

# **Konsortium HarmoS Fremdsprachen**

## **FREMDSPRACHEN**

### **Wissenschaftlicher Kurzbericht und Kompetenzmodell**

Provisorische Fassung (vor Verabschiedung der Standards)

Stand : 18. Juni 2009

#### **Wissenschaftliches Konsortium:**

Universität de Fribourg (Leading house) | Günther Schneider, Peter Lenz, Thomas Studer  
(Hauptverantwortliche des Kurzberichts)

Pädagogische Hochschule Zürich | Sandra Hutterli, Ruth Keller

Universität Bern | Verena Tunger

Institut de Recherche et de Documentation | Claudine Brohy

Pédagogique (IRDP), Neuchâtel | Daniel Elmiger

Haute Ecole pédagogique de Fribourg | Christine Müller

Pädagogische Hochschule Zentralschweiz Luzern | Monika Mettler

# Inhaltsverzeichnis

1	Ausgangslage .....	3
1.1	Vorbemerkung .....	3
1.2	Auftrag .....	3
1.3	Das Konsortium – Zusammensetzung und Arbeitsweise .....	3
2	Grundlagen .....	4
2.1	Kompetenzmodell .....	4
2.2	Überprüfung des Konstrukts (Methodologie) .....	10
3	Die empirischen Untersuchungen in HarmoS .....	13
3.1	Überblick .....	13
3.2	Umfang der Untersuchungen .....	14
3.3	Testinstrumente .....	16
3.3.1	Grundsätzliche Überlegungen zur Überprüfung von Kompetenzen .....	16
3.3.2	Zur Testentwicklung in HarmoS .....	19
3.4	Hörverstehen .....	21
3.5	Leseverstehen .....	24
3.5.1	Konzeptuelle Grundlagen und Konstrukt .....	24
3.5.2	Instrumente .....	28
3.5.3	Datenerhebung .....	33
3.5.4	Korrektur .....	34
3.5.5	Ergebnisse und Analyse der Ergebnisse .....	34
3.5.6	Bilanz und Ausblick .....	42
3.5.7	Zusammenfassung .....	43
3.6	Schreiben .....	44
3.6.1	Ausgangspunkt .....	44
3.6.2	Instrumente (Aufgaben) .....	45
3.6.3	Datenerhebung .....	48
3.6.4	Rating .....	49
3.6.5	Ergebnisse der Untersuchung .....	51
3.6.6	Analysen zu den Ergebnissen .....	55
3.6.7	Stand der Arbeit und Ausblick .....	59
3.7	Sprechen .....	60
3.7.1	Ausgangspunkt und Grundlagen .....	60
3.7.2	Instrumente (Aufgaben) .....	61
3.7.3	Datenerhebung .....	66
3.7.4	Rating .....	66
3.7.5	Analysen und Ergebnisse .....	68
3.7.6	Ausblick: Überprüfung der Sprechkompetenz im Bildungsmonitoring .....	71
4	Kompetenzbereiche, die nicht empirisch untersucht wurden .....	71
4.1	Sprachmittlung .....	71
4.2	Überfachliche Kompetenzen .....	74

4.2.1	Allgemeines .....	74
4.2.2	Kulturelle und interkulturelle Kompetenzen .....	75
4.2.3	Methodische Kompetenzen .....	82
4.2.4	Literatur zu den überfachlichen Kompetenzen .....	86
5	Die Expertenvorschläge für Basisstandards .....	91
6	Literatur.....	93

# 1 Ausgangslage

## 1.1 Vorbemerkung

Das vorliegende Dokument will einen Gesamtüberblick über das Projekt HarmoS Fremdsprachen vermitteln. Der Schwerpunkt dieses eher wissenschaftlich ausgerichteten Berichts liegt auf den empirischen Untersuchungen, die im Rahmen des Projekts zu den Kompetenzbereichen Leseverstehen, Schreiben und Sprechen durchgeführt wurden (Kap. 3) sowie auf Überlegungen zur Sprachmittlung und zu überfachlichen Kompetenzen (Kap. 4). Eingerahmt werden diese Kapitel durch eine Skizze zur Ausgangslage des Projekts (Kap. 1), zu den Grundlagen der Untersuchungen (Kompetenzmodell und Methodologie, Kap. 2) sowie durch eine kurze Schilderung des Vorgehens beim Setzen der Basis-Bildungsstandards (Kap. 5).

## 1.2 Auftrag

Das Konsortium HarmoS Fremdsprachen (FS) hat von der EDK den Auftrag übernommen, für die wichtigsten Fremdsprachenfächer in der obligatorischen Schule (d. h. Französisch und Englisch in der Deutschschweiz, sowie Deutsch und Englisch in der Westschweiz) Expertenvorschläge für Bildungsstandards auszuarbeiten, die sich auf zu erreichende Basiskompetenzen am Ende der Primarstufe einerseits und am Ende der Sekundarstufe I andererseits beziehen. Diese Vorschläge sollen auf einem didaktisch und, soweit möglich, empirisch validierten Kompetenzmodell gründen. Die empirische Validierung sollte auch genutzt werden, um gesicherte Anhaltspunkte darüber zu gewinnen, was heutige Schülerinnen und Schüler, die sich an diesen Nahtstellen befinden (also Sechst- und Neuntklässler/innen) leisten können.

Die Situation des Konsortiums Fremdsprachen ist insofern speziell, als die Basisstandards für künftige Schülerinnen und Schüler festgelegt werden sollen, die unter veränderten Bedingungen Fremdsprachen gelernt haben werden. Die Umsetzung der neuen Szenarien für den Fremdsprachenunterricht gemäss Beschluss der Erziehungsdirektorenkonferenz vom 25. März 2004 steht nämlich grösstenteils erst noch bevor. In diesem Beschluss ist insbesondere der Beginn des Fremdsprachenunterrichts geregelt: Spätestens in der heutigen dritten Klasse soll mit der ersten Fremdsprache begonnen werden und spätestens in der fünften Klasse mit der zweiten Fremdsprache. Eine dieser Sprachen soll eine Landessprache der Schweiz sein, die andere Englisch. Am Ende der obligatorischen Schulzeit sollen die Schülerinnen und Schüler beide Fremdsprachen vergleichbar gut können. Gleichzeitig mit der Einführung von zwei Fremdsprachen auf der Primarstufe wird eine Erneuerung des Fremdsprachenunterrichts angestrebt. Zu diesem Zweck werden auch neue Lehrpläne und neue Unterrichtsmaterialien entwickelt<sup>1</sup>. Über die Auswirkungen der Neuerungen – was werden die Schüler/innen, die unter den neuen Bedingungen gelernt haben, an den Nahtstellen der obligatorischen Schulzeit in den Schulfremdsprachen leisten können? – können zurzeit nur hypothetische Annahmen gemacht werden, die sich auf Forschungsliteratur, Expertenkonsens und Extrapolierungen von Ergebnissen stützen, die im gegenwärtigen Fremdsprachenunterricht erhoben wurden.

## 1.3 Das Konsortium – Zusammensetzung und Arbeitsweise

Das Lern- und Forschungszentrum Fremdsprachen der Universität Freiburg i. Üe. ist Vertragspartner der EDK und so genanntes *Leading House* des Fremdsprachenkonsortiums (Projektverantwortung: Günther Schneider; Projektleitung: Peter Lenz, Thomas Studer; Sekretariat: Eva Wieden Keller). Partner sind – in der Reihenfolge ihrer Beteiligung – die Pädagogische Hochschule Zürich (Sandra Hutterli, Ruth Keller), die Universität Bern (Verena Tunger), das Institut de recherche et de documentation pédagogique (IRDP) in Neuchâtel (Daniel Elmiger, zeitweise auch Claudine Brohy), die Pädagogische Hochschule Freiburg (Christine Müller-Tragin), sowie die Pädagogische Hochschule Zentralschweiz in Luzern (Monika Mettler). Einen Beobachterstatus hatte die Pädagogische

---

<sup>1</sup> zu den je verschiedenen Zielen und zum Zusammenspiel von Standards, Lehrplänen und Lehrmitteln vgl. z. B. Caspari 2005

Fachhochschule Graubünden in Chur (Luigi Menghini). Die Mitarbeit am Projekt wurde nur teilweise durch das Projektbudget der EDK finanziert; die genannten Institutionen mussten beträchtliche Mittel selbst aufbringen.

Das Konsortium wurde von einem Fach-Beirat begleitet, der mehrmals zu zentralen Entscheidungen konsultiert wurde, insbesondere auch bei der Festlegung der Anforderungen, welche die Basisstandards stellen sollten. Mitglieder des Beirats waren Bruno Dahinden, Kt. Thurgau; Mirjam Egli, EDK und Uni BS; Simona Pekarek, Uni NE; Michel Nicolet, Sekretär der CIIP; Véronique Roncoroni-Arletta, später Brigitte Jörimann Vancheri, Kt. TI; Victor Saudan, Sprachenverantwortlicher der NW EDK; Daniel Stotz, PHZH; Iwar Werlen, Uni BE; Susanne Wokusch, HEP VD und UNIL.

Für die Durchführung der Tests zum Sprechen (mündliche Produktion und Interaktion) sowie die Korrektur und Beurteilung der Schülerleistungen wurden temporäre Hilfskräfte angestellt. Anlässlich der Entwicklung der Testaufgaben und bei der Festlegung der Basisstandards haben einige Fachdidaktikerinnen und Fachdidaktiker, die weder am Konsortium beteiligt noch Mitglieder des Beirats waren, als Gutachter bzw. Diskussionspartner mitgewirkt. Die Vorerprobung der Testaufgaben und die Durchführung der Feldtests zum Sprechen waren nur mit der Unterstützung von zahlreichen Lehrerinnen und Lehrern und deren Schulleitungen und Klassen möglich.

Die verschiedenen Mitarbeiterinnen und Mitarbeiter des Konsortiums setzten ihre Arbeitsschwerpunkte primär hinsichtlich der Kompetenzbereiche (v.a. Lesen, Schreiben, Sprechen), die überprüft und beschrieben werden sollten. Es gab also zum Beispiel keine regionalen Untergruppen oder solche, die sich nur mit einer der drei Zielsprachen beschäftigten. Die Projektleitung arbeitete auch in der Strategiegruppe der EDK mit und traf ihre Entscheidungen in Bezug auf die Hauptuntersuchung in Zusammenarbeit der HarmoS-Methodologiegruppe. Die Mitglieder dieser Gruppe waren auch hauptsächlich für die statistische Auswertung der Daten zuständig. Die Zusammenarbeit verlief insgesamt gut. In der Anfangsphase machte sich der Umstand negativ bemerkbar, dass die Methodologiegruppe zu spät ins Spiel gebracht wurde und sich selbst zuerst auch konstituieren musste. In der Phase vor dem offiziellen Abgabetermin für die Berichte der Konsortien (Dezember 2007) fehlte dann die Zeit für verschiedene Analysen im Zusammenhang mit der Auswertung der Hauptuntersuchung, weil die Methodologiegruppe alle Konsortien parallel bedienen musste. Im Jahr 2008 holte das Konsortium Fremdsprachen einen Teil dieser Analysen dann in Eigenverantwortung nach. Als grosser Zeitverlust erwies sich für das Konsortium Fremdsprachen letztlich die vielversprechende Idee der HarmoS-Projektleitung vom Frühjahr 2006, die Hauptuntersuchung erst ein Jahr später durchzuführen, dafür aber mehr vertiefende Untersuchungen zu ermöglichen. Daraufhin stellte das Konsortium Fremdsprachen anstehende Arbeiten zurück und arbeitete eine neue Planung aus, die aber im Herbst 2006 von den Erziehungsdirektoren dann doch abgelehnt wurde. Dies war nicht nur aus zeitlich-planerischer, sondern auch aus inhaltlicher Sicht sehr bedauerlich, denn ergänzende qualitative Untersuchungen zu den Tests, auf die wegen der Ablehnung der neuen Planung verzichtet werden musste, hätten vermutlich Wesentliches zur Validierung des Kompetenzmodells und der Tests, mit denen in Zukunft Standards überprüft werden sollen, beitragen können (vgl. Banerjee 2004). Für die Auftraggeberschaft überwog aber offenbar das Argument, dass eine repräsentative Untersuchung zum heutigen Zeitpunkt auch mit einem verfeinerten Validierungskonzept keine exakten Anhaltspunkte für die Kompetenzen der zukünftigen Schülerinnen und Schüler, die im Rahmen der neuen Szenarien lernen werden, liefern könnte.

Eine Konsequenz dieser Sichtweise – und auch der Zeitknappheit, die sich eingestellt hatte – war, dass die Fremdsprachen in den repräsentativen Hauptuntersuchungen in der Deutschschweiz und in der Westschweiz schliesslich nur mit zwei Kompetenzbereichen vertreten waren: Leseverstehen und Schreiben. Allerdings hat das Generalsekretariat der EDK für einen späteren Zeitpunkt, an dem die neuen Unterrichtsszenarien gegriffen haben werden, die Möglichkeit einer weiteren grösseren Untersuchung in Aussicht gestellt.

## **2 Grundlagen**

### **2.1 Kompetenzmodell**

Der Auftrag der EDK an die Konsortien sieht eine Verbindung von Kompetenzmodell und Standards vor. Es soll ein Kompetenzmodell entwickelt und (teilweise) empirisch validiert werden, auf

dessen Grundlage die Basis-Anforderungen an die Lernenden verschiedener Stufen bestimmt und kommuniziert werden können.

Die Erwartungen der EDK an ein Kompetenzmodell basieren vom Ansatz her auf der Expertise, die als Grundlagenpapier für Bildungsstandards der deutschen Kultusministerkonferenz (KMK) verfasst wurde (Klieme et al. 2003)<sup>2</sup>. In dieser Studie wird "Kompetenz" – in Anlehnung an Weinert – folgendermassen definiert:

In Übereinstimmung mit Weinert (2001, S. 27f.) verstehen wir unter Kompetenzen die bei Individuen vorhandenen oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können. (Klieme et al. 2003: 72 [im Original fett])

Kompetenz wird damit im Wesentlichen als (kognitive) Problemlösungsfähigkeit verstanden, die in Abhängigkeit von bestimmten nicht-kognitiven "Bereitschaften und Fähigkeiten" eingesetzt wird. Der Kompetenzbegriff, auf den sich Klieme et al. berufen, ist recht nah am Modell des *Gemeinsamen europäischen Referenzrahmens für Sprachen* (Europarat 2001), auf das sich das Konsortium Fremdsprachen primär bezieht. Im Anhang zur Klieme-Studie selbst (Klieme et al. 2003: 146-156) werden das Modell des *Referenzrahmens* und die dazugehörigen Kompetenzbeschreibungen (Deskriptoren) als „eine gute Ausgangsbasis für die Entwicklung von Aufgaben, Tests und anderen Prüfungsverfahren“ vorgestellt.

Das Konsortium Fremdsprachen hat den grossen Vorteil, dass es sich bei seinen Entwicklungsarbeiten bereits auf ein Kompetenzmodell beziehen kann, das im „*Gemeinsamen europäischen Referenzrahmen*“ selbst und in verschiedenen, darauf aufbauenden Arbeiten teilweise ausführlich spezifiziert ist und zudem in unterschiedlichen Kreisen eine recht grosse Akzeptanz genießt. Im schulischen Bereich ist dieses Kompetenzmodell (ansatzweise) vor allem durch das *Europäische Sprachenportfolio*, aber auch durch einige neuere kantonale oder regionale Lehrpläne bekannt geworden.

Die Frage, ob das Modell des *Referenzrahmens* überhaupt als Ausgangspunkt der Arbeiten genommen werden soll, stellt sich in HarmoS auch deshalb nicht, weil dieses (oder zumindest Teile davon) in der Fremdsprachendidaktik und besonders auch im Bereich des Sprachentestens (und der bekannten Sprachdiplome) eine so grosse Verbreitung erfahren hat, dass ein Verzicht auf diesen Bezug geradezu irritieren müsste. Die Erziehungsdirektorenkonferenz als Auftraggeber erwartet zudem, dass sich die HarmoS-Standards explizit auf den *Referenzrahmen* und die im Rahmen von *IEF/lingualevel* (Lenz/Studer 2007) geschaffenen zielgruppenspezifischen Differenzierungen (Altersgruppe der 11-16-Jährigen) und Ergänzungen bezieht. Hinzu kommt, dass das *Leading House* seinerzeit an der Entwicklung der Kompetenzbeschreibungen der Referenzniveaus und an wichtigen Weiterentwicklungen auf der Basis des *Referenzrahmens* (besonders *IEF/lingualevel*, aber auch Benchmarkings von Referenzleistungen) ganz wesentlich beteiligt war und das zugrunde liegende Modell inhaltlich mitträgt.

Sprachkompetenz wird im *Referenzrahmen* in erster Linie gesehen als Sprachverwendungskompetenz, die Fähigkeit also, unter Einbezug von Sprache zu handeln und dabei Aufgaben unterschiedlicher Art zu bearbeiten bzw. zu lösen. Sprache steht zwar in diesem Modell im Mittelpunkt, dies aber nicht in einem reduktionistischen Sinn: Sprachverwendungskompetenz wird als eine umfassende menschliche Fähigkeit gezeigt, die den „ganzen“ Menschen mit all seinen sprachlichen und nichtsprachlichen Ressourcen einbezieht, unabhängig davon, ob diese kognitiver oder anderer, z. B. affektiver, Natur sind.

Das folgende Schema gibt einen Überblick über das Sprachverwendungsmodell des *Europäischen Referenzrahmens*:

---

<sup>2</sup> Vgl. dazu jetzt auch die umfassende Expertise von Oelkers / Reusser (2008), die für diesen Bericht allerdings nicht mehr berücksichtigt werden konnte.

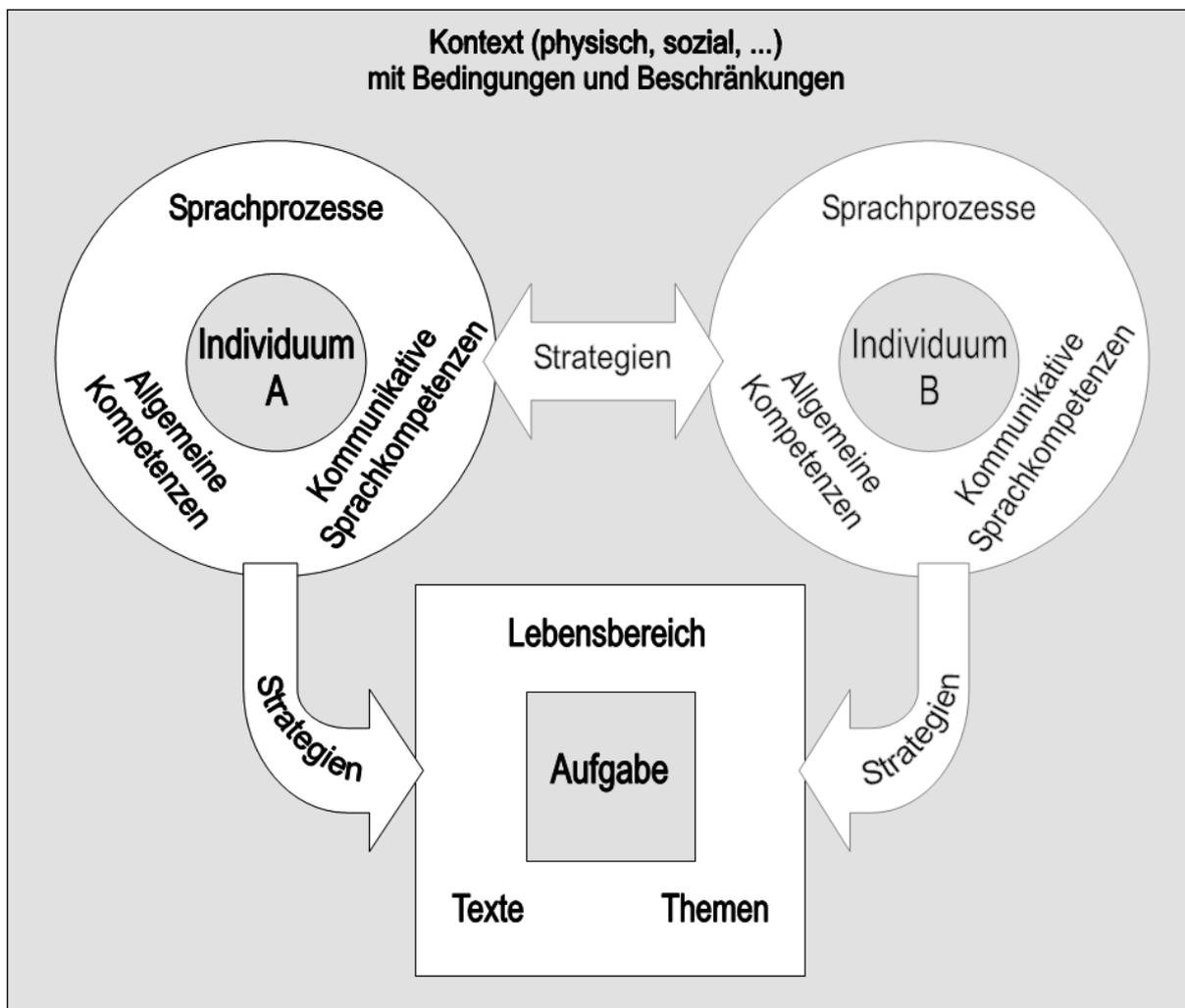


Abb. 1: Sprachverwendung gemäss *Europäischem Referenzrahmen*

Dieses Schema kann folgendermassen gelesen werden: Kommunikative Sprachverwendung findet unter den spezifischen Bedingungen des jeweiligen Kontexts statt. Die kommunizierenden Individuen verfügen über vielfältige Kompetenz-Ressourcen unterschiedlicher Art, die sich zu allgemeinen Kompetenzen und spezifischen kommunikativen Sprachkompetenzen gruppieren lassen. Wenn Menschen Sprache produzieren oder rezipieren, setzen kognitive Prozesse ein, die speziell geeignet sind, Sprache zu verarbeiten (Sprachprozesse). Strategien werden gebraucht, um die Sprachaktivitäten – Interaktion mit anderen und mündliche oder schriftliche Produktion und Rezeption von Texten – zu steuern, was mehr oder weniger bewusst geschieht. Sprachliche Kommunikationsaufgaben haben bestimmte Merkmale, durch die sie sich grob charakterisieren lassen: den Lebensbereich, in dem sie angesiedelt sind – Bildung (Schule), öffentliches Leben usw. –, die mündlichen oder schriftlichen Texte, die im Zusammenhang mit der Aufgabe bearbeitet werden müssen, die Themen, von denen diese Texte handeln. Bei interaktiven Sprachaktivitäten besteht ein wesentlicher Teil der Aufgabe in der Interaktion mit anderen beteiligten Individuen (z. B. in Diskussionen zur Entscheidungsfindung; in schriftlicher Korrespondenz). Viele rezeptive und produktive Sprachaktivitäten werden jedoch idealtypisch als Vorgänge gesehen, die sich zwischen einem Individuum und einer Aufgabe abspielen. Dabei sind andere Individuen nur indirekt, als sozialer Kontext, an der Bearbeitung der Aufgaben beteiligt (siehe Abbildung 1, grau schattiertes Individuum)<sup>3</sup>.

<sup>3</sup> Im *Referenzrahmen* selbst findet sich die folgende, sehr konzise Charakterisierung des Sprachverwendungsmodells bzw. des „handlungsorientierten Ansatzes“: „Sprachverwendung – und dies schließt auch das Lernen einer Sprache mit ein – umfasst die Handlungen von Menschen, die als Individuen und als gesellschaftlich Handelnde eine Vielzahl von *Kompetenzen* entwickeln, und zwar *allgemeine*, besonders aber *kommunikative Sprachkompetenzen*. Sie greifen in verschiedenen *Kontexten* und unter verschiedenen *Bedingungen und Beschränkungen* auf diese Kompetenzen zurück, wenn sie *sprachliche Aktivitäten* ausführen, an denen (wiederum) *Sprachprozesse* beteiligt sind, um *Texte* über

Die Kompetenzen, die Sprachverwendungskompetenz ausmachen bzw. zur Sprachverwendungskompetenz beitragen, werden im *Referenzrahmen* unter unterschiedlichen Gesichtspunkten und in unterschiedlicher Ausführlichkeit aufgeschlüsselt.

- 1) Das Kernstück der Kompetenzbeschreibung im *Referenzrahmen* bildet die Charakterisierung von **kommunikativer Handlungsfähigkeit** auf verschiedenen Niveaus mittels grösstenteils empirisch skaliertes<sup>4</sup> (d. h. Kompetenzniveaus zugeordneter) Beschreibungen von "kommunikativen Aktivitäten und Strategien" (Kap. 4.4). Sie illustrieren, was jemand typischerweise auf einem Niveau tun kann, welche *kommunikativen Sprachaktivitäten* er oder sie bewältigen kann. Der *Referenzrahmen* ordnet die kommunikativen Aktivitäten und Strategien vier verschiedenen sprachlichen Handlungsmodi zu (Rezeption, Interaktion, Produktion und Sprachmittlung<sup>5</sup>), wobei jeweils zwischen mündlichem und schriftlichem Sprachgebrauch (Kanal) unterschieden wird.

		Kommunikationsmodus			
		Produktion	Interaktion	Rezeption	Mediation
Kanal	mündlich	■	■	■	■ ■ ■
	schriftlich	■	■	■	■ ■ ■

Abb. 2: Ordnungssystem für sprachliche Aktivitäten nach Kommunikationsmodus und Kanal (in Anlehnung an den *Europäischen Referenzrahmen*)

„Rezeption schriftlich“ in dieser Darstellung entspricht beispielsweise der kommunikativen Sprachfertigkeit Leseverstehen. Insgesamt gibt es im Referenzrahmen über 50 Skalen für kommunikative Aktivitäten und Strategien.

- 2) Weiter werden im *Referenzrahmen* die **Kompetenzressourcen** beschrieben (Kap. 5), die nötig sind, um kommunikative Sprachaktivitäten auszuführen. Dabei wird zwischen allgemeinen und spezifisch sprachlichen Kompetenzen unterschieden:

a) Allgemeine Kompetenzen:

- Wissen (*savoirs*): Weltwissen (z. B. Geschichte, Geografie); soziokulturelles Wissen (z. B. über den Alltag; das kulturelle Leben; Werte; interpersonale Beziehungen; die Körpersprache); interkulturelles Bewusstsein (z. B. bez. Ähnlichkeiten und Unterschieden zwischen Gemeinschaften);
- Können (*savoir-faire*): praktische Fertigkeiten (z. B. Verhalten gemäss den gesellschaftlichen Konventionen; berufliche Fertigkeiten); interkulturelle Fertigkeiten (z. B. zwei Kulturen zueinander in Beziehung setzen können);
- Persönlichkeitsbezogene Kompetenz (*savoir-être*): z. B. Motivation; kognitiver Stil; Werte;
- Lernfähigkeit (*savoir-apprendre*): z. B. Sprach- und Kommunikationsbewusstsein; Lern-techniken; Fertigkeiten im Umgang mit Neuem.

---

bestimmte *Themen* aus verschiedenen *Lebensbereichen* (Domänen) zu produzieren und/oder zu rezipieren. Dabei setzen sie *Strategien* ein, die für die Ausführung dieser *Aufgaben* am geeignetsten erscheinen. Die Erfahrungen, die Teilnehmer in solchen kommunikativen Aktivitäten machen, können zur Verstärkung oder zur Veränderung der Kompetenzen führen." (Europarat 2001: 21)

<sup>4</sup> Diese Entwicklungsarbeiten wurden hauptsächlich im Rahmen des Nationalen Forschungsprogramms zur „Wirksamkeit unserer Bildungssysteme“ geleistet. Die Projektverantwortung lag beim heutigen Lern- und Forschungszentrum Fremdsprachen der Universität Freiburg (LeFoZeF) (vgl. z. B. North & Schneider 1998; Schneider & North 2000; North 2000). (Das LeFoZeF ist inzwischen selbst wieder Geschichte. Die neue Struktur heisst Departement für Mehrsprachigkeits- und Fremdsprachenforschung der Universität Freiburg.)

<sup>5</sup> Grundsätzlich geht es bei Sprachmittlungsaktivitäten darum, Mitteilungen, die man in einer Sprache gehört oder gelesen hat, für jemanden, der diese Sprache nicht versteht, in einer anderen für ihn verständlichen Sprache mündlich oder schriftlich weiterzugeben. Eigene Gedanken und Absichten stehen dabei im Hintergrund.

b) Kommunikative Sprachkompetenzen (im engeren Sinn):

- Linguistische Kompetenzen: z. B. lexikalische; grammatische; phonologische;
- Soziolinguistische Kompetenzen: z. B. sprachliche Kennzeichnung sozialer Beziehungen; Höflichkeitskonventionen; Verwendung von unterschiedlichen sprachlichen Registern; Erkennen von unterschiedlichen Dialekten und Akzenten;
- Pragmatische Kompetenzen: z. B. Diskurskompetenz; funktionale Kompetenz; Fähigkeit, unterschiedliche Interaktionsschemata zu verwenden; wichtige qualitative Aspekte von pragmatischer Kompetenz sind Flüssigkeit und Genauigkeit des Ausdrucks.

Auf diese Kompetenzressourcen wird in unterschiedlicher Auswahl und in unterschiedlichem Grad bei allen sprachlichen Aktivitäten zurückgegriffen. Umgekehrt lösen wohl alle Erfahrungen beim sprachlichen Handeln Lernprozesse aus und beeinflussen so wiederum Umfang und Qualität der Kompetenzressourcen.

Ähnlich wie zu der kommunikativen Handlungsfähigkeit gibt es im *Referenzrahmen* auch zu den kommunikativen Sprachkompetenzen eine Reihe von empirisch ermittelten Skalen, die dem Niveausystem A1 bis C2 zugeordnet sind, darunter zum Beispiel solche zu Wortschatzspektrum, grammatischer Korrektheit, Flexibilität und Flüssigkeit.

- 3) Als weiteren, im Grunde den Kompetenzressourcen übergeordneten, Kompetenzbereich des *Referenzrahmens* kann man die **Strategien der Handlungssteuerung** sehen, welche den Zugriff auf die verschiedenen Kompetenzressourcen steuern, sodass konkrete Kommunikationsaufgaben gelöst werden können (Strategien für Planung, Ausführung, Kontrolle und evtl. Reparatur von Handlungen; vgl. dazu Bachman & Palmer 1996: 63).

Das Beschreibungssystem des *Referenzrahmens* berücksichtigt ausdrücklich, dass Entwicklung und Verwendung der kommunikativen Kompetenz nicht nur sprachliche Dimensionen implizieren, sondern beispielsweise auch soziokulturelle Bewusstheit, interkulturelle Fertigkeiten, Erfahrung und Phantasie, Motivation, affektive Beziehungen, Lernenlernen usw. „Kompetenzen“ werden definiert als „Summe des (deklarativen) Wissens, der (prozeduralen) Fertigkeiten und der persönlichkeitsbezogenen Kompetenzen und allgemeinen kognitiven Fähigkeiten [den Strategien], die es einem Menschen erlauben, Handlungen auszuführen“ (Europarat 2001: 21).

Für zahlreiche Bereiche fehlen im *Referenzrahmen* die Skalen, zum Beispiel solche zum soziokulturellen Wissen, zu interkulturellen Fertigkeiten, zu Lernstrategien, aber auch solche zu bestimmten Bereichen der kommunikativen Handlungsfähigkeit: zur Sprachmittlung, zur Lektüre literarischer Texte oder auch zum Telefonieren. Das bedeutet aber eben keineswegs, wie teilweise kurzschlüssig angenommen wird, dass solche Bereiche für den *Referenzrahmen* unwichtig wären.

Bei einigen der Bereiche, für die Skalen fehlen, spricht nichts Grundsätzliches gegen eine Skalenbildung im Sinne der existierenden Sprachkompetenzskalen (zum Beispiel bei der Sprachmittlung oder anderen Bereichen der kommunikativen Handlungsfähigkeit). Andere Kompetenzaspekte dürften kaum skalierbar sein, jedenfalls nicht parallel zu den Dimensionen kommunikativer Handlungsfähigkeit (v.a. die „allgemeinen Kompetenzen“).

Die nicht skalierbaren Kompetenzen der kommunizierenden Menschen können im *Referenzrahmen* nicht aus dem Modell herausfallen, weil sie letztlich bei jedem kommunikativen Handeln auch im Spiel sind und zum Gelingen oder Scheitern der Kommunikation beitragen. Offensichtlich grosse Bedeutung haben nicht-skalierbare Kompetenzen beim ausserschulischen Handeln und zum Beispiel auch für den Erfolg bei Sprachtests.

Ursprünglich bestand im Rahmen von HarmoS Fremdsprachen die Absicht, das Kompetenzmodell im Bereich der Sprachmittlung, der (inter-)kulturellen Kompetenzen und der Mehrsprachigkeitskompetenz weiterzuentwickeln, weil es sich dabei um Bereiche handelt, die für den schulischen Fremdsprachenunterricht zunehmend wichtiger werden.

Für die Sprachmittlung konnte vor allem konzeptuelle Arbeit geleistet werden; eine empirische Überprüfung dieser Ansätze war aber nicht möglich. Für wesentliche Weiterentwicklungen in den Bereichen (inter-)kulturelle Kompetenz und Mehrsprachigkeitskompetenz wären eigenständige Projekte nötig gewesen (z. B. Nationalfondsprojekte im Rahmen des NFP 56). Es kamen aber keine solchen zustande. Immerhin konnten aber für (inter-)kulturelle und auch für methodische

Kompetenzen Vorarbeiten geleistet werden, die in allfälligen HarmoS-Folgeprojekten sowie in der Lehrplan- und Lernmittelarbeit genutzt werden können (vgl. Kap. 4.2, überfachliche Kompetenzen).

Als Basis für die Arbeiten in HarmoS wurden neben dem *Referenzrahmen* selbst auch Weiterentwicklungen und Konkretisierungen des *Referenzrahmens* einbezogen. Dies betrifft insbesondere:

- Ergebnisse des Deutschschweizer Projekts *IEF* (Verlagsprodukt *lingualevel*), v.a. die Beschreibungen von kommunikativer Handlungsfähigkeit, die an die Bedürfnisse und Möglichkeiten von ca. 11- bis 16-jährigen Schülerinnen und Schülern angepasst wurden; zudem ergänzende Kriterien zur (feineren) Beschreibung der sprachlichen Merkmale von mündlichen und schriftlichen Leistungen (Lenz / Studer 2007; 2004);
- auf dem *Referenzrahmen* aufbauende und ihn ergänzende Materialien aus Projekten des Europarats, z. B. das Handbuch für die Zuordnung von Sprachprüfungen (Council of Europe 2003); eingestufte und kommentierte Beispiele von Aufgaben und Leistungen zum Lese- und Hörverstehen sowie zum Sprechen und Schreiben, usw.<sup>6</sup>;
- wissenschaftliche Arbeiten zum *Referenzrahmen*;
- das *Europäische Sprachenportfolio* und Arbeiten zum *Sprachenportfolio*.

Weitere wichtige Quellen waren eine von der EDK in Auftrag gegebene Analyse aller kantonalen Lehrpläne, neuere Fremdsprachenlehrwerke sowie theoretische und empirische Forschung zum Sprachenlernen und Sprachentesten.

Der Beitrag der Ergebnisse von HarmoS zur Kompetenzbeschreibung ist vor dem Hintergrund eines bereits gut ausgebauten Kompetenzmodells zu sehen. Durch die Tests neu hinzugekommen sind in erster Linie zahlreiche Beispiele von Schülerleistungen in den Kompetenzbereichen Sprechen (*Teilnahme an Gesprächen/Zusammenhängendes Sprechen*) und Schreiben, sowie Testaufgaben zum Leseverstehen, die den Niveaus zugeordnet werden konnten und nun die Niveaus und deren Anforderungen konkret illustrieren (vgl. das Vernehmlassungsdokument zu den Basisstandards in den Fremdsprachenfächern). Die Basis für die Niveaubeschreibungen und damit auch für die Beschreibung der Standards bilden weiterhin die früheren Entwicklungen; die Ergebnisse aus den HarmoS-Tests haben zu Modifikationen und Ergänzungen der existierenden Kompetenzbeschreibungen geführt und sind eine Quelle für die weitere Konkretisierung der Beschreibungen (vgl. besonders Kap. 2.1 „Kompetenzmodell“ in diesem Bericht. ).

In diesem Zusammenhang erscheint es sinnvoll, kurz auf grundsätzlich unterschiedliche Typen von Skalen und Beschreibungen zu verweisen (vgl. Europarat 2001: 46f.), die im Zusammenhang mit Standards und deren Überprüfung relevant sind:

- Benutzerorientierte Skalen: Sie beschreiben das typische Verhalten von Lernenden auf bestimmten Niveaus, indem sie vorwiegend positiv ausdrücken, was jemand tun kann.
- Beurteilungsorientierte Skalen: Sie formulieren typischerweise Aspekte der Qualität der erwarteten Leistung und lenken damit das Urteil von Beurteilenden.

Die Niveaubeschreibungen, die in HarmoS FS verwendet wurden, enthalten Elemente dieser beiden Skalentypen.

- Aufgabenorientierte Skalen (Skalen für Testautoren): Sie beschreiben üblicherweise kommunikative Aufgaben, die ziemlich direkt in Testaufgaben umzusetzen sind.
- Test- und Testaufgabenspezifikationen (für Testautoren): Diese technischen Spezifikationen regeln beispielsweise Testumfang und Testzusammensetzung, Optionen bezüglich des Aufgabenformats (welche offenen und/oder geschlossenen Formate?) und viele weitere Einzelheiten, die in erwünschter (konstruktrelevanter) oder unerwünschter (konstruktirrelevanter) Weise den Test und dessen Ergebnisse beeinflussen können. Weir (2005b: 46) gibt einen Überblick darüber, welche Aspekte im Prinzip in Bezug auf Testaufgaben zum Sprechen geregelt werden müssen:

---

<sup>6</sup> Diese Materialien sind auf der Website des Europarats zugänglich: [http://www.coe.int/t/dg4/linguistic/CADRE\\_FR.asp](http://www.coe.int/t/dg4/linguistic/CADRE_FR.asp)

<b>Merkmale der Aufgabe:</b> <ul style="list-style-type: none"> <li>- Zweck der Aufgabenstellung</li> <li>- Antwortformat, Erwartungen an Produkt</li> <li>- Zeit, Länge</li> <li>- Personenkonstellation(en)</li> <li>- Sprachliche Merkmale: <ul style="list-style-type: none"> <li>- lexikalisch</li> <li>- strukturell</li> <li>- funktional (z. B. Dialogtypen)</li> </ul> </li> <li>- Vorausgesetztes Weltwissen</li> </ul>	<b>Interlokutor(en):</b> <ul style="list-style-type: none"> <li>- Anzahl</li> <li>- Sprechgeschwindigkeit</li> <li>- Akzent(e)</li> <li>- Grad der Vertrautheit</li> <li>- Geschlecht</li> </ul>	<b>Durchführungsbestimmungen:</b> <ul style="list-style-type: none"> <li>- Räumlichkeiten</li> <li>- Konkrete Durchführung (Vorbereitung usw.)</li> </ul> <b>Beurteilung:</b> <ul style="list-style-type: none"> <li>- Beurteiler (Konsistenz, Strenge)</li> <li>- Beurteilungskriterien</li> </ul>
---	--	--

Abb. 3: Mögliche Quellen von konstruktrelevanter und konstruktirrelevanter Varianz in Tests zum Sprechen (nach Weir 2005b, 46)

Bei den Arbeiten in HarmoS FS standen die Untersuchungen zu den Kompetenzbereichen im Vordergrund. Das Sichern von Erkenntnissen und, soweit möglich, das Ableiten von Elementen für Test- und Testaufgabenspezifikationen aus den HarmoS-Untersuchungen gehörten aber mit zu den Zielen der Entwicklungsarbeiten, und zwar aus zwei Gründen: Zum einen sind Test- und Testaufgabenspezifikationen unauflösbar verbunden mit der Konstruktvalidierung und der Interpretation der Testergebnisse, denn die Testaufgaben bilden die entscheidende Nahtstelle zwischen Überprüfung und Beschreibung von Kompetenz; grundsätzlich ist die Rechtfertigung der Interpretationen, die aus Testergebnissen gezogen werden, eine permanente, zentrale Herausforderung auch für die Personen und Institutionen, die sich in HarmoS mit dem Sprachentesten befassen. Zum andern können Folgeuntersuchungen und das künftige Bildungsmonitoring von Einsichten in die Bedingungen und das Funktionieren von Tests und Testaufgaben profitieren. In diesem Bericht werden Aspekte von Test- und Testaufgabenspezifikationen bei der Darstellung der Untersuchungen zu den verschiedenen Kompetenzbereichen diskutiert, beim Lesen z. B. in Kap. 3.3.1 und schwerpunktmässig in 3.5.5.

## 2.2 Überprüfung des Konstrukts (Methodologie)

Eine der Hauptaufgaben des HarmoS-Fremdsprachenkonsortiums lag darin, für die Zielgruppen relevante Kompetenzbereiche zu bestimmen und dazu didaktisch und testtechnisch geeignete Überprüfungsverfahren zu entwickeln und zu erproben. Die konzeptuelle und praktische Ausgangsbasis für diese Arbeiten bildete primär der *Gemeinsame europäische Referenzrahmen für Sprachen* sowie Ergebnisse aus darauf aufbauenden Projekten, insbesondere *IEF/lingualevel*.

In Bezug auf einzelne zu vertiefende Aspekte des Kompetenzmodells bestand im Zusammenhang mit den Validierungsuntersuchungen ein spezifisches Forschungsinteresse:

- Identifizieren von schwierigkeitsbestimmenden Faktoren – oder Wechselwirkungen zwischen Faktoren – in Kommunikationsaufgaben (*Task*-Faktoren);
- Erkennen von Bezügen zwischen Prozess- und Handlungsdimensionen: Inwieweit bestimmen die kognitiven Prozesse, die bei der Bearbeitung von bestimmten Aufgaben ablaufen (z. B. Suchverhalten, Interpretieren, Folgern) den Handlungserfolg bei kommunikativen Aufgaben?;
- Identifizieren von Alterseffekten bei Sechst- und Neuntklässlern.

Durch die konzeptuellen und empirischen Arbeiten in HarmoS sollte die Ausgangsbasis (*Referenzrahmen, IEF/lingualevel* u. a.) nicht nur operationalisiert, sondern, soweit möglich, auch einer kritischen Prüfung unterzogen, ergänzt sowie gegebenenfalls modifiziert werden.

Allerdings konnten diese hochgesteckten Ziele, wie angesichts der Forschungslage und aufgrund des redimensionierten Projekts nicht anders zu erwarten war, nur teilweise erreicht werden.

Die Operationalisierung der Kompetenzbeschreibungen des *Referenzrahmens* in Form von Testaufgaben ist ein Gebiet, in dem grundsätzlich viel Entwicklungsarbeit zu leisten ist. Hinsichtlich der Operationalisierung von niveaubezogenen Kompetenzbeschreibungen für das Testen sind zur Zeit auf internationaler Ebene zahlreiche Arbeiten im Gange, v.a. im Zusammenhang mit dem oben erwähnten Handbuch für die Zuordnung von Sprachprüfungen zum *Referenzrahmen* (Council of

Europe 2003; Council of Europe, im Druck). Gerade die Stärken der Kompetenzbeschreibungen des *Referenzrahmens* (Verständlichkeit für Laien, Plausibilität, Kürze) werden manchmal umgekehrt als Schwäche angesehen: Testspezialisten beispielsweise vermissen in den Beschreibungen eine durchgehende Berücksichtigung eines festen Sets von Bestimmungsfaktoren (Weir 2005a); ihnen fehlen Inventare von Texten, Textmerkmalen, Wortschatz- und Grammatikelementen, von erschwerenden und erleichternden Rahmenbedingungen usw., auf welche sie bei der Aufgabenentwicklung zurückgreifen könnten. Allerdings sind solche Merkmale keine statischen Größen, und sie lassen sich kaum je *per se* einem bestimmten Niveau zuordnen. Auszugehen ist vielmehr von komplexen Wechselwirkungen zwischen text- und aufgabenbezogenen Merkmalen einerseits und Kompetenzen von Lernenden beim Bearbeiten kommunikativer Aufgaben andererseits (vgl. u. a. Bygate et al. 2001).

Wenn aber einerseits das Funktionieren von Tests verstanden werden soll und andererseits aus Schülerleistungen in konkreten Tests Schlüsse auf eine vorhandene und in einem bestimmten Rahmen auch verallgemeinerbare Kompetenz gezogen werden sollen<sup>7</sup>, dann ist trotz der offensichtlichen Komplexität der Schüler-Aufgaben-Konstellation ein prinzipiengeleitetes Herangehen nötig. Letztlich geht es, wie immer beim Testen, darum, dass die so genannte Konstruktvalidität gewährleistet und, im Prinzip, auch aufgezeigt werden kann.

Testentwickler im Bildungsbereich berufen sich hinsichtlich ihres Validitätsverständnisses häufig auf Messick<sup>8</sup>. Elementar wichtig für die Validität von Tests sind nach Messick die Verhinderung von Konstrukt-Unterrepräsentation und von Konstrukt-irrelevanter Varianz, sowie die testtechnisch gute Qualität des Tests und die Angemessenheit der Folgerungen aus den Testergebnissen (Messick 1996: 10; 12ff.). Konstrukt-Unterrepräsentation liegt dann vor, wenn ein Test das Konstrukt nicht "abdeckt", weil zum Beispiel die Items einseitig ausgewählt wurden oder nicht in genügender Anzahl im Test vorhanden waren. Solchermassen lückenhafte Daten erlauben kaum Aussagen über das Konstrukt insgesamt. Konstrukt-irrelevante Varianz ergibt sich zum Beispiel dann, wenn ein Testformat ein bestimmtes Geschick verlangt, das in der Population ungleich verteilt ist und das nichts mit dem zu messenden Konstrukt zu tun hat; komplizierte Computermanipulationen in einem Sprachtest würden vermutlich zu unerwünschter Varianz führen. Weiter kann konstrukt-irrelevante Varianz auch durch nicht hinreichend objektive Formen der Bewertung und Beurteilung entstehen, zum Beispiel bei mündlichen Prüfungen (mehr dazu in Lenz 2006: 210f.).

In HarmoS wurde auf vielfältige Weise versucht, den Forderungen von Messick nachzukommen. Auf Einzelheiten dazu wird in den Teilen zu den Entwicklungsarbeiten in den einzelnen Kompetenzbereichen näher eingegangen.

In vielen Prüfungen, die beanspruchen, sich auf den *Referenzrahmen* zu beziehen, werden heute Sprachkompetenzen (genauer: die sprachverwendungsbezogene *language proficiency*) primär mittels sprachfertigkeitsspezifischer (*skill-specific*) Kommunikationsaufgaben (*communicative tasks*) überprüft, so zum Beispiel in den *Cambridge*-, *DELFDALF*- oder *Goethe*-Diplomprüfungen. Die Unterteilung der Sprachverwendungskompetenz nach kommunikativen Sprachfertigkeiten in Hörverstehenshandlungen, Handlungen im Bereich der mündlichen Interaktion usw. hat sich in der Praxis weitgehend bewährt. Es besteht allerdings ein gewisses Risiko, dass durch diese Systematik umfassendere Handlungen, welche Sprachaktivitäten aus verschiedenen Fertigkeitsbereichen umfassen, tendenziell nicht als Einheit wahrgenommen und nicht als solche getestet werden. In der Praxis werden z. B. Fertigkeiten in diesem Sinne kombiniert, wenn aus mündlichen und schriftlichen Quellen Informationen zusammengesucht und dann in mündlicher oder schriftlicher Form weitervermittelt werden, was in bestimmten Kontexten eine durchaus relevante Aufgabenstellung

---

<sup>7</sup> d. h. Schlüsse in Bezug auf das „Konstrukt“: die latenten Könnens- oder Wissensbereiche, über die etwas ausgesagt werden soll

<sup>8</sup> Einflussreich spezifisch im Bereich des *language testing* sind Bachman und Palmer mit ihrem Qualitätsbegriff der *test usefulness* (Bachman & Palmer 1996, 17ff.). *Test usefulness* wird dabei als Konglomerat gesehen, das verschiedene Qualitätsaspekte von Tests zusammennimmt und gegeneinander abwägt, die in anderen Konzeptionen zum Teil als Antagonismen interpretiert wurden: Reliabilität, Konstruktvalidität, Authentizität, Interaktivität (werden die gewünschten Kompetenzen angesprochen und aktiviert?), *impact* und Praktikabilität. Die beiden letzten sind testexterne Qualitätskriterien, welche die Auswirkungen auf das Sprachenlehren und -lernen (*impact*) bzw. die Grenzen der praktischen Machbarkeit (in diesem Ansatz das hierarchiehöchste Kriterium) betreffen.

sein kann. Auch das HarMoS-Fremdsprachenkonsortium ordnet sich mit seinen Tests in die Tradition eines **aufgabenbasierten kommunikativen Sprachentestens** ein<sup>9</sup>. Konkret überprüft wurden zwar nur Kompetenzen in den Fertigungsbereichen Leseverstehen, Schreiben und, in weit geringem Umfang, auch Sprechen – Niveaubeschreibungen werden aber, auf der Grundlage von früheren Untersuchungen, auch zum Hörverstehen und versuchsweise zur Sprachmittlung vorgelegt. Dem Fertigkeitenmodell entsprechend werden die Kompetenzen der Schülerinnen und Schüler, und auch die Basisstandards, nicht *global* für Französisch, Englisch oder Deutsch beschrieben, sondern z. B. in Bezug auf die Leseverstehens- oder Schreibkompetenz in den verschiedenen Sprachen.

Um die Konstruktdeckung innerhalb jedes der Fertigungsbereiche besser gewährleisten zu können, wurden für die Testentwicklung neben den Fertigungsbereichen zusätzlich funktional-linguistische Ordnungskategorien eingeführt, denen die Testaufgaben bzw. -aufgabenteile zugeordnet wurden. Es handelt sich dabei um fünf *Diskurstypen* (auch *rhetorische Funktionen* genannt): Informieren/Beschreiben, Auffordern/Veranlassen, Erzählen/Berichten, Meinungsäußerung/Argumentieren, Unterhalten von Beziehung. Durch diese Ordnungskategorien sollte beispielsweise gewährleistet werden, dass es beim Schreiben nicht nur Aufgaben(teile) gab, in denen beschrieben oder erzählt werden musste, sondern auch solche, in denen es darum ging, die Meinung zu äussern und zu begründen. Spezifisch beim Leseverstehen wurden zusätzlich andere, möglicherweise konstruktrelevante Grundunterscheidungen eingeführt, die an die Unterscheidung nach Leseprozessen bei PISA anknüpfen.

Die handlungsorientierten Testaufgaben, die in HarMoS eingesetzt wurden, haben grundsätzlich einen „holistischen“ Charakter, d. h. ihre Bearbeitung involviert eine grosse Zahl von Kompetenzen in nur schwer zu durchschauender Komplexität. Dies wird allerdings in Kauf genommen, weil davon ausgegangen wird, dass ein reduktionistisches, auf Einzelkompetenzen fokussiertes Testen (*discrete-point testing*) zwar vordergründig klarere Ergebnisse liefern würde, dass aber die Generalisierungsleistung von „diskreten“ Ergebnissen hin zu Aussagen über das Verhalten beispielsweise auch bei ausserschulischen Kommunikationsanforderungen damit noch nicht erbracht wäre. Testaufgaben, die bereits wesentliche Parallelen zu Kommunikationsaufgaben aufweisen, wie man sie ausserhalb der Testsituation antrifft, ermöglichen eher plausible Extrapolationen.

Bei der Aufgabenkonstruktion wurden neben den eben beschriebenen funktionalen Kriterien auch solche Gesichtspunkte einbezogen, die sich auf die weiter oben beschriebenen allgemeinen Kompetenzen und kommunikativen Sprachkompetenzen beziehen. Die Anforderungen an die allgemeinen Kompetenzen der Schülerinnen und Schüler (z. B. das vorausgesetzte Weltwissen oder die kognitive Komplexität eines Themas) wurden, soweit wie möglich, ebenso kontrolliert wie diejenigen aus dem Bereich der kommunikativen Sprachkompetenzen (z. B. Differenziertheit des Wortschatzes und Komplexität der Strukturen in Lesetexten). Zudem spielten die sprachlichen Ressourcen in den Rastern zur Beurteilung der Qualität der mündlichen und schriftlichen Leistungen als Beurteilungsgesichtspunkte eine wichtige Rolle.

Die statistische Analyse der verschiedenartigen und sehr umfangreichen Daten und die darauf aufbauende Interpretation der Ergebnisse der HarMoS-Tests erwiesen sich als sehr aufwändig und konnten nicht immer in der wünschbaren Breite und Tiefe erfolgen. Weitere Analysen und Interpretