, 6, 9); <math>r_{\text{e}} = .79</math></b>	
Item 4:	„freundlich“ (.78)
Item 6:	„großzügig“ (.75)
<b>2. Faktor: egozentrisches Verhalten (Item 2, 10, 11, 16); <math>r_{\text{e}} = .60</math></b>	
Item 10:	„ungeschuldig“ (.60)
Item 11:	„stur“ (.56)
<b>3. Faktor: Kompetenz/Fleiß (Item 8, 12, 14); <math>r_{\text{e}} = .56</math></b>	
Item 12:	„leistungsfähig“ (.65)
Item 14:	„tüchtig“ (.73)
<b>b) Einstellung (L2)</b>	
nur 13 Items (Items 7, 13 und 15 wurden ausgeschlossen)	
3 Faktoren (14.5%, 13.0% und 11.0% der Gesamtvarianz); $r_{\text{e}} = .61$	
<b>1. Faktor: egozentrisches Verhalten (Item 2, 10, 11, 16); <math>r_{\text{e}} = .68</math></b>	
Item 10:	„ungeschuldig“ (.67)
Item 11:	„stur“ (.74)
<b>2. Faktor: prosoziales Verhalten (Item 4, 5, 6, 9); <math>r_{\text{e}} = .69</math></b>	
Item 4:	„freundlich“ (.68)
Item 6:	„großzügig“ (.65)
<b>3. Faktor: Kompetenz/Fleiß (Item 8, 12, 14); <math>r_{\text{e}} = .60</math></b>	
Item 12:	„leistungsfähig“ (.51)
Item 14:	„tüchtig“ (.72)

Wie die Tabelle 3 zeigt, ergeben sich nach Ausschluss von drei Items jeweils dreifaktorielle Lösungen, wobei die drei Faktoren im Fall der L1-Skala 42% und im Fall der L2-Skala 38.5% der Gesamtvarianz aufklären und sich zudem in ihrer relativen Größenordnung unterscheiden: Während bezogen auf die L1 „prosoziales Verhalten“ eindeutig am wichtigsten ist, scheint bezogen auf die L2 der Faktor „egozentrisches Verhalten“ die bedeutsamste Dimension zu sein – jedoch dicht gefolgt von „prosozialem Verhalten“. Zudem weisen die Items 1 und 3 nur sehr geringe Ladungen auf und können auch nicht eindeutig einem bestimmten Faktor zugeordnet werden. [7]

Auch nach dem Ausschluss der Items 7, 13 und 15 sind die Reliabilitäten sowohl der beiden Gesamtskalen als auch der Einzelskalen wenig zufrieden stellend und bestenfalls für gruppenbezogene Entscheidungen ausreichend. Auffallend ist zudem, dass die Einzelskalen in einigen Fällen reliabler als die Gesamtskalen sind, was als weiteres Indiz für die Mehrdimensionalität der untersuchten Items gewertet werden kann (vgl. z.B. Tremblay 2001: 242). In Anbetracht der geringen Reliabilitäten und der vermutlich mehrdimensionalen Struktur ist zu prüfen, ob die Einzelskalen nicht durch Hinzufügung weiterer Items reliabler gemacht werden können.

Tabelle 4 zeigt für die einzelnen Teilstichproben die Mittelwerte ( $M$ ) und Stichprobenumfänge ( $n$ ) der auf 13

Items reduzierten Einstellungsskalen (L1, L2). Die Mittelwerte unterscheiden sich nur wenig; die Reliabilitäten liegen zwischen .45 und .80. Angesichts der geringen Reliabilitäten wurde auf eine zufallskritische Überprüfung der beobachteten Mittelwertunterschiede verzichtet.

Tabelle 4: Einstellungsskalen (L1 und L2) (Spannweite: 13 – 52 Punkte)

Gruppe	Ort	Zielsprache	Niveau	L1		L2	
				n	M	n	M
1	Duisburg	Englisch	Studienanfänger	81	31.9	83	37.9
2	Bochum	Französisch	vorwiegend 1. Jahr	68	32.8	72	37.7
3	Portsmouth	Deutsch	2. Jahr	22	33.0	18	36.0
4	Portsmouth	Deutsch	4. Jahr	24	36.8	23	38.1
5	Portsmouth	Spanisch	2. Jahr	38	33.3	38	38.2
6	Portsmouth	Spanisch	4. Jahr	49	33.7	49	36.0
7	Portsmouth	Französisch	2. Jahr	15	32.8	15	37.3
8	Portsmouth	Französisch	4. Jahr	16	35.4	16	35.4
1 + 2	Deutschland	Englisch Französisch		149	32.3	155	37.8
3 – 8	Portsmouth	Deutsch, Franzö- sisch, Spanisch		164	34.1	159	36.9
Gesamt				313	33.2	314	37.3

Insgesamt gesehen muss angesichts der wenig zufrieden stellenden psychometrischen Eigenschaften von einer Benutzung der vorliegenden Einstellungsskalen zunächst einmal abgeraten werden. Allerdings haben die beiden Skalen sicherlich noch ein gewisses Verbesserungspotential. Zudem sollten die Resultate früherer Untersuchungen, in denen die hier benutzten Einstellungsskalen z.T. ohne eine Analyse der Reliabilität eingesetzt worden sind (vgl. die Literaturhinweise in [Abschnitt 2.2.3](#)), kritisch hinterfragt werden.

-12-

Trotz dieser Einschränkungen möchte ich abschließend die signifikanten Korrelationen der Einstellungsskalen mit den übrigen TESTATT-Instrumenten präsentieren. Die Daten sind in Tabelle 5 aufgeführt.

