

- Tesch, Bernd; Leupold, Eynar & Köller, Olaf (Hrsg.) (2008), *Bildungsstandards Französisch: konkret. Sekundarstufe I: Grundlagen, Aufgabenbeispiele und Unterrichts Anregungen*. Cornelsen Scriptor: Berlin.
- Vollmer, Helmut J. (2006), Bildungsstandards von oben - Bildungsstandards von unten. In: *Der Fremdsprachliche Unterricht*. Englisch, 40 (2006) 81, 12-16.
- Vollmer, Helmut J. (2006a), Kompetenzaufgaben als Forschungs- und Evaluationsinstrument. Statement zur 26. Frühjahrskonferenz zur Erforschung des Fremdsprachensunterrichts. In: Bausch, Karl-Richard; Burwitz-Melzer, Eva; Königs, Frank G. & Krumm, Hans-Jürgen (Hrsg.), *Aufgabenorientierung als Aufgabe*. Tübingen: Narr, 244-255.
- Weinert, Franz E. (1999), *Concepts of competence (Contribution within the OECD project Definition and selection of competencies: Theoretical and conceptual foundations (DeSeCo))*. Neuchâtel: DeSeCo.
- Weinert, Franz E. (2001), Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In: Weinert, Franz E. (Hrsg.), *Leistungsmessungen in Schulen*. Weinheim und Basel: Beltz Verlag, 17-31.
- Zydati, Wolfgang (2005), *Bildungsstandards und Kompetenzniveaus im Englischunterricht. Konzepte, Empirie, Kritik und Konsequenzen*. Frankfurt/M.: Peter Lang.
- Zydati, Wolfgang (2007a), „was Sie über die Bildungsstandards in Deutschland wissen sollten...“. In: *Babylonia* 15 (4/2007), 15-18.
- Zydati, Wolfgang (2007b), *Deutsch-Englische Züge in Berlin (DEZIBEL). Eine Evaluation des bilingualen Sachfachunterrichts an Gymnasien*. Frankfurt/M.: Peter Lang.
- Zydati, Wolfgang (2008), SMS an KMK: Standards mit Substanz! Kulturelle Inhalte, Mediation zwischen Sprachsystem und Sprachhandeln, Kritikfähigkeit – auch im Fremdsprachenunterricht. *Beiträge zur Fremdsprachenvermittlung*, Sonderheft 13 (2008), 13-34.

## Vernünftige Erwartungen: Referenzrahmen, Kompetenzniveaus, Bildungsstandards

Erwin Tschirner<sup>1</sup>

In 2003, Germany instituted National Educational Standards (*Bildungsstandards*) for the lower secondary school system (grades 9 and 10) and in 2007 it decided also to develop National Standards for the upper secondary school system (*Abitur*). For foreign languages, both sets of standards are or will be based to a large extent on the Common European Framework of Reference for Languages. This paper focuses on expected proficiency levels in speaking. The only large-scale study to date establishing proficiency levels at school is the DESI study which looked at scholastic achievement in German and English in grade 9 across all school types. The test used to evaluate speaking proficiency in English was the *PhonePass Set-10* test. This paper discusses the validity of the results obtained and questions the rationality of some of the speaking proficiency expectations for foreign languages at German secondary schools.

Der Gemeinsame europäische Referenzrahmen (GeR) wird mehr und mehr zum Maßstab von zu erreichenden Fremdsprachenkompetenzen und fungiert zunehmend im Sinne von Bildungsstandards. Doch welche GeR-Niveaus lassen sich mit dem schulischen Fremdsprachenunterricht realistisch erreichen? Hierzu liegen kaum belastbare Erfahrungen vor. Die DESI-Studie (Schröder, Harsch & Nold 2006) erhebt den Anspruch, für die 9. Klasse zumindest für die Sprechkompetenz diese Ergebnisse geliefert zu haben, ob sie das wirklich geleistet hat, soll ein Thema dieses Beitrags sein. Insgesamt scheinen Lehrpläne im fremdsprachlichen Bereich oft immer noch eher von Wunschen geprägt zu sein. Das Ziel B2 in allen Fertigkeiten außer im Lesen (dort C1) steht zum Beispiel in Bayern nicht nur bei Englisch als erster Fremdsprache in den Lehrplänen, sondern auch bei Russisch als dritter Fremdsprache (Bayerisches Staatsministerium für Unterricht und Kultus 2003). Internationale Studien, insbesondere zur Entwicklung der Sprechkompetenz, zeigen allerdings, dass zum Erreichen fremdsprachlicher Kompetenzen große Zeitspannen notwendig sind, die überdies von der Intensität der Sprachlernerfahrungen sowie der Distanz von Ausgangssprache zu Zielsprache abhängen (Rifkin 2003; Swender 2003; Tschirner 2007a).

Während es für den GeR nur sehr wenige empirische Studien gibt, was Lernerleistungen betrifft, gibt es für die *Oral Proficiency Guidelines* des *American Council on the Teaching of Foreign Languages* (ACTFL) eine Reihe von Unter-

<sup>1</sup> Korrespondenzadresse: Prof. Dr. Erwin Tschirner, Universität Leipzig, Herder-Institut, Beethovenstraße 15, 04107 Leipzig, Email: tschirner@uni-leipzig.de

suchungen, die sich mit dem Faktor Zeit auseinandergesetzt haben (für einen Überblick siehe Tschirner & Heilenman 1998 und Tschirner 2007a). Mit Blick auf eine vernünftige Lernzielsystematik könnte sich eine Orientierung an Ergebnissen dieser Studien als fruchtbar erweisen. Im Folgenden sollen deshalb die GeR-Skalen mit den ACTFL-Skalen verglichen werden, um zu zeigen, dass sie große Gemeinsamkeiten aufweisen und dass die empirischen Ergebnisse im Zusammenhang mit den ACTFL-Skalen auch für Bildungsstandards, die auf GeR-Skalen aufbauen, eine Grundlage bilden können. Im zweiten Teil werden diese Ergebnisse im Hinblick auf die Faktoren Lernzeit und Sprachendistanz präsentiert. Im dritten Teil werden die Ergebnisse der DESI-Studie vorgestellt und sie werden mit den Anforderungen standardisierter internationaler Englischtests und mit den Ergebnissen des zweiten Teils verglichen. Da diese Ergebnisse mit den Ergebnissen der DESI-Studie im Zusammenhang mit der Entwicklung der Sprechkompetenz unvereinbar sind, wird im vierten Teil der *PhonePass Set-10-Test*, der in der DESI-Studie für die Erfassung der Sprechkompetenz verwendet wurde, genau untersucht und es wird argumentiert, dass dieser Test keine valide Basis für GeR-Niveaus darstellt. Im fünften und letzten Teil wird versucht, auf Basis der Ergebnisse zum ACTFL *Oral Proficiency Interview* (OPI) realistische Kompetenzziele für die Entwicklung der Sprechkompetenz im schulischen Fremdsprachenunterricht zu formulieren.

## 1. Vergleich GeR und ACTFL

Die Kompetenzbeschreibungen des GeR sind das Produkt des Forschungsprojekts "Evaluation und Selbstevaluation der Fremdsprachenkompetenzen an Schnittstellen des Schweizer Bildungssystems" unter Leitung von Günther Schneider und Brian North. Dabei wurden existierende Skalen dekonstruiert (u.a. auch die ACTFL-Skala) und die daraus gewonnenen Kompetenzbeschreibungen von Schweizer Fremdsprachenlehrerinnen und -lehrern in mehreren Schritten in Schwierigkeitsniveaus eingeteilt. Das Schweizer Projekt ergab insgesamt neun relativ gleichmäßig verteilte Schwierigkeitsniveaus, nämlich A1, A2, A2+, B1, B1+, B2, B2+, C1 und C2 (Schneider 2001).

Das ACTFL *Oral Proficiency Interview* (OPI) ist eine Weiterentwicklung des OPI des amerikanischen *Foreign Service Institute* (FSI), das 1956 entwickelt als die Mutter aller mündlichen Kompetenzskalen gilt (Chalhoub-Deville & Fulcher 2003; Fulcher 1997; Herzog 2003; Liskin-Gasparro 2003). Die Skalen wurden mehrmals, zuletzt 1999, überarbeitet und unterscheiden zwischen zehn Kompetenzniveaus, *Novice Low, Mid, High, Intermediate Low, Mid, High, Advanced Low, Mid, High* und *Superior* (Breiner-Sanders et al. 2000). Die Skalen des Re-

ferenzrahmens und der *Oral Proficiency Guidelines* beschreiben die Entwicklung mündlicher Handlungsfähigkeit von Anfängern bis zu sehr fortgeschrittenen Fremdsprachensprechern.

Die GeR-Skala baut explizit auf dem Modell der kommunikativen Kompetenz von Bachman (1990) auf (Schneider 2001), die ACTFL-Skala ist mit ihm kompatibel (Tschirner 2001). Beide Skalen beruhen auf vielfältigen und breit angelegten Expertenurteilen und damit auf intersubjektiven empirischen Beobachtungen. Beide Skalen haben die Skala des FSI der USA als Grundlage, die ACTFL-Skala direkt und die GeR-Skala indirekt über die Dekonstruktion mündlicher Kompetenzbeschreibungen, von denen viele auf der FSI-Skala aufbauen. Da beide Skalen die Diskussion um Bildungsstandards in ihren jeweiligen Heimatmärkten dominieren, diese Heimatmärkte aber umfangreich miteinander vernetzt und verbunden sind, ist es ein dringendes Desiderat, Korrelationen zwischen beiden Skalen herzustellen.

Vandergrift (2006), der für einen *Common Framework of Reference for Languages of Canada* bestehende Skalen analysiert, korreliert das Niveau *Novice High* mit A1, *Intermediate High* mit A2, *Advanced Low* mit B1, *Advanced Mid* mit B2, *Advanced High* mit C1 und *Superior* mit C2. Zu einem sehr ähnlichen Ergebnis kommt Tschirner (2005), der nur *Intermediate High* etwas höher einstuft, nämlich B1, und dadurch *Intermediate Mid* mit A2 gleichsetzt. North (Majima 2006) setzt ebenfalls *Intermediate High* mit B1 und *Advanced Mid* mit B2 gleich, meint aber *Superior* wäre C1. Tschirner (2007b) geht von den ursprünglich neun Niveaus des Referenzrahmens aus und schlägt auf der Basis einer Mikroanalyse der Niveaubeschreibungen beider Skalen die folgende Äquivalenztabelle vor (s. Tabelle 1). Eine Gegenüberstellung der Niveaubeschreibungen findet sich in Anlage 1.

GeR	A1	A2	A2+	B1	B1+	B2	B2+	C1	C2
ACTFL	NH	IL	IM	IH	AL	AM		AH	S

Tabelle 1: Äquivalenztabelle zwischen GeR und ACTFL

Beide Skalen gehen von einer wachsenden Kompetenz in vier Aspekten aus: in der Länge und Komplexität von Äußerungen, in der Bandbreite und Art der Themen, in der Bandbreite und Komplexität der Sprechakte und in der Beherrschung der formalen Seite der Sprache, ihrer Aussprache, Grammatik, Lexik und soziokulturellen Angemessenheit. Am Beispiel *Oral Proficiency Guidelines* soll dies kurz erläutert werden. Die organisatorische Kompetenz (nach Bachmann 1990) entwickelt sich über die funktionale Aneinanderreihung von Wörtern (*Novice*) zur Fähigkeit, einfache Sätze zu bilden und aneinander zu reihen (*Interme-*

diate) zur Fähigkeit, einfache (*Advanced*) und komplexe Texte (*Superior*) zu produzieren, die sich sowohl durch ihre Länge wie durch den Grad ihrer Verknüpftheit (Kohäsion) unterscheiden.

Die thematische Kompetenz entwickelt sich von elementaren Bedürfnissen und Konzepten (*Novice*) über grundlegende persönliche Bereiche (*Intermediate*) zur Fähigkeit, über aktuelle (*Advanced*) und abstrakte Themen (*Superior*) sprechen zu können. Die funktionale Kompetenz entwickelt sich vom einfachen Reagieren mit Wörtern und Phrasen auf einfache Fragen (*Novice*) über die Fähigkeit, mit Hilfe von einfachen Sätzen zu persönlichen Themen Fragen zu stellen und Antworten zu geben (*Intermediate*), zur Fähigkeit, deskriptive und narrative Texte zu persönlichen und über das Persönliche hinausgehende Erlebnisse, Ereignisse und Pläne (*Advanced*) und abwägende, argumentierende, hypothesenbildende Texte zu konkreten und abstrakten Themen zu produzieren (*Superior*).

Die formale Kompetenz entwickelt sich von der Fähigkeit, minimal verständlich zu sprechen (*Novice*) über die Fähigkeit, Kohärenz durch ein einfaches Verwenden oder Aneinanderreihen von verständlichen Äußerungen auf Satzniveau (*Intermediate*) zur Fähigkeit, einfache Registerunterscheidungen konsequent durchzuhalten, temporale und aktionale Kohärenz durch konsequente Verwendung von Tempusformen sowie die Informationsstruktur von deskriptiven und narrativen Texten durch Wortstellung, Kasus, Anaphorik und Konnexion für Zuhörer verständlich zu machen (*Advanced*) und schließlich zur Fähigkeit, komplexe abstrakte Texte in gehobener distanzierender und differenzierender Sprechweise ohne ablenkende Aussprache-, Grammatik-, Lexik- und Registerauffälligkeiten weitgehend fehlerfrei zu produzieren (*Superior*).

Die erste großangelegte Studie, in der Kompetenzniveaus von amerikanischen Bachelor-Absolventen mit einem Abschluss in einer Fremdsprachenphilologie untersucht wurden, stammt aus dem Jahr 1967 (Carroll 1967). In den seit dieser Studie verstrichenen 40 Jahren gab es eine Vielzahl von Arbeiten, die sich u.a. mit der Validität und Reliabilität des OPI auseinandersetzten (Malone 2003; Surface & Dierdorff 2003; Watanabe 1998), mit der Frage, wie lange es dauert, um ein bestimmtes Niveau bei unterschiedlichen Ausgangs- und Zielsprachen zu erreichen (u.a. Swender 2003; Tschirner & Heilenman 1998), welche grammatischen Strukturen auf welchen Niveaus beherrscht werden (u.a. Magnan 1988; Mikhailova 2006; Norris 1996; Rifkin 2002; Rubio 2003; Tschirner 1996; Watanabe 2003) und welche Korrelationen es mit anderen Tests gibt (u.a. Halleck & Moder 1995; Kenyon & Tschirner 2000). Ich möchte im Folgenden die Ergebnisse herausgreifen, die sich mit Fragen der Zeit, Sprachlernerfahrung und Sprachdistanz beschäftigt haben.

## 2. Unterrichtsstunden, Sprachlernerfahrungen, Sprachdistanz

Eine größere Anzahl von Studien zum OPI hat sich damit auseinandergesetzt, wie viele Unterrichtsstunden und welche Art von Sprachlernerfahrungen benötigt werden, um ein bestimmtes Niveau zu erreichen. Je nach Ausgangs- und Zielsprache und nach Lernergruppe werden dabei unterschiedlich viele Unterrichtsstunden benötigt. Das amerikanische *Foreign Service Institute* (Foreign Service Institute 2006) unterteilt Fremdsprachen bei Ausgangssprache Englisch in drei Schwierigkeitskategorien. Zur Kategorie I (leichte Sprachen) gehören romanische und germanische Sprachen außer Deutsch, zur Kategorie II (schwere Sprachen) gehören slavische Sprachen und Sprachen wie Hindi, Neugriechisch, Ungarisch, Türkisch sowie Vietnamesisch und zur Kategorie III (superschwere Sprachen) gehören Arabisch, Chinesisch, Japanisch und Koreanisch. Der Unterricht findet in sehr kleinen Gruppen (maximal sechs Teilnehmer) unter intensiven Bedingungen statt (22 Stunden Unterricht plus 18 Stunden aufgabenbasiertes eigenverantwortliches Arbeiten). Die Kursteilnehmer sind im Durchschnitt 40 Jahre alt, haben mindestens einen Masterabschluss und sind hochmotiviert. Unter diesen Bedingungen erreichen sie im Sprechen und Lesen das Niveau *Advanced High* bei leichten Sprachen nach durchschnittlich 960 Arbeitsstunden (inkl. 600 Unterrichtsstunden), bei schweren Sprachen nach durchschnittlich 1760 Arbeitsstunden (inkl. 1100 Unterrichtsstunden) und bei superschweren Sprachen nach durchschnittlich 3520 Arbeitsstunden (inkl. 2200 Unterrichtsstunden), wobei im Normalfall ein ganzes Jahr im Land der Zielsprache verbracht wird. Für Deutsch werden im Durchschnitt 1200 Arbeitsstunden (inkl. 750 Unterrichtsstunden) angesetzt (s. Tabelle 2).

	I	Deutsch	II	III
AH	960	1200	1760	3520

Tabelle 2: Stundenverteilung am *Foreign Service Institute* (USA) zum Erreichen des Niveaus *Advanced High* (AH) bei Sprachen der Kategorien I, II und III sowie des Deutschen

Empirische Studien zum OPI bestätigen und ergänzen diese Vorgaben. Die unteren Niveaus bis *Intermediate Mid* (IM) werden bei verwandten Sprachen relativ schnell erreicht, nämlich *Intermediate Low* (IL) nach 150 Stunden und IM nach 300 Stunden (Tschirner & Heilenman 1998). Der Sprung auf das Niveau *Intermediate High* (IH) und darüber hinaus stellt jedoch eine große Hürde dar und wird oft nur über einen Auslandsaufenthalt von sechs Monaten und mehr erreicht. Rifkin (2003) fand ein Niveau von *Advanced Low* (AL) oder bes-

ser nur bei Bachelor-Absolventen in Russisch, die mindestens ein Semester in Russland studiert hatten.

Swender (2003) fasst die Ergebnisse von fünf amerikanischen Elite-Universitäten zusammen, die über einen Zeitraum von fünf Jahren (1998-2002) ihre Studierenden (insgesamt 501) regelmäßig zum Abschluss ihres Bachelorstudiums einer Fremdsprachenphilologie von ACTFL testen ließen. Viele dieser Studierenden hatten dabei an einem Austauschprogramm im Land der Zielsprache, das im Normalfall ein akademisches Jahr beinhaltete, teilgenommen. 17% der Studierenden der Fächer Hispanistik, Französisistik und Italianistik (insgesamt 442 Studierende) waren am Ende ihres Bachelorstudiums IM, 57% waren IH oder AL und 26% waren *Advanced Mid* (AM) oder besser. Recht ähnliche Ergebnisse erzielten Studierende der Germanistik und Russistik (insgesamt 39 Studierende): 26% waren IM, 44% waren IH/AL und 30% waren AM oder besser. Etwas weniger gute, aber sehr akzeptable Ergebnisse erzielten Studierende der Sinologie und Japanologie (insgesamt 20 Studierende): 35% waren IM, 50% waren IH/AL und 15% waren AM oder besser.

### 3. Deutsch Englisch Schülerleistungen International (DESI)

Wie die Studien zur mündlichen Kompetenz amerikanischer Studierender zeigen, ist das Niveau *Intermediate High* (B1) das durchschnittliche Niveau am Ende eines Bachelorstudiums, selbst wenn ein ganzes Jahr davon im Land der Zielsprache studiert wird. *Advanced Mid* (B2) wird nur vom oberen Drittel erreicht und meist nur, wenn ein Jahr im Land der Zielsprache studiert wurde. Die DESI-Studie nun kommt zu deutlich anderen Ergebnissen (Schröder, Harsch & Nold 2006). Was mögliche Gründe dafür sind und wie belastbar diese Ergebnisse sind, soll in diesem Abschnitt erörtert werden.

Die DESI-Studie ist eine nationale Studie zur Erfassung der sprachlichen Leistungen in Deutsch und Englisch von Schülerinnen und Schülern der 9. Klassen aller Schularten. Sie wurde 2001 von der Kultusministerkonferenz (KMK) in Auftrag gegeben und soll eine nationale Ergänzung zur PISA-Studie darstellen. Unter Federführung des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF) wurden von einem Wissenschaftlerkonsortium knapp 11.000 Schüler zu Beginn und am Ende des Schuljahres 2003/04 getestet. Für das Fach Englisch wurden die mündliche Sprechfähigkeit, Hör- und Leseverstehen, kreatives Schreiben, Aspekte der Sprachbewusstheit, interkulturelle Kompetenz und die Fähigkeit zur Rekonstruktion von Texten getestet (Kultusministerkonferenz 2006).

Der einzige Test, der nach Aussage des DESI-Konsortiums direkt Niveaus des GeR abbildet, ist der Sprechtest. Danach erreichen zwei Drittel der Schüler am Ende der 9. Jahrgangsstufe das Niveau A2, ein Drittel erreicht B1 und neun Prozent erreichen B2. Bei dem dabei verwendeten Test handelt es sich um den *PhonePass Set-10*-Test, einen computerbasierten Test, der laut Aussage des DESI-Konsortiums an den Niveaus des Referenzrahmens normiert wurde. Deshalb könnten die Zielsetzungen der KMK-Bildungsstandards für den Hauptschulabschluss explizit überprüft werden (DESI-Konsortium 2006).

Im Folgenden sollen zuerst die Ergebnisse der DESI-Studie mit standardisierten, internationalen Tests des Englischen verglichen werden, um auf ein Glaubwürdigkeitsproblem dieser Ergebnisse hinzuweisen. Im Anschluss wird der *PhonePass Set-10*-Test vorgestellt und seine "Normierung" am GeR diskutiert, wobei argumentiert werden wird, dass es keine zuverlässige Normierung des *PhonePass Set-10*-Tests an GeR-Niveaus gegeben hat.

Tannenbaum & Wylie (2005) normierten in einer umfangreichen Studie den Test of English as a Foreign Language (TOEFL), den Test of Spoken English (TSE), den Test of Written English (TWE) und den Test of English for International Communication (TOEIC) an den Niveaus B1 und C1 des GeR. Tabelle 3 zeigt diese Ergebnisse ebenso wie die Cambridge-Zertifikate, die den einzelnen GeR-Niveaus zugeordnet sind, nämlich den Key English Test (KET) (A2), den Preliminary English Test (PET) (B1), das First Certificate in English (FCE) (B2), das Certificate in Advanced English (CAE) (C1) und das Certificate of Proficiency in English (CPE) (C2). Tabelle 4 enthält weiterhin die Ergebnisse des DESI-Sprechtests für alle 9. Klassen ebenso wie separat für die 9. Klasse Gymnasium (Schröder et al. 2006). Schließlich enthält die Tabelle die in Abschnitt 2 vorgestellten Ergebnisse amerikanischer Bachelor-Absolventen der Fachrichtung Germanistik nach ca. 700 Unterrichtsstunden, die zusätzlich zu ihrem amerikanischen Studium ein Austauschstudienjahr in einem deutschsprachigen Land verbracht hatten (Swender 2003).

GeR	A1	A2	B1	B2	C1	C2
TSE			45		55	
TWE			4,5		5,5	
TOEFL (Papier)			457		560	
TOEIC			550		880	
Cambridge Zertifikate		KET	PET	FCE	CAE	CPE
DESI 9. Kl. alle	28%	32%	25%	8%	1%	0,5%
DESI 9. Kl. Gym.	5%	20%	50%	24%	1%	0,5%
BA (700 h) + 1 Jahr D		26%	44%	30%		

Tabelle 3: Vergleich internationaler Englischtests und Zertifikate mit DESI-Ergebnissen und den Ergebnissen der OPI-Studie von Swender (2003)

Nach den DESI-Ergebnissen würden 72% aller Neuntklässler den *Key English Test* der *University of Cambridge ESOL Examinations* bestehen, 40% den *Preliminary English Test* und fast 10% würden sogar das *First Certificate in English* (FCE) erwerben können, von Gymnasialschülern würde letzteres sogar von jedem/jeder Vierten erworben werden. Um das FCE zu erwerben, werden ca. 600 Unterrichtsstunden in kleinen Gruppen mit hochmotivierten Lernern und muttersprachlichen Lehrern, in denen der gesamte Unterricht auf Englisch stattfindet, benötigt. Lerner, die das FCE erworben haben,

- understand texts from a wide variety of sources;
- use English to make notes while someone is speaking in English;
- talk to people about a wide variety of topics;
- understand people talking in English on radio or television programmes (University of Cambridge ESOL Examinations).

40% aller Hauptschüler und 75% aller Gymnasialschüler würden am Ende der 9. Klasse mindestens 457 Punkte im TOEFL oder 550 Punkte im TOEIC erreichen. Schließlich wären die DESI-Ergebnisse für Gymnasialschüler am Ende der 9. Klasse den Ergebnissen von amerikanischen Bachelor-Absolventen eines Philologiestudiums, die ein Jahr im Land der Zielsprache studiert haben, sehr ähnlich. All dies deutet auf ein Glaubwürdigkeitsproblem der DESI-Ergebnisse hin. Deshalb soll im folgenden Teil der Test, der zu diesen Ergebnissen geführt hat, untersucht werden.

#### 4. *PhonePass Set-10*

Der *PhonePass SET-10*-Test ist ein kommerzieller Test der *Ordinate Corporation* in Menlo Park, Kalifornien. Er verwendet Spracherkennungstechnologie und ist vollkommen automatisiert, nicht nur der Test selbst, sondern auch die Bewertung verläuft vollautomatisch. Er wird über das Telefon gegeben und dauert ca. 10 Minuten. Es handelt sich bei diesem Test in erster Linie um einen Aussprachetest, obwohl die Hersteller behaupten, dass er "facility in spoken English" evaluiert, weil es hochautomatisierter Hör- und Sprechroutinen bedürfe, um gut abzuschneiden. Der *PhonePass SET-10* besteht aus fünf Teilen, wovon nur die ersten vier bewertet werden (Ordinate 2007).

In Teil 1 haben die Testteilnehmer zwölf Sätze vor sich liegen, die in lose miteinander verbundenen Vierergruppen eingeteilt sind, um Missverständnisse zu vermeiden. Die Sätze sind vom Vokabular und den grammatischen Strukturen her relativ einfach gehalten. Hier ist ein Beispiel für eine Vierergruppe.

- (1) Traffic is a huge problem in Southern California.
- (2) The endless city has no coherent mass transit system.
- (3) Sharing rides was going to be the solution to rush-hour traffic.
- (4) Most people still want to drive their own cars, though.

Von diesen zwölf Sätzen werden acht per Zufallsauswahl ausgewählt, die die Testteilnehmer vorlesen müssen.

In Teil 2 müssen 16 Äußerungen wiederholt werden, die von unterschiedlichen US-amerikanischen Sprechern vorgesprochen werden. Ihre Länge besteht aus drei bis 15 Wörtern. Diese Äußerungen werden nur gehört. Hier sind drei Beispiele:

- (5) War broke out.
- (6) It's supposed to rain tomorrow, isn't it?
- (7) There are three basic ways in which a story might be told to someone.

Teil 3 besteht aus einfachen gesprochenen Fragen, die rezeptiven und produktiven Wortschatz überprüfen sollen. Diese Fragen setzen kein spezielles Wissen voraus und wurden so formuliert, dass sie ein 12-jähriger Muttersprachler des Englischen problemlos beantworten kann. In jeder Frage sind mindestens drei bis vier lexikalische Einheiten enthalten. Als Antworten werden Kurzwörter erwartet. Hier sind drei typische Fragen:

- (8) What season comes before spring?  
 (9) What is frozen water called?  
 (10) Does a tree usually have fewer trunks or branches?

Im *PhonePass Set-10-Test* besteht dieser Teil aus 24 Fragen, für die DESI-Studie wurde er auf 16 Fragen reduziert.

Teil 4 besteht aus 10 Sätzen, deren Bestandteile durcheinandergewürfelt sind und die von den Testteilnehmern in die richtige Reihenfolge gebracht werden müssen. Es handelt sich dabei jeweils um relativ kurze, einfache Sätze, die in drei Einheiten aufgeteilt werden. Auch diese Einheiten werden nur gehört. Hier sind drei typische Vorgaben:

- (11) in / bed / stay  
 (12) Ralph / this photograph / could convince  
 (13) we wondered / would fit in here / whether the new piano

Teil 5 fließt nicht in die Ergebnisse ein, sondern muss durch menschliche Bewerter beurteilt werden. Es handelt sich hier um zwei Fragen, die in jeweils 30 Sekunden beantwortet werden sollen. Im Originaltest werden hier Fragen zu "family life" und "personal choices" gestellt. Das DESI-Konsortium hat für diesen Teil eigene Fragen entwickelt, die auf ihre Zielgruppe zugeschnitten waren, und insgesamt drei Fragen gestellt, die in jeweils 20 Sekunden beantwortet werden sollten. Auch in der DESI-Studie floss dieser Teil nicht in die Ergebnisse ein.

Der Wortschatz und die grammatischen Strukturen der Testsätze wurden empirisch ermittelt. Es wurden insgesamt 540 Gespräche zwischen US-amerikanischen Muttersprachlern aufgezeichnet. Diese Gespräche wurden sorgfältig ausgewählt, um Männer und Frauen und alle wichtigen Regiolekte des amerikanischen Englisch gleichmäßig zu repräsentieren. Testsätze wurden von Kontrollpersonen in Großbritannien und Australien gegengelesen, um Amerikanismen lexikalischer Art zu vermeiden.

Die Antworten werden nach einem internen Algorithmus, der nicht veröffentlicht ist, anhand von vier Kriterien ausgewertet: Satzbau, Wortschatz, Flüssigkeit und Aussprache. Punkte für Satzbau werden bei den Testteilen "Sätze nachsprechen" und "Satzteile ordnen" vergeben, Punkte für Wortschatz beim Testteil "Kurzantworten", Punkte für Flüssigkeit bei den Testteilen "Vorlesen", "Sätze nachsprechen" und "Satzteile ordnen", und Punkte für Aussprache ebenfalls bei den Testteilen "Vorlesen", "Sätze nachsprechen" und "Satzteile ordnen" (s. Tabelle 4).

	Satzbau	Wortschatz	Flüssigkeit	Aussprache
Vorlesen				
Nachsprechen				
Kurzantworten				
Satzteile ordnen				

Tabelle 4: *PhonePass Set-10-Test*teile und -Bewertungskriterien

Satzbau und Wortschatz ergeben 50% der Gesamtpunktzahl, wobei die Punkte für das Vorkommen erwarteter Wörter in den Äußerungen der Testteilnehmer vergeben werden. Aussprache und Flüssigkeit ergeben ebenfalls 50% der Gesamtpunktzahl. Punkte für Aussprache werden für die korrekte Aussprache von Konsonanten und Vokalen und korrekte Betonung von Silben in Wort- und Satzzusammenhängen vergeben. Korrektheit wird definiert anhand der Aussprache und Betonung von gebildeten, erwachsenen Muttersprachlern des amerikanischen Englisch. Flüssigkeit wird daran gemessen, wie lange es dauert, bis der Testteilnehmer zu sprechen beginnt, an seiner Sprechgeschwindigkeit sowie der Position und Länge seiner Pausen.

Es ist unklar, was der Test genau misst, sicherlich nicht produktive mündliche Handlungskompetenz. Im Teil "Sätze vorlesen" wird Aussprache und Flüssigkeit bewertet, im Teil "Sätze nachsprechen" ebenfalls Aussprache und Flüssigkeit, teilweise auch Hörverständnis und Wortschatz, im Teil "Kurzantworten" wird Hörverständnis und Wortschatz bewertet und im Teil "Satzteile ordnen" ebenfalls Hörverständnis und Wortschatz ebenso wie Aussprache und Flüssigkeit. Produktive mündliche Handlungsfähigkeit wird nur im Teil 5 elizitiert, der allerdings nicht in die Bewertung einfließt. Selbst als Aussprachetest ist der *PhonePass Set-10-Test* nicht unumstritten. Hincks (2001) fand zum Beispiel heraus, dass Testteilnehmer, die Testsätze schneller vorlasen, bessere Ergebnisse erzielten.

Zu *Versant for English*, wie der *PhonePass Set-10* seit 2006 genannt wird, gibt es nur wenig publiziertes Material. Es gibt einige wenige Artikel, vor allem aus dem Bereich Spracherkennung, in denen er als Aussprachetest erwähnt wird (z.B. de Wet, van der Walt & Niesler 2007), und drei Artikel von Hincks, die sich kritisch mit seiner Rolle als Aussprachetest auseinandersetzen (Hincks 2001, 2002, 2003). Zwei Studien hatten das Ziel, den *PhonePass Set-10-Test* mit dem GeR zu korrelieren (de Jong & Bernstein 2001; Suziki & Harada 2004). Diesen Studien möchte ich mich jetzt zuwenden.

Die erste Studie wurde 2001 von John de Jong und Jared Bernstein auf der *7th European Conference on Speech Communication and Technology* in Aalborg, Dänemark, präsentiert und in den Kongressakten veröffentlicht (de Jong & Bernstein 2001). Jared Bernstein ist der Begründer der *Ordinate Corporation* und jetzige Präsident. In dieser Studie wurde Teil 5, "Offene Fragen", mit den anderen vier Teilen korreliert. Der Teil "Offene Fragen" besteht aus zwei persönlichen Fragen mit Antworten, die bis zu 30 Sekunden dauern können. Diese Antworten wurden von drei Bewertern beurteilt. Einer kam aus den Niederlanden, einer aus der Schweiz und einer aus Großbritannien. Nur einer war ein professioneller Bewerter von Englischtests. Diese Bewerter wurden anhand von sechs Videoaufnahmen von Sprechern auf unterschiedlichen GeR-Niveaus geschult, eine Videoaufnahme einer Aufgabe pro Niveau. Im Anschluss wurden sie gebeten, sechs Audioaufnahmen aus dem Testteil "Offene Fragen" des *PhonePass Set-10* zu bewerten und sich auf eine Bewertung zu einigen. Die GeR-Deskriptoren wurden vereinfacht und auf den *PhonePass Set-10*-Test zugeschnitten (s. Anhang 2). Der Testteil "Offene Fragen" wurde anschließend von 121 Probanden bewertet. Die Interrater-Reliabilität war mit 0,95 sehr hoch, was auch damit zusammenhängt, dass es sich nur um sehr wenige Bewerter handelte, die überdies mit trennscharfen und einfachen Deskriptoren arbeiten konnten.

Die Korrelation zwischen dem Testteil "Offene Fragen" mit den Teilen 1-4 war mit 0,84 moderat hoch. Dies bedeutet allerdings nur, dass Probanden, die höhere Punktzahlen in den Teilen 1-4 erreichten, auch höhere Punktzahlen im Teil 5 bekamen. Es bedeutet nicht, dass Punktzahlen in den Teilen 1-4 mit dem Referenzrahmen korrelieren, da die vereinfachten Deskriptoren der Studie die Deskriptoren des Referenzrahmens nur unzureichend abbilden. Weiterhin variieren angenommene Referenzrahmenniveaus im Vergleich zu den *PhonePass*-Ergebnissen erheblich. Wie das Streudiagramm der Studie zeigt (de Jong & Bernstein 2001: *Figure 3*), variieren A1-Resultate von 38-59, A2 von 43-62, B1 von 49-80, B2 von 58-80 und C1 von 69-80. Desweiteren wurden Probanden vor allen auf den Niveaus unter A1 bis B1 eingestuft (84%), nur elf Probanden wurden mit B2 bewertet, 6 mit C1 und 2 mit C2. Schließlich ist es wenig plausibel, dass eine zuverlässige Einstufung eines Testteilnehmers auf einer Kompetenzskala auf der Basis von maximal einer Minute gesprochener Sprache möglich ist. Aus all diesen Gründen komme ich zu dem Schluss, dass mit dieser Studie keine zuverlässige Normierung des *PhonePass Set-10*-Tests auf Grundlage des GeR stattgefunden hat.

Auch die zweite Studie wurde von *Ordinate* initiiert und finanziert und folgte einem ähnlichen Muster (Suziki & Harada 2004). Die automatisch bewerteten Teilbereiche wurden mit dem Teilbereich "Offene Fragen" korreliert. An dieser Studie nahmen 268 Nichtmuttersprachler und 35 Muttersprachler teil. Beurteilt

wurden die "Offenen Fragen" von sechs Bewertern. Die Interrater-Reliabilität war mit 0,89 hoch, ebenso die Korrelation von 0,88 zwischen den automatisch und den von Bewertern bewerteten Teilen. Allerdings trägt auch hier die hohe Korrelation. Völlige Übereinstimmung zwischen Bewertern war 63%, ein Auseinanderklaffen um eine Stufe war bei 30% der Fall. Interessant ist auch, dass vier der Muttersprachler mit B1 bewertet wurden und vier weitere mit B2. Das Streudiagramm wiederum ergibt eine große Bandbreite von Ergebnissen pro Kompetenzniveau, nämlich 25-51 für A1, 25-56 für A2, 31-77 für B1, 41-76 für B2, 51-80 für C1 und 71-76 für C2 bei Nichtmuttersprachlern des Englischen. Ungewöhnlich scheint auch die Tatsache zu sein, dass die Studie von 2004 (Suziki & Harada) *Ordinate* zu anderen Trennpunkten bewegte (*Ordinate* 2007) (s. Tabelle 5).

	A1	A2	B1	B2	C1	C2
2001	40-49	50-55	56-61	62-67	68-72	73-80
2006	26-35	36-46	47-57	58-68	69-78	79-80

Tabelle 5: Korrelierung des *PhonePass Set-10*-Tests am GeR anhand zweier Studien der *Ordinate Corporation*

Da auch diese Studie mit den gleichen Problemen behaftet ist wie die vorhergehende, u.a. bewertbare Texte maximal im Umfang von einer Minute, ungeklärtes Verhältnis von Deskriptoren der Studie zu den Deskriptoren des GeR, Ballung von Ergebnissen auf den Niveaus von unter A1 bis B1, erhebliche Variation der *PhonePass*-Punkte auf GeR-Niveaus, ist zu bezweifeln, ob die neuen Trennpunkte zuverlässiger als die alten sind. Im Grunde genommen sind beide Korrelationen gleich unzuverlässig. Damit sind auch die Ergebnisse der DESI-Studie, was die Sprechkompetenz betrifft, unzuverlässig und können nicht mit dem GeR korreliert werden.

## 5. Vernünftige Erwartungen

Ich habe versucht zu zeigen, dass die DESI-Studie im Hinblick auf die mündliche Sprachkompetenz keine belastbaren Ergebnisse für erreichbare GeR-Niveaus geliefert hat. Um diese zu bekommen, werden dringend empirische Studien benötigt, ohne die die Bildungsstandards auf tönernen Füßen stehen. Was die Sprechkompetenz betrifft, deuten amerikanische Studien darauf hin, dass die deutschen Bildungsstandards zu ehrgeizig sind, selbst was die erste Fremdsprache betrifft. Auf völlig tönernen Füßen stehen allerdings die Annah-

men, dass es keine Rolle spielt, wie lange man eine Sprache lernt, oder dass alle Sprachen gleich schnell gelernt werden können. Bildungsstandards, die für alle Sprachen mehr oder weniger dieselben Ziele für das Abitur vorgeben, können deshalb nicht erreicht werden. Im Folgenden versuche ich, auf der Basis von Studien, die sich mit dem Erwerb der Sprechkompetenz beschäftigt haben, Hypothesen für den schulischen Fremdspracherwerb aufzustellen, die empirisch zu überprüfen wären. Ausgangspunkt meiner Überlegungen sind die bayerischen Vorgaben für das Gymnasium.

Tabelle 6 zeigt in Zeile 1 die Klassenstufen für das Abitur nach zwölf Jahren, Zeile 2 die kumulativen Stunden für die erste Fremdsprache, Zeile 3 das erwartete Niveau am Ende der Klassenstufe und Zeile 4 ein aus meiner Sicht realistisches Niveau. Die Zeilen 5-7 zeigen das gleiche für die zweite Fremdsprache ab Klasse 6, die Zeilen 8-11 für die dritte Fremdsprache ab Klasse 8 und die Zeilen 12-14 für die spät beginnende Fremdsprache in Klassen 11 und 12.

1	Klassenstufe	5.	6.	7.	8.	9.	10.	12.
2	1. FS/Stunden	200	360	520	640	760	880	1200
3	Erwartetes Niveau	A1	A1+	A2	A2+	B1	B1+	B2+
4	Realistisches Niveau	A1	A2	A2+			B1	B1+
5	2. FS/Stunden		160	320	480	600	720	1040
6	Erwartetes Niveau		A1	A2	A2+	B1	B1+	B2
7	Realistisches Niveau		A1	A2	A2+		B1	B1+
8	3. FS/Stunden				160	320	480	800
9	Erwartetes Niveau				A2	A2+	B1+	B2
10	Realistisches Niveau				A2	A2+	A2+	(B1)
11	Russisch				A1+	A2	A2+	(B1)
12	Spätbeginnend							360
13	Erwartetes Niveau							B1
14	Realistisches Niveau							A2+

Tabelle 6: Stundenzahlen, erwartete und realistische Niveaus für die Sprechkompetenz in Fremdsprachen an Gymnasien

Es ist sicherlich nicht unrealistisch, anzunehmen, dass selbst oder gerade bei Kindern nach 200 oder gar 360 Stunden in Bezug auf die mündliche Handlungsfähigkeit in der ersten Fremdsprache das Niveau A2 erreicht werden kann, insbesondere wenn der Sprachunterricht bereits in der Grundschule begonnen hat. Damit sollte am Ende der 7. Klasse auch das Niveau A2+ erreicht werden können. Das Niveau B1 allerdings ist eine Schwelle, die nicht einfach zu überwinden ist. Dazu bedarf es einer intensiven Sprachlernerfahrung, die oft nur über einen längeren Auslandsaufenthalt gemacht werden kann. Ein einsprachiger

Unterricht mit vielen intensiven Sprachbaderfahrungen über Video/Fernsehen/Books-on-Tape und/oder mehrere Monate im Land der Zielsprache könnte es ebenfalls möglich machen, diese Schwelle zu überwinden. Selbst für die besten Schüler ist dies ohne einen längeren Auslandsaufenthalt wahrscheinlich nicht vor Ende der 10. Klasse zu erwarten. Das Niveau B2 oder B2+ ist ohne ein Jahr im Land der Zielsprache wenig realistisch und selbst dann nur bei ehrgeizigen Lernern, die sich nicht mit *Basic Interpersonal Communication Skills* (BICS) zufrieden geben, sondern sich aktiv und mit hohen eigenen Sprechanteilen um *Cognitive Academic Language Proficiency* (CALP) (Cummins 1979) bemühen. Für die zweite Fremdsprache gilt im Grunde genommen das Gleiche.

Für die dritte Fremdsprache ist das Niveau A2+ nach zwei Jahren durchaus realistisch, das Niveau B1 wird aber wahrscheinlich – ohne längeren Auslandsaufenthalt – selbst bis zum Abitur nur von wenigen erreicht werden. Bei Russisch und anderen nicht-affinen Sprachen ist das Niveau A2+ erst am Ende der 10. Klasse realistisch und wird sicherlich nicht von allen erreicht. B1 im Abitur wird wahrscheinlich nur von Schülern erreicht, die mindestens ein halbes Jahr in russischsprachiger Umgebung verbracht haben. Für spät beginnende Fremdsprachen (11. und 12. Klasse) ist das Maximum, was ohne längeren Auslandsaufenthalt oder familiäre Voraussetzungen erreicht werden kann, das Niveau A2+. Selbst dieses Niveau wird sicherlich nicht von allen erreicht.

Diese Hypothesen gelten für die Sprechkompetenz. Die Annahme, dass sich unterschiedliche Fertigkeiten unterschiedlich schnell entwickeln (Tschirner 1996), wird vom GeR durch die getrennte Darstellung der Fertigkeiten unterstützt. Dem tragen u.a. die bayerischen Bildungsstandards Rechnung, indem sie für die Lesekompetenz höhere Ziele ansetzen, beim Abitur in der ersten Fremdsprache z.B. C1. Es ist durchaus möglich, dass die in den Bildungsstandards verankerten Niveaus beim Lesen und Hören und mit Abstrichen beim Schreiben erreicht werden können. Sollten die Erfahrungen beim Erwerb der Sprechkompetenz auf den Erwerb der anderen Fertigkeiten übertragbar sein, so wäre auch dies wahrscheinlich nur dann der Fall, wenn die Fertigkeiten direkt und intensiv geübt werden können, was z.B. ein hohes Lesepensum in der Fremdsprache voraussetzen würde, einen aufgabenorientierten Unterricht, der fast ausschließlich in der Fremdsprache stattfindet, und häufige direkte Kontakte mit Sprechern/Schreibern der Zielsprache.

Ausgang meiner Überlegungen war, dass es wichtig ist, realistische Erwartungen in Bezug auf erreichbare Kompetenzniveaus im schulischen Fremdsprachenunterricht zu haben. Die Formulierung von Bildungsstandards auf Grundlage des GeR hat Handlungskompetenzen erneut in den Vordergrund gerückt. Es sollte dabei allerdings keine Diskrepanz zwischen Erwartetem und Möglichem



aufzutreten. Es gibt eine Reihe von Studien, die in Deutschland noch nicht genug wahrgenommen werden, die eine erste Annäherung an realistische Niveaus bilden können. Zwingend erforderlich sind empirische Studien, die zeigen, was an deutschen Schulen unter welchen Bedingungen erreichbar ist. Einen ersten Schritt in diese Richtung hat die DESI-Studie gemacht. Die Ergebnisse zur Sprechkompetenz sind aus meiner Sicht allerdings nicht belastbar und bedürfen dringend der Überprüfung. Dies schränkt den Wert der DESI-Studie an sich nicht ein, sondern weist vor allem darauf hin, dass, was die Ergebnisse der mündlichen Kompetenz betrifft, genauso vorsichtig argumentiert werden muss und genauso wenig der Referenzrahmen erwähnt werden kann, wie die DESI-Studie dies auch bei den anderen Fertigkeiten macht.

Die Schlussfolgerungen der DESI-Studie allerdings, dass im Englischunterricht viel mehr Englisch gesprochen werden muss, inklusive von den Schülern selbst, dass viel mehr Input mit Hilfe von audiovisuellen Lehrmaterialien, u.a. auch authentischen Materialien geliefert werden muss, dass viel mehr Wert auf das Hörverständnis gelegt werden muss, dass bilinguale Angebote einen Königsweg beim Lernen einer Fremdsprache darstellen können, sehe ich nicht in Frage gestellt. Ich würde hinzufügen, dass Auslandsaufenthalte und andere intensive Sprachlernerfahrungen unter Immersionsbedingungen (Klassenfahrten, *summer camps* und selbstverständlich bilingualer Sachfachunterricht) verstärkt in die Planung einbezogen werden müssen, wenn die anspruchsvollen Ziele im Fremdsprachenunterricht von mehr als nur einer kleinen Gruppe von Schülern erreicht werden sollen.

Eingang des revidierten Manuskripts 21.10.2008

## Literaturverzeichnis

- Bachman, Lyle (1990), *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bayerisches Staatsministerium für Unterricht und Kultus (Hrsg.) (2003), *Lehrplan für Gymnasien in Bayern*. Wolnzach: Kastner AG.
- Breiner-Sanders, Karen; Lowe, Pardee, Jr.; Miles, John & Swender, Elvira (2000), ACTFL proficiency guidelines – speaking, revised. *Foreign Language Annals* 33, 13-18.
- Carroll, John B. (1967), Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals* 1/2, 131-151.
- Chalhoub-Deville, Micheline & Fulcher, Glenn (2003), The Oral Proficiency Interview: A research agenda. *Foreign Language Annals* 36, 498-506.
- Cummins, Jim (1979), Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism* 19, 121-129.
- DESI-Konsortium (2006), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Zentrale Befunde der Studie Deutsch-Englisch-Schülerleistungen-International (DESI)*. Frankfurt am Main: Dipf.
- Foreign Service Institute (2006), *Language Continuum*. Washington: Foreign Service Institute. [Online: <http://fsitraining.state.gov/training/Language%20Continuum.pdf>. 2.10.2008].
- Fulcher, Glenn (1997), The testing of L2 speaking. In: Clapham, Caroline & Corson, David (Hrsg.) (1997), *Encyclopedia of language and education, vol. 7: Language testing and assessment*. Dordrecht: Kluwer, 75-85.
- Halleck, Gene & Moder, Carol (1995), Testing language and teaching skills of international teaching assistants: The limits of compensatory strategies. *TESOL Quarterly* 29, 733-758.
- Herzog, Martha (2003), Impact of the proficiency scale and the Oral Proficiency Interview on the foreign language program at the Defense Language Institute Foreign Language Center. *Foreign Language Annals* 36, 566-571.
- Hincks, Rebecca (2001), Using speech recognition to evaluate skills in spoken English. *Working Papers 49. Papers from the Fonetik 2001*. Lund: Lund University Department of Linguistics, 58-61. [Online: <http://www.speech.kth.se/~hincks/papers/fon01.pdf>. 2.10.2008]
- Hincks, Rebecca (2002), Speech recognition for language teaching and evaluating: A study of existing commercial products. *Proceedings from ICSLP 2002*. Denver: Center for Spoken Language Research, 773-776. [Online: <http://www.speech.kth.se/~hincks/papers/icslp02xxx.pdf>. 2.10.2008].
- Hincks, Rebecca (2003), Speech technologies for pronunciation feedback and evaluation. *ReCALL* 15, 3-20.
- de Jong, John H.A.L & Bernstein, Jared (2001), Relating *PhonePass* overall scores to the Council of Europe Framework level descriptors. In: Dalsgard, Paul; Lindberg, Børge; Benner, Hendrik & Tan, Zheng-hua (Hrsg.) (2001), *Proceedings of Eurospeech 2001 Scandinavia, 7th European Conference on Speech Communication and Technology*, 2803-2806.

- Kenyon, Dorry & Tschirner, Erwin (2000), The rating of direct and semi-direct oral proficiency interviews: Comparing performance at lower proficiency levels. *Modern Language Journal* 84, 85-101.
- Kultusministerkonferenz (2006), Stellungnahme der Kultusministerkonferenz zu den Ergebnissen der Studie "Deutsch Englisch Schülerleistungen International" (DESI). Bonn, 3.3.2006. [Online: [http://www.kmk.org/aktuell/pm060303\\_desi.htm](http://www.kmk.org/aktuell/pm060303_desi.htm). 2.10.2008].
- Liskin-Gasparro, Judith (2003), The ACTFL Proficiency Guidelines and the Oral Proficiency Interview: A brief history and analysis of their survival. *Foreign Language Annals* 36, 483-490.
- Magnan, Sally (1988), Grammar and the ACTFL Oral Proficiency Interview: Discussion and data. *Modern Language Journal* 72, 266-276.
- Majima, Junko (Hrsg.) (2006), Panel discussion with Brian North, Johanna Panthier, Neil Jones, Lynne Parmenter and Lukas Wertenschlag. *Osaka University of Foreign Studies Japan-Europe International Symposium. A new direction in foreign language education: The potential of the Common European Framework of Reference for Languages*. Osaka, 5. März 2006. [Online: <http://homepage.mac.com/jmajima1/bukosite/cef/Symposium.html>. 2.10.2008].
- Malone, Margaret (2003), Research on the Oral Proficiency Interview: Analysis, synthesis, and future directions. *Foreign Language Annals* 36, 491-497.
- Mikhailova, Julia (2006), Description in Russian: How the syntactical complexity of description in the OPI is different from description in the SOPI. *Working Papers in Slavic Studies*, vol. 6. Columbus: Ohio State University.
- Norris, John (1996), *A validation study of the ACTFL guidelines and the German Speaking Test*. University of Hawai'i at Manoa: Unpublished master's thesis.
- Ordinate (2007), *Versant for English Technical Manual*. Harcourt Assessment, Inc. [Online: [http://harcourtassessment.com/hai/Images/dotCom/VersantTest/Versant\\_English\\_TechManual.pdf](http://harcourtassessment.com/hai/Images/dotCom/VersantTest/Versant_English_TechManual.pdf). 2.10.2008].
- Rifkin, Benjamin (2002), A case study of the acquisition of narration in Russian: An investigation at the intersection of three disciplines. *Slavic and East European Journal* 46, 465-82.
- Rifkin, Benjamin (2003), Oral proficiency learning outcomes and curricular design. *Foreign Language Annals* 36, 582-588.
- Rubio, Fernando (2003), Structure and complexity of oral narratives in advanced-level Spanish: A comparison of three learning backgrounds. *Foreign Language Annals* 36, 546-554.
- Schneider, Günther (2001), Kompetenzbeschreibungen für das Europäische Sprachenportfolio. *Fremdsprachen Lehren und Lernen* 30, 193-214.
- Schröder, Konrad; Harsch, Claudia & Nold, Günter (2006), DESI. Die sprachpraktischen Kompetenzen unserer Schülerinnen und Schüler im Bereich Englisch. Zentrale Befunde. *Neusprachliche Mitteilungen* 03/2006, 11-32.
- Surface, Eric & Dierdorff, Erich (2003), Reliability and the ACTFL Oral Proficiency Interview: Reporting indices of interrater consistency and agreement for 19 languages. *Foreign Language Annals* 36, 507-519.

- Suzuki, Masanori & Harada, Yasunari (2004), A common testing framework for measuring spoken language skills of non-native speakers. *IWLeL 2004: An interactive workshop on language e-learning*, 115-122. [Online: <http://dSPACE.wul.waseda.ac.jp/dSPACE/bitstream/2065/1402/1/13.pdf>. 2.10.2008].
- Swender, Elvira (2003), Oral proficiency testing in the real world: Answers to frequently asked questions. *Foreign Language Annals* 36, 520-535.
- Tannenbaum, Richard & Wylie, E. Caroline (2005), Mapping English language proficiency test scores onto the Common European Framework. *TOEFL Research Reports* 80. Princeton, NJ: ETS. [Online: <http://www.ets.org/Media/Research/pdf/RR-05-18.pdf>. 2.10.2008].
- Tschirner, Erwin (1996), Scope and sequence: Rethinking beginning foreign language instruction. *Modern Language Journal* 80, 1-14.
- Tschirner, Erwin (2001), Die Evaluation fremdsprachlicher mündlicher Handlungskompetenz: Ein Problemaufriss. *Fremdsprachen Lehren und Lernen* 30, 87-115.
- Tschirner, Erwin (2005), Das ACTFL OPI und der Europäische Referenzrahmen. *Babylonia* 2/2005, 50-55.
- Tschirner, Erwin (2007a), The development of oral proficiency in a four-week intensive immersion program in Germany. *Die Unterrichtspraxis / Teaching German* 40, 111-117.
- Tschirner, Erwin (2007b), The Common European Framework of Reference and the ACFTL Oral Proficiency Guidelines: A comparison. Paper presented at the IES Faculty Seminar: Language Placement and Assessment, Granada, Spain, April 19-22, 2007.
- Tschirner, Erwin & Heilenman, L. Kathy (1998), Reasonable expectations: Oral proficiency goals for intermediate students of German. *Modern Language Journal* 82, 147-158.
- University of Cambridge ESOL Examinations. First Certificate of English. [Online: <http://www.cambridgeesol.org/exams/general-english/fce.html>. 20.10.2008].
- Vandergrift, Laurens (2006), Proposal for a Common Framework of Reference for Languages for Canada. Canadian Heritage. [Online: [http://www.pch.gc.ca/progs/lo-ol/pubs/new-nouvelles\\_perspectives/tm\\_e.cfm](http://www.pch.gc.ca/progs/lo-ol/pubs/new-nouvelles_perspectives/tm_e.cfm). 2.10.2008].
- Watanabe, Suwako (1998), Concurrent validity and application of the ACTFL Oral Proficiency Interview in a Japanese language program. *The Journal of the Association of Teachers of Japanese* 32/1, 22-38.
- Watanabe, Suwako (2003), Cohesion and coherence strategies in paragraph-length and extended discourse in Japanese Oral Proficiency Interviews. *Foreign Language Annals* 36, 555-565.
- de Wet, Febe; van der Walt, Christa & Niesler, Thomas (2007), Automatic large-scale oral language proficiency assessment. *Proceedings of the eighth annual conference of the International Speech Communication Association (Interspeech)*, Antwerp, Belgium, August 2007, 218-221. [Online: [http://www.dsp.sun.ac.za/~trn/reports/dewet+vanderwalt+niesler\\_interspeech07.pdf](http://www.dsp.sun.ac.za/~trn/reports/dewet+vanderwalt+niesler_interspeech07.pdf). 2.10.2008].

## Anlage 1: Gegenüberstellung ausgewählter Deskriptoren des Gemeinsamen europäischen Referenzrahmens (GeR) und der Oral Proficiency Guidelines (ACTFL)

GeR			ACTFL
A1	Has a very basic repertoire of words and simple phrases related to personal details and particular concrete situations. Shows only limited control of a few simple grammatical structures and sentence patterns in a memorized repertoire. Can ask and answer questions about personal detail.	Conversation is restricted to topics such as basic personal information, basic objects and a limited number of activities, preferences and immediate needs. NH speakers respond to simple, direct questions or requests for information. They are able to express personal meaning by relying heavily on learned phrases or recombinations of these.	NH
A2	Uses basic sentence patterns with memorised phrases to communicate limited information in simple everyday situations. Can make self understood in very short utterances, even though pauses, false starts and reformulation are very evident. Can answer questions and respond to simple statements.	Able to handle successfully a limited number of uncomplicated communicative tasks by creating with the language in straightforward social situations. Able to express personal meaning by combining and recombining into short sentences what they know. Speech is characterized by frequent pauses, ineffective reformulations and self-corrections.	IL
A2+	Can initiate, maintain and close simple, restricted face-to-face conversation, asking and answering questions on topics of interest, pastimes and past activities. Can interact with reasonable ease in structured situations but participation in open discussion is fairly restricted.	Able to handle successfully a variety of uncomplicated communicative tasks expressing personal information covering self, family, home, daily activities, interests and personal preferences as well as physical and social needs. Able to ask a variety of questions to obtain simple information to satisfy basic needs.	IM

B1	Has enough language to get by, to express self with some hesitation and circumlocution on topics such as family, hobbies and interests, work, travel, and current events. Can link a series of shorter discrete simple elements into a connected linear sequence of points.	Able to handle successfully many uncomplicated tasks and social situations requiring an exchange of basic information related to work, school, recreation, particular interests and areas of competence. Able to narrate and describe using connected discourse of paragraph length with some consistency.	IH
B1+	Can exploit a basic repertoire of strategies to keep a conversation going. Can give brief comments on others' views during discussion. Has a sufficient range of language to describe unpredictable situations. Generally good control of structures though with noticeable mother tongue influences.	Able to handle a variety of communicative tasks, although somewhat haltingly at times. Demonstrate the ability to narrate and describe in all major time frames but control of aspect may be lacking at times. Structure of the dominant language is still evident.	AL
B2	Able to give clear descriptions and express viewpoints on most general topics using some complex sentence forms to do so. Can use a limited number of cohesive devices to link utterances into clear, coherent discourse. Shows a relatively high degree of grammatical control. Does not make errors that cause misunderstanding.	Able to handle with ease and confidence a large number of communicative tasks. Able to narrate and describe in all major time frames by providing a full account, with good control of aspect. Speech is marked by substantial flow with much accuracy, clarity and precision. Intended message is conveyed without misrepresentation.	AM
C1	Has a good command of a broad range of language to express herself clearly in an appropriate style on a wide range of general, academic or professional topics. Can produce clear, well-structured speech, showing controlled use of organizational patterns, connectors and cohesive devices. Errors are rare.	Able to consistently explain in detail and narrate fully and accurately with linguistic ease, confidence and competence. Can provide a structured argument to support their opinions and they may construct hypotheses, but patterns of error appear. Can discuss some topics abstractly but are more comfortable	AH

		discussing a variety of topics concretely.	
C2	Can create coherent and cohesive discourse making full and appropriate use of a variety of organizational patterns and a wide range of connectors and other cohesive devices. Maintains consistent grammatical control of complex language, even when attention is otherwise engaged.	Explain their opinions on a number of topics such as social and political issues, and provide structured argument to support their opinions. Able to construct and develop hypotheses to explore alternative possibilities. Demonstrate virtually no pattern of error in the use of basic structures.	S

## Anlage 2: Vereinfachte GeR-Deskriptoren zur Normierung des *PhonePass Set-10-Tests*

C2	Can express him/herself spontaneously at length with a natural colloquial flow. Consistent grammatical and phonological control of a wide range of complex language, including appropriate use of connectors and other cohesive devices.
C1	Can express him/herself fluently and spontaneously, almost effortlessly, with a smooth flow of language. Clear, natural pronunciation. Can vary intonation and stress for emphasis. High degree of accuracy; errors are rare. Controlled use of connectors and cohesive devices.
B2	Can produce stretches of language with a fairly even tempo; few noticeably long pauses. Clear pronunciation and intonation. Does not make errors which cause misunderstanding. Clear, coherent, linked discourse, though there may be some "jumpiness".
B1	Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair may be very evident. Pronunciation is intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur. Reasonably accurate use of main repertoire associated with more predictable situations. Can link discrete, simple elements into a connected sequence.
A2	Can make him/herself understood in very short utterances, even though pauses, false starts and reformulation are very evident. Pronunciation is generally clear enough to be understood despite a noticeable foreign accent. Uses some simple structures correctly, but still systematically makes basic mistakes. Can link groups of words with simple connectors like "and", "but" and "because".
A1	Can manage very short, isolated, mainly pre-packaged utterances. Much pausing to search for expressions, to articulate less familiar words. Pronunciation is very foreign.

# Selbstreflexion und Evaluierung von Fremdsprachenlernprozessen in der Projektarbeit

Sabine Hoffmann<sup>1</sup>

The relationship between awareness and learning German was examined in project work undertaken with students of the degree course in International Media Studies at the University of Palermo. For many years now reflection upon and evaluation of learning processes have been regarded as an integral part of foreign language teaching. This article gives a detailed and differentiated account of research carried out over the last ten years into the role of awareness in second and foreign language learning. Focusing on this aspect of consciousness in Richard Schmidt's definition and re-examining it in the light of Klaus Holzkamp's learning theory, a theoretical approach is worked out which links the concept of awareness at its three levels to interest in foreign language learning. This concept underlies the qualitative longitudinal research carried out in the academic year 2006/07, from which two cases and a decisive phase of the project have been chosen in order to exemplify individual levels of consciousness in relation to group learning of German.<sup>2</sup>

Selbstreflexion und Evaluierung sind Bestandteile der Konzepte, mit denen sich die Fremdsprachenforschung seit über einem Jahrzehnt auseinandersetzt und die mittlerweile auch konkrete Wirkungen auf die Unterrichtspraxis zeigen. Mit der Frage, inwieweit sie auch in der Projektarbeit umgesetzt werden können, befasst sich der vorliegende Beitrag, der auf Ergebnisse einer empirischen Untersuchung im Fach *Deutsch als Fremdsprache* an einer süditalienischen Universität rekurriert. Nach einem Überblick über den Forschungsstand werden die erkenntnistheoretischen Grundlagen der Studie und deren Voraussetzungen erläutert, um daran anschließend an zwei Einzelfällen Stufen von Bewusstmachung von Lernprozessen in der Gruppe zu exemplifizieren.

## 1. Reflexion im Spannungsfeld zwischen individuellem und kollektivem Lernen

### 1.1 Sprach(lern)bewusstheit als Forschungsgegenstand

Die zunehmende Relevanz der Kognitionswissenschaften und die Tendenz, die in diesem Bereich gewonnenen Erkenntnisse auch auf andere Disziplinen auszuweiten, lenkte in den 90er Jahren die Aufmerksamkeit der Fremdsprachenfor-

1 Korrespondenzadresse: Dr. Sabine Hoffmann, Via Domenico Scinà 105, I-90139 Palermo, E-Mail: sabinehoffmann@katamail.com

2 Der vorliegende Aufsatz entspricht in Teilen dem gleichnamigen Vortrag auf dem DGFF-Kongress in Gießen 2007.