Tabelle 5: Signifikante Korrelationen zwischen den Einstellungsskalen und den übrigen TESTATT-Instrumenten ( $p < .05$ )

	deutsche Studierende	englische Studierende	Gesamt
Einstellung (L1) mit Einstellung (L2)	.181 n = 142; p = .031		
Einstellung (L1) mit C-Test (DELTA)	-.247 (Duisburg) n = 79; p = .028		
Einstellung (L2) mit Fremdspra- chenverwendungsangst	-.182 n = 148; p = .027	-.202 n = 156; p = .012	-.205 n = 304; p = .000
Einstellung (L2) mit Motivation	.291 n = 149; p = .000		
Einstellung (L2) mit C-Test			.147 n = 230; p = .026

Die Zusammenhänge zwischen den Einstellungsskalen und den anderen TESTATT-Instrumenten in Tabelle 5 entsprechen weitgehend den theoretisch und empirisch begründeten Erwartungen. Sie sind jedoch insgesamt gesehen nur sehr schwach – z.T. bedingt durch die geringe Reliabilität der Einstellungsskalen. Könnte man z.B. die Reliabilität der Einstellungsskala (L2) für die Gesamtgruppe von .61 auf .90 erhöhen, dann würde die Korrelation z.B. mit der Fremdsprachenverwendungsangst ( $r_{tt} = .93$ ) in der Gesamtgruppe auf  $-.249$  steigen

(vgl. zu dieser minderungskorrigierten Korrelation Lienert & Raatz 1994: 260).

Sieht man von der Möglichkeit eines Methodeneffekt auf Grund des identischen Itemformats einmal ab, dann ist bei der Interpretation der Interkorrelationen der TESTATT-Persönlichkeitsskalen weiterhin zu berücksichtigen, dass eine substantielle Korrelation auch ein Indiz dafür sein kann, dass sich die jeweiligen theoretischen Konstrukte und/oder deren Operationalisierungen nicht hinreichend voneinander unterscheiden. Vor diesem Hintergrund ist es zu begrüßen, dass die Korrelationen keine substantiellen Größenordnungen erreichen. Zum anderen ist es natürlich möglich, dass die Persönlichkeitsattribute trotz hinreichender konzeptueller Differenzierung dazu tendieren, jeweils ähnliche Ausprägungen anzunehmen. Auf jeden Fall verringert jedoch eine substantielle Interkorrelation die Effektivität der Merkmale als gemeinsame Prädiktoren der fremdsprachlichen Leistung (es sein denn, es handelt sich um Suppressor-Variablen).

Die Korrelationen in Tabelle 5 lassen sich u.a. folgendermaßen interpretieren:

Je positiver die Einstellung (L1),

- desto positiver auch die Einstellung (L2),
- desto geringer der C-Test-Punktwert (DELTA; nur Duisburg).

Je positiver die Einstellung (L2),

- desto geringer die Fremdsprachenverwendungsangst,
- desto höher die Motivation und,
- desto höher der C-Test-Punktwert.

#### 5.4 Sprachlernmotivation

Die Tabelle 6 zeigt die Resultate der Faktorenanalysen der 10 Items umfassenden Skala „Sprachlernmotivation“. Nach Ausschluss von Item 8, das eine Kommunalität von 0 aufwies, ergab sich in der Gesamtgruppe ein einziger Faktor, der insgesamt 48.2% Varianz aufklärt. Auch bei den englischen Studierenden (Gruppe 3 bis 8) ergab sich lediglich ein einziger relativ starker Faktor, der 57% an Varianz aufklärt. Die Reliabilitäten sind mit  $r_{tt} = .85$  und  $r_{tt} = .92$  (sehr) zufrieden stellend. Bei den deutschen Studierenden (Gruppe 1 + 2) konnten dagegen zwei als intrinsische/integrative Orientierung bzw. extrinsische/instrumentelle Orientierung interpretierbare Faktoren extrahiert werden, die allerdings zusammen nur 27.4% an Varianz aufklären. Es konnte somit lediglich für die Teilgruppe der deutschen Studierenden die mit der Zusammenstellung der Items angezielte zweidimensionale Skalenstruktur reproduziert werden. Dies steht im Gegensatz u.a. zur ELPS-Studie, wo die von Coleman (1996: Kap. 5) berechneten Faktorenanalysen auch bei englischsprachigen Studierenden eine mehrfaktorielle Motivationsstruktur ergeben haben. Allerdings misst die vorliegende Skala die Sprachlernmotivation der deutschen Studierenden nur relativ ungenau. Dies gilt insbesondere für die Teilskala „extrinsische/instrumentelle Orientierung“ ( $r_{tt} = .50$ ). Zudem ist die Korrelation zwischen den beiden Skalen mit .31 relativ gering und die Reliabilität der Gesamtskala deutlich höher als die Reliabilität der Einzelskalen. Beide Befunde können als Hinweis auf eine eher eindimensionale Struktur oder auch als Hinweis auf eine spezifische Antworttendenz auf Seiten der Befragten interpretiert werden (vgl. z.B. Brown 2001: 175f; Hinz, Brähler, Geyer & Körner 2003).

Insgesamt sprechen die Ergebnisse der Faktorenanalysen m.E. dafür, die vorliegende Motivationsskala als eindimensional zu behandeln, d.h. zunächst einmal nicht zwischen einer intrinsisch-integrativen Orientierung und einer extrinsisch-instrumentellen Orientierung zu differenzieren. Allerdings sollte geprüft werden, ob nicht durch Hinzunahme weiterer Items hinreichend reliable Subskalen konstruiert werden können und ob sich auch in diesem Fall bei deutschen und englischen Studierenden unterschiedliche Faktorstrukturen ergeben.

*Tabelle 6: Sprachlernmotivation*

*Resultate der Faktorenanalysen (Maximum Likelihood Methode, Varimax Rotation)*

**A) Gesamtgruppe**

2 Faktoren (24.8% und 22.3 der Gesamtvarianz);  $r_{\alpha} = .85$ .

Das Item 8

„...weil Sie mehr Respekt erhalten, wenn Sie verschiedene Fremdsprachen sprechen.“

hat eine Kommunalität von 0. Ohne dieses Item:

**1 Faktor (48.2% der Gesamtvarianz);  $r_{\alpha} = .88$**

Entscheidung: Ausschluss von Item 8 aus der Skala und aus den weiteren Berechnungen

**B) Gruppe 1 und 2: deutsche Studenten**

2 Faktoren (16.6% und 10.8% der Gesamtvarianz);  $r_{\alpha} = .64$

**1. Faktor: intrinsische/integrative Orientierung (Items 4, 6, 7, 9, 10);  $r_{\alpha} = .59$**

Item 10: „...weil Sie gerne in dem Land leben möchten, wo diese Sprache gesprochen wird.“ (.59)

Item 4: „...um das Land und die Lebensweisen, wo diese Sprache gesprochen wird, besser zu verstehen.“ (.59)

**2. Faktor: extrinsische/instrumentelle Orientierung (Item 1, 2, 3, 5);  $r_{\alpha} = .50$**

Item 3: „...um eine höhere Bildung zu erlangen.“ (.48)

Item 1: „...für Ihren zukünftigen Beruf.“ (.46)

Interkorrelation der Skalen: .31

Es gibt nur minimale Unterschiede zwischen den deutschen Teilstichproben.

**C) Gruppe 3 bis 8: englische Studenten**

**1 Faktor (57% der Gesamtvarianz);  $r_{\alpha} = .92$**

Es gibt nur minimale Unterschiede zwischen den englischen Teilstichproben.

Tabelle 7 zeigt die Mittelwerte, Standardabweichungen und Reliabilitäten der Motivationskala für die verschiedenen Teilstichproben. Es handelt sich dabei um die Ausgangsskala mit 10 Items. Der Mittelwertunterschied zwischen den deutschen ( $n = 167$ ) und den englischen Studierenden ( $n = 165$ ) ist hochsignifikant ( $p < .00$ ). Die Mittelwertunterschiede innerhalb der Gruppe der deutschen und der englischen Studierenden liegen dagegen im Zufallsbereich. Interessant ist die Tatsache, dass die deutschen Studierenden deutlich höhere Motivationswerte als die englischen Studierenden aufweisen, was möglicherweise den tendenziell geringeren Stellenwert von Fremdsprachenkenntnissen bei Lernern mit Englisch als Muttersprache widerspiegelt.

*Tabelle 7: Sprachlernmotivation (Spannweite der Skala: 10 – 40 Punkte)*

Gruppe	Ort	Zielsprache	Niveau	n	$r_{tt}$	M	SD
1	Duisburg	Englisch	Studienanfänger	91	.70	21.6	4.3
2	Bochum	Französisch	vorwiegend 1. Jahr	76	.58	21.8	3.6
3	Portsmouth	Deutsch	2. Jahr	24	.84	18.1	6.7
4	Portsmouth	Deutsch	4. Jahr	23	.89	16.6	7.9
5	Portsmouth	Spanisch	2. Jahr	38	.91	18.8	8.3
6	Portsmouth	Spanisch	4. Jahr	49	.89	15.0	7.9
7	Portsmouth	Französisch	2. Jahr	15	.90	14.5	8.9
8	Portsmouth	Französisch	4. Jahr	16	.85	17.2	7.4
1 + 2	Deutschland	E, F	1. Jahr	167	.64	21.7	4.0
3 bis 8	Portsmouth	D, F, S	2. und 4. Jahr	165	.89	16.7	8.0

In der Tabelle 8 sind die signifikanten Korrelationen zwischen der Motivationskala und den übrigen TESTATT-Instrumenten aufgeführt. Die Berechnungen basieren wie bereits in Tabelle 7 auf der nicht reduzierten Ausgangsskala.

**Tabelle 8: Signifikante Korrelationen zwischen der Sprachlernmotivationskala und den übrigen TESTATT-Instrumenten ( $p < .05$ )**

	deutsche Studierende	englische Studierende	Gesamt
Motivation mit Extraversion	.194 n = 153; p = .016	.258 n = 157; p = .001	.175 n = 310; p = .002
Motivation mit C-Test: Bochum	.283 n = 75; p = .014		
Motivation mit C-Test: Duisburg	.200 n = 89; p = .060		
Motivation mit Fremdsprachenverwendungsangst	-.164 n = 155; p = .042		
Motivation mit Einstellung (L2)	.291 n = 149; p = .000		

Die Zusammenhänge zwischen der Motivationskala und den anderen TESTATT-Instrumenten in Tabelle 8 sind insgesamt gesehen nur schwach; sie entsprechen jedoch völlig den theoretisch und empirisch begründeten Erwartungen. Interpretiert man den Zusammenhang zwischen Motivation und C-Test-Leistung in der Gruppe der Duisburger Studierenden im Sinne einer gerichteten Hypothese, dann ist auch diese Beziehung signifikant ( $p = .03$ ).

-15-

Die Korrelationen lassen sich folgendermaßen interpretieren:

Je höher die Motivation,

- desto höher die Extraversion,
- desto höher der C-Test-Punktwert (nur deutsche Studierende),
- desto geringer die Fremdsprachenverwendungsangst,
- desto positiver die Einstellung (L2).

Auffallend ist wiederum, dass bei den englischen Studierenden lediglich die Korrelation zwischen Motivation und Extraversion einen signifikanten Wert erreicht. Auch dieser Befund zeigt, dass sich deutsche und englische Studierende offensichtlich in der Struktur des Konstrukts „Sprachlernmotivation“ unterscheiden.

## 6 Zusammenfassung und Perspektiven

Es konnten mit Hilfe der Faktorenanalysen fünf Skalen definiert werden: Eine Extraversion/Introversion-Skala mit 15 Items, eine Skala der Fremdsprachenverwendungsangst mit 11 Items, eine Skala für die Einstellung gegenüber den Sprechern der eigenen Sprache mit 13 Items, eine Skala für die Einstellung gegenüber den Sprechern der Zielsprache mit ebenfalls 13 Items und eine Sprachlernmotivations-Skala mit 9 Items. Es konnten Subskalen identifiziert werden und zu bereits aus der Literatur bekannten Konstrukten in Beziehung gesetzt werden. Allerdings waren die Resultate nicht in allen Fällen eindeutig. So ergab sich z.B. bei der Sprachlernmotivation je nach der Gruppe der Studierenden entweder eine einfaktorielle oder eine zweifaktorielle Lösung. Auch die Reliabilitäten erwiesen sich nicht in allen Fällen als zufrieden stellend. Letzteres gilt insbesondere in Bezug auf die beiden Einstellungsskalen.

Insgesamt gesehen erscheint der Schluss gerechtfertigt, dass das Hauptziel des TESTATT-Projekts, nämlich die Entwicklung ökonomischer Messinstrumente im Hinblick auf den studienbezogenen Auslandsaufenthalt, weitgehend erreicht ist.

Allerdings muss diese Aussage aus einer Reihe von Gründen relativiert werden:

- Die Untersuchungsstichproben waren nicht repräsentativ,
- die Stichprobenumfänge waren recht gering,
- es wurden lediglich bivariate Korrelationen berechnet ohne Berücksichtigung möglicher komplexer Interdependenzen,
- es konnten – u.a. wegen der geringen Stichprobenumfänge – nur bedingt die für die jeweilige Fragestellung optimalen statistischen Verfahren verwendet werden.

Deshalb ist es notwendig,

- die Instrumente in größerem Umfang im Rahmen geeigneter Untersuchungen einzusetzen,
- die sich ergebenden Daten zu sammeln und
- auf dieser Datenbasis und mit Hilfe komplexerer Methoden die gezogenen tentativen Schlussfolgerungen zu überprüfen. [8]

---

-16-

Hierbei sind jedoch eine Reihe von wichtigen methodologischen Aspekten zu beachten. So setzen z.B. die im vorliegenden Aufsatz berechneten Pearson Produkt-Moment-Korrelationen (streng genommen) Intervallskalenniveau voraus. Bei den vorliegenden vierstufigen Likert-Skalen handelt es sich jedoch um Ordinalskalen. Olsson, Drasgo & Dorans (1982) stellen in diesem Zusammenhang fest:

“...product moment correlations computed from categorical variables can be seriously misleading. This is particularly evident when rating scale items, with varying number of categories and varying degrees of skewness, are used to assess substantively important constructs. Interitem correlations and item-total correlations can be seriously distorted by discrete measurement of latent continuous variables.” (346f) [9]

Zudem setzt der Signifikanztest der Maximum-Likelihood-Faktorenanalyse zur Bestimmung der Zahl der Faktoren eine multivariate Normalverteilung voraus. Diese Annahme ist jedoch bei der vorliegenden Skalenstruktur nicht erfüllt. Auch im Fall der Summenwerte ist zu fragen, inwieweit die Annahme von Intervallskalen sowie einer bivariaten Normalverteilung (beim Test der Signifikanz der Korrelationen) erfüllt ist. Zur Überprüfung der Dimensionalität der Likert-skalierten Items sollten deshalb anstelle von Pearson Produkt-Moment-Korrelationen die von Olsson (1979) vorgeschlagenen polychorischen Korrelationen berechnet werden – z.B. mit Hilfe der PRELIS2-Software im Rahmen von LISREL (vgl. auch Schumacker & Lomax 1996: 18; 104f). Die resultierende asymptotische Kovarianzmatrix kann dann z.B. mit Hilfe der Methode der gewichteten kleinsten Quadrate weiter analysiert werden, die asymptotisch-verteilungsfreie Schätzer liefert und damit keine multivariate Normalverteilung voraussetzt. Allerdings benötigt man für die Berechnung der Schätzer hohe Stichprobenumfänge (es müssen Momente vierter Ordnung ermittelt werden), weshalb das Verfahren im vorliegenden Fall nicht angewendet werden konnte (vgl. zum Problem des Stichprobenumfangs und asymptotisch verteilungsfreier Schätzer z.B. Backhaus, Erichson, Plinke & Weiber 2003: 264f; Browne 1984; Curran, West & Finch 1996; Kunnan 1998: 300; Satorra 1990). Entsprechend könnte auch im Fall der Summenscores vorgegangen werden, wobei im Fall der Korrelation zwischen C-Test-Summenscore und Persönlichkeitssummenscore polyserielle Korrelationen berechnet werden sollten (vgl. Olsson, Drasgow & Dorans 1982; Schumacker & Lomax 1996: 18; 104f).

Die Berechnung von Produkt-Moment-Korrelationen im Fall von Likert-Skalen und die unzureichende Beachtung der Voraussetzung einer (multivariaten) Normalverteilung bei bestimmten Varianten der Faktorenanalyse oder auch von Strukturgleichungsmodellen wie LISREL ist im Übrigen in der empirischen Fremdsprachenforschung weithin üblich. Dies gilt mit Einschränkungen z.B. auch für die kanadischen Untersuchungen im Rahmen des “socio-educational model of second language learning” von Gardner und anderen Forschern.

Weiterhin sollten die Skalen und Tests nicht nur auf der Basis der klassischen Testtheorie, sondern auch mit Hilfe probabilistischer Messmodelle (Item-Response-Theorie) analysiert werden (zu den probabilistischen Messmodellen vgl. z.B. Bond & Fox 2001; Embretson & Reise 2000; Rost 1996; Rost, Carstensen & von Davier 1999; Steyer & Eid 2001). In der fremdsprachenerwerbsspezifischen Persönlichkeitsforschung gibt es bisher kaum entsprechende Arbeiten. Für die Analyse von ordinalskalierten Daten existiert eine Reihe verschiedener Modelle, wie z.B. das *Partial-Credit-Modell*, das *diskrete Ratingskalen-Modell* und das *kontinuierliche Ratingskalen-Modell*. Ein Vergleich der drei genannten Modelle bezogen auf die Skalierung von C-Tests findet sich bei Eckes (2004). Weitere Item-Response-Modelle für geordnete Antwortkategorien werden z.B. von Embretson & Reise (2000: Kap. 5) und Thissen, Nelson, Rosa & McLeod (2001) behandelt.

Eine Analyse von Skalen und Tests auf der Basis der klassischen Testtheorie und probabilistischen Messmodellen führt zwar häufig zu vergleichbaren Resultaten (vgl. z.B. Fan 1998). Letztere bieten jedoch



neben Vorteilen bei der Interpretation der Ergebnisse zusätzliche Informationen, wie z.B. Angaben zur individuenspezifischen Messgenauigkeit in bestimmten Skalenbereichen, Informationen zur optimalen Zahl der Antwortkategorien, Hinweise auf unskalierbare Personen oder auch auf Personengruppen mit bestimmten Urteilstendenzen (vgl. auch Andrich, de Jong & Sheridan 1997; Nachtigall, Kroehne, Funke & Steyer 2003; Reise & Henson 2003; Rost, Carstensen & von Davier 1999). Es ist geplant, diese Vorteile bei der Weiterentwicklung der TESTATT-Instrumente zu nutzen.

### Anmerkungen

- [1] Teile der Abschnitte 1-4 des vorliegenden Beitrags stellen eine gekürzte und überarbeitete Fassung der entsprechenden Abschnitte aus Grotjahn, Raatz & Wockenfuß (2004) dar.
- [2] Projektleitung der ELPS-Studie sowie des Anschlussprojekts „TESTATT“: Jim Coleman (University of Portsmouth), Rüdiger Grotjahn (Ruhr-Universität Bochum), Christine Klein-Braley und Ulrich Raatz (Universität Duisburg). Beide Projekte wurden vom British Council und dem DAAD finanziell unterstützt. Einen detaillierten Überblick über die ELPS-Studie gibt Coleman (1996).
- [3] In der ELPS-Studie wurde z.T. nur eine einzige Frage zu einem Konstrukt gestellt.
- [4] Hinweise zu multivariaten Beziehungen zwischen Variablen im Bereich des Lernens und Gebrauchs von Fremd- und Zweitsprachen finden sich z.B. in Gardners "socio-educational model of languages learning" und MacIntyres Model der Bereitschaft zur Kommunikation (*willingness to communicate*) (vgl. z.B. Gardner, Tremblay & Masgoret 1997; MacIntyre & Charos 1996; MacIntyre, Baker, Clément & Donovan 2002; MacIntyre, Clément, Dörnyei, & Noels 1998 sowie auch den [Beitrag von McIntosh & Noels 2004](#) in dieser Nummer).
- [5] Das EPI scheint sowohl in der Psychologie als auch in der Fremdsprachenforschung am häufigsten zur Messung von Extraversion/Introversion eingesetzt worden zu sein. Zu weiteren in der Fremdsprachenforschung verwendeten Verfahren zur Messung von Extraversion/Introversion vgl. Grotjahn, Raatz & Wockenfuß (2004).
- [6] Die Mummendey-Skala wurde auch von Müßig-Trapp & Schnitzer (1997) zu Grunde gelegt. Die Autoren haben mit Hilfe von Cluster- und Faktorenanalysen die Skala auf 17 Items mit fünf Antwortstufen reduziert. Die reduzierte Skala erwies sich als hoch reliabel (Cronbachs Alpha = 0.84).
- [7] Ein Grund für das idiosynkratische Verhalten von Item 1 könnte sprachlicher Natur sein. Für die deutsche Version des Fragebogens wurde das englische Adjektiv "emotional" mit „gefühlvoll“ übersetzt. Gefühlvoll hat jedoch für manche deutsche Muttersprachler eine in Richtung auf „rührselig“ gehende Konnotation. Adäquatere Übersetzungsäquivalente wären vermutlich „gefühlsbetont“, „empfindsam“ oder auch „emotional“. Es ist zu überprüfen, ob eine entsprechende Übersetzung zu einer Verbesserung der psychometrischen Eigenschaften von Item 1 führt.
- [8] Aus diesem Grunde werden die vorgestellten Skalen und C-Tests auch nur mit Einschränkungen für den wissenschaftlichen Gebrauch freigegeben. Fremdsprachenforscher, Linguisten, Pädagogen, Psychologen oder sonstige wissenschaftlich interessierte Kolleginnen und Kollegen können die Instrumente unter Angabe von Institution und Verwendungszweck und mit der Verpflichtung, die Testergebnisse in anonymisierter Form zurückzumelden, über folgende e-Mail-Adresse anfordern: [Raatz@uni-duisburg.de](mailto:Raatz@uni-duisburg.de) (Prof. Dr. Ulrich Raatz). Das TESTATT-Team hofft, dass auf diese Weise so viele Daten zusammenkommen, dass in naher Zukunft gesichertere und differenziertere Aussagen gemacht werden können.
- [9] Vgl. auch Olsson (1979: 443): "... application of factor analysis to discrete data may lead to incorrect conclusions regarding the number of factors, and to biased estimates of the factor loadings, especially when the distributions of the observed variables are skewed in opposite directions. This is mainly due to biased estimates of the correlations. Thus, there seems to be some need for correlation estimates which are more viable when the observed data are ordinal with only a few steps." Zur Analyse von Likert-Skalen vgl. auch Muthén & Kaplan (1992).

### Literaturverzeichnis

- Amelang, Manfred & Bartussek, Dieter. (1997). *Differentielle Psychologie und Persönlichkeitsforschung* (4. überarb. u. erw. Aufl.). Stuttgart: Kohlhammer.
- Andrich, David, de Jong, John H. A. L. & Sheridan, Barry A. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In Jürgen Rost & Rolf Langeheine (Hrsg.), *Applications of latent*

*trait and latent class models in the social sciences* (S. 59-70). Münster: Waxmann.

- Backhaus, Klaus, Erichson, Bernd, Plinke, Wulff & Weiber, Rolf. (2003). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (10., neu bearb. und erw. Aufl.). Berlin: Springer-Verlag.
- Bem, Daryl J. & Allen, Andrea. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review*, 81(6), 505-520.
- Bond, Trevor G. & Fox, Christine M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.
- Borkenau, Peter & Ostendorf, Fritz. (1993). *NEO-Fünf-Faktoren Inventar nach Costa und McCrae: Handanweisung*. Göttingen: Hogrefe.
- Brown, James D. (2001). *Using surveys in language learning*. Cambridge: Cambridge University Press.
- Brown, James D., Robson, Gordon & Rosenkjar, Patrick R. (2001). Personality, motivation, anxiety, strategies, and language proficiency of Japanese students. In Zoltán Dörnyei & Richard W. Schmidt (Hrsg.), *Motivation and second language acquisition* (S. 361-398). Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Browne, Michael W. (1984). Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 127-141.
- Coleman, James A. (1996). *Studying languages: A survey of British and European students. The proficiency, background, attitudes and motivations of students of foreign languages in the United Kingdom and Europe*. London: CILT.
- Coleman, James A. (1997). Residence abroad within language study. *Language Teaching*, 30, 1-20.
- Curran, Patrick J., West, Stephen G. & Finch, John F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16-29.
- Dewaele, Jean-Marc & Furnham, Adrian. (1999). Extraversion: The unloved variable in applied linguistics research. *Language Learning*, 49(3), 509-544.

---

-19-

- Dörnyei, Zoltán. (2001). *Teaching and researching motivation*. Harlow, England: Longman.
- Dörnyei, Zoltán. (Hrsg.). (2003). *Attitudes, orientations, and motivations in language learning*. Malden, MA: Blackwell.
- Dörnyei, Zoltán & Schmidt, Richard W. (Hrsg.). (2001). *Motivation and second language acquisition*. Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Eckes, Thomas. (2004). Rasch-Modelle zur C-Test-Skalierung. In Rüdiger Grotjahn (Hrsg.), *The C-Test: Theory, empirical research, applications*. Frankfurt am Main: Lang. [im Druck]
- Eggert, Dietrich. (1983). *Eysenck-Persönlichkeits-Inventar (E-P-I)*. Göttingen: Hogrefe.
- Embretson, Susan E. & Reise, Steven. (2000). *Item response theory for psychologists*. Hillsdale, NJ: Erlbaum.
- Fan, Xitao. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58(3), 357-381.
- Finkbeiner, Claudia. (2001). Zur Erforschung attitudinaler und affektiver Faktoren beim Lehren und Lernen fremder Sprachen. In Claudia Finkbeiner & Gerhard W. Schnaitmann (Hrsg.), *Lehren und Lernen im Kontext empirischer Forschung und Fachdidaktik* (S. 352-375). Donauwörth: Auer.
- Ganschow, Leonore & Sparks, Richard. (2001). Learning difficulties and foreign language learning: A review of research and instruction. *Language Teaching*, 34(1), 79-98.
- Gardner, Robert C. (1985). *Social psychology and language learning: The role of attitudes and motivation*. London: Arnold.
- Gardner, Robert C., Kirby, D. M., Gorospe, F. H. & Villamin, A. C. (1972). Ethnic stereotypes: An alternative assessment technique, the stereotype differential. *The Journal of Social Psychology*, 87, 259-267.
- Gardner, Robert C. & Lambert, Wallace E. (1972). *Attitudes and motivation in second-language learning*. Rowley, Mass.: Newbury House.
- Gardner, Robert C., Tremblay, Paul F. & Masgoret, Anne-Marie. (1997). Towards a full model of second language learning: An empirical investigation. *Modern Language Journal*, 81, 344-362.

- Grotjahn, Rüdiger. (1986). Der Bochumer Einstufungstest ‚Französisch‘. In Seminar für Sprachlehrforschung der Ruhr-Universität Bochum (Hrsg.), *Probleme und Perspektiven der Sprachlehrforschung. Bochumer Beiträge zum Fremdsprachenunterricht in Forschung und Lehre* (S. 313-324). Frankfurt am Main: Scriptor.
- Grotjahn, Rüdiger. (1995). Der C-Test: State of the Art. *Zeitschrift für Fremdsprachenforschung*, 6(2), 37-60.
- Grotjahn, Rüdiger. (Hrsg.). (2004). *The C-test: Theory, empirical research, applications*. Frankfurt am Main: Lang. [im Druck]
- Grotjahn, Rüdiger, Raatz, Ulrich & Wockenfuß, Verena. (2004). Das Projekt "Tests and Attitude Scales for the Year Abroad" (TESTATT): Theoretische Basis und einige empirische Resultate. In Wolfgang Börner & Klaus Vogel (Hrsg.), *Emotion und Kognition im Fremdsprachenunterricht*. Tübingen: Narr. [im Druck]

---

-20-

- Grotjahn, Rüdiger, Raatz, Ulrich, Wockenfuß, Verena & Coleman, James A. (2001). *New tools to investigate student residence abroad*. Paper presented at the CILT Research Seminar: New research into residence abroad (June 11, 2001), London.
- Hambleton, Ronald K. & de Jong, John H. A. L. (Hrsg.). (2003). Advances in translating and adapting educational and psychological tests [Special Issue]. *Language Testing*, 20(2).
- Hinz, Andreas, Brähler, Elmar, Geyer, Michael & Körner, Annett. (2003). Urteileffekte beim NEO-FFI. *Diagnostica*, 49(4), 157-163.
- Horwitz, Elaine K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, 21, 112-126.
- Horwitz, Elaine K., Horwitz, Michael B. & Cope, Jo A. (1991/1986). Foreign language classroom anxiety. In Elaine K. Horwitz & Dolly J. Young (Hrsg.), *Language anxiety: From theory and research to classroom implications* (S. 27-36). Englewood Cliffs, NJ: Prentice Hall [reprinted from: *The Modern Language Journal*, 70, 1986, 125-132].
- Kiany, G. Reza. (1997). Personality and language learning: The contradiction between psychologists and applied linguists. *I.T.L. Review of Applied Linguistics*, 115-116, 111-136.
- Kleppin, Karin. (2001). Motivation. Nur ein Mythos? (I). *Deutsch als Fremdsprache*, 38(4), 219-225.
- Kleppin, Karin. (2002). Motivation. Nur ein Mythos? (II). *Deutsch als Fremdsprache*, 39(1), 26-30.
- Kunnan, Antony J. (1998). An introduction to structural equation modelling for language assessment research. *Language Testing*, 15(3), 295-332.
- Lienert, Gustav A. & Raatz, Ulrich. (1994). *Testaufbau und Testanalyse* (5., völlig Neubearb. und erw. Aufl.). Weinheim: Beltz, Psychologie Verlags Union.
- MacIntyre, Peter D. (1999). Language anxiety: A review of the research for language teachers. In Dolly J. Young (Hrsg.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere* (S. 24-45). Boston: McGraw-Hill.
- MacIntyre, Peter D., Baker, Susan C., Clément, Richard & Donovan, Leslie A. (2002). Sex and age effects on willingness to communicate, anxiety, perceived competence, and L2 motivation among junior high school French immersion students. *Language Learning*, 52(3), 537-564.
- MacIntyre, Peter D. & Charos, Catherine. (1996). Personality, attitudes, and affect as predictors of second language communication. *Journal of Language and Social Psychology*, 15(1), 3-26.
- MacIntyre, Peter D., Clément, Richard, Dörnyei, Zoltán & Noels, Kimberly A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545-562.

---

-21-

- Masgoret, Anne-Marie & Gardner, Robert C. (2003). Attitudes, motivation, and second language learning: A meta analysis of studies conducted by Gardner and associates. *Language Learning*, 53(1), 123-163.
- Matthews, Gerald & Deary, Ian J. (1998). *Personality traits*. Cambridge: Cambridge University Press.
- McIntosh, Cameron N. & Noels, Kimberly A. (2004). Self-determined motivation for language learning: The role of need for cognition and language learning strategies. *Zeitschrift für Interkulturellen*

*Fremdsprachenunterricht [online]*, 9(2) [<http://www.ualberta.ca/~german/ejournal/Mcintosh2.htm>].

- Mummendey, Hans D. (1995). *Die Fragebogen-Methode: Grundlagen und Anwendung in Persönlichkeits-, Einstellungs- und Selbstkonzeptforschung* (2., korr. Aufl.; 3., unveränd. Aufl. 1999). Göttingen: Hogrefe, Verlag für Psychologie.
- Müßig-Trapp, Peter & Schnitzer, Klaus. (1997). *Vorbereitung auf Europa durch Mobilität und Internationalisierung des Studiums*. Bonn: Bundesministerium für Bildung und Forschung [auch abrufbar unter <http://www.his.de/Abt2/Auslandsstudium/Mobilitaet/INDEX.HTM>].
- Muthén, Bengt O. & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45, 19-30.
- Nachtigall, Christof, Kroehne, Ulf, Funke, Friedrich & Steyer, Rolf. (2003). (Why) should we use SEM? Pros and cons of Structural Equation Modeling. *Methods of Psychological Research Online*, 8(2), 1-22 [<http://www.mpr-online.de>].
- Noels, Kimberly A., Pelletier, Luc G., Clément, Richard & Vallerand, Robert J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning*, 50(1), 57-85.
- Olsson, Ulf. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4), 443-460.
- Olsson, Ulf, Drasgow, Fritz & Dorans, Neil J. (1982). The polyserial correlation coefficient. *Psychometrika*, 47(3), 337-347.
- Onwuegbuzie, Anthony J., Bailey, Phillip & Daley, Christine E. (2000). The validation of three scales measuring anxiety at different stages of the foreign language learning process: The Input Anxiety Scale, the Processing Anxiety Scale, and the Output Anxiety Scale. *Language Learning*, 50(1), 87-117.
- Osgood, Charles E., Suci, George J. & Tannenbaum, Percy H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Pickett, Lorraine. (1993). *The effects of cultural distancing on attitudes and motivation in foreign language learning*. Unpublished M.A. thesis, University of Portsmouth.

---

-22-

- Raatz, Ulrich, Grotjahn, Rüdiger & Wockenfuß, Verena. (2004). Das TESTATT-Projekt: Entwicklung von C-Tests zur Evaluation des Fremdsprachenstudiums. In Rüdiger Grotjahn (Hrsg.), *The C-test: Theory, empirical research, applications*. Frankfurt am Main: Lang. [im Druck]
- Reise, Steven P. & Henson, James M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment*, 81(2), 93-103.
- Riemer, Claudia. (2001). Zur Rolle der Motivation beim Fremdsprachenlernen. In Claudia Finkbeiner & Gerhard W. Schnaitmann (Hrsg.), *Lehren und Lernen im Kontext empirischer Forschung und Fachdidaktik* (S. 376-398). Donauwörth: Auer.
- Rodríguez, Máximo & Abreu, Orángel. (2003). The stability of general foreign language classroom anxiety across English and French. *The Modern Language Journal*, 87(3), 365-374.
- Rost, Jürgen. (1996). *Lehrbuch Testtheorie, Testkonstruktion*. Bern: Huber.
- Rost, Jürgen, Carstensen, Claus H. & von Davier, Matthias. (1999). Sind die Big Five Rasch-skalierbar? Eine Reanalyse der NEO-FFI-Normierungsdaten. *Diagnostica*, 45(3), 119-127.
- Satorra, Albert. (1990). Robustness issues in structural equation modelling: A review of recent developments. *Quality & Quantity*, 24, 367-386.
- Schumacker, Randall E. & Lomax, Richard G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Erlbaum.
- Schuman, Howard & Presser, Stanley. (1996). *Questions and answers in attitude surveys*. London: Sage.
- Schwibbe, Gudrun & Schwibbe, Michael. (1987). Validierungsuntersuchungen zum Duisburg English Language Test for Advanced Students (DELTA). In Rüdiger Grotjahn, Christine Klein-Braley & Douglas K. Stevenson (Hrsg.), *Taking their measure: The validity and validation of language tests* (S. 255-274). Bochum: Brockmeyer.
- Smit, Arnold, Kelderman, Henk & van der Flier, Henk. (2003). Latent trait latent class analysis of an Eysenck

personality questionnaire. *Methods of Psychological Research Online*, 8(3), 23-50 [<http://www.mpr-online.de>].

Sparks, Richard & Ganschow, Leonore. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics*, 21, 90-111.

Stevens, James. (1996). *Applied multivariate statistics for the social sciences* (3. Aufl.). Mahwah, NJ: Erlbaum.

Steyer, Rolf & Eid, Michael. (2001). *Messen und Testen* (2. Aufl.). Berlin: Springer-Verlag.

-23-

Stiensmeier-Pelster, Joachim & Rheinberg, Falko. (Hrsg.). (2003). *Diagnostik von Motivation und Selbstkonzept*. Göttingen: Hogrefe [= Tests und Trends. Jahrbuch der pädagogischen Diagnostik. N.F., Bd. 2].

Tabachnick, Barbara G. & Fidell, Linda S. (1989). *Using multivariate statistics* (2. Aufl.). Cambridge: Harper & Row.

Thissen, David, Nelson, Lauren, Rosa, Kathleen & McLeod, Lori D. (2001). Item response theory for items scored in more than two categories. In David Thissen & Howard Wainer (Hrsg.), *Test scoring* (S. 141-186). Mahwah, NJ: Erlbaum.

Tremblay, Paul F. (2001). Research in second language motivation: Psychometric and research design considerations. In Zoltán Dörnyei & Richard W. Schmidt (Hrsg.), *Motivation and second language acquisition* (S. 239-255). Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.

Verhoeven, Ludo & Vermeer, Anne. (2002). Communicative competence and personality dimensions in first and second language learners. *Applied Psycholinguistics*, 23(3), 361-374.

Willis, Frank M., Doble, Gordon, Sankarayya, Umasnkar & Smithers, Alan. (1977). *Residence abroad and the study of modern languages*. University of Bradford: Modern Language Centre.

Wood, James M., Tataryn, Douglas J. & Gorsuch, Richard L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, 1(4), 354-365.

**Anhang 1: Einstellungsskala (L1)** im [Word-Format](#) oder [PDF-Format](#)

**Anhang 2: Sprachlernmotivation** im [Word-Format](#) oder [PDF-Format](#)

Copyright © 2004 *Zeitschrift für Interkulturellen Fremdsprachenunterricht*

**Grotjahn, Rüdiger.** (2004). "Tests and Attitude Scales for the Year Abroad" (TESTATT): Sprachlernmotivation und Einstellungen gegenüber Sprechern der eigenen und der fremden Sprache. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* [Online], 9(2), 23 pp. Erhältlich unter <http://www.ualberta.ca/~german/ejournal/Grotjahn2.htm>

[Zurück zur [Leitseite](#)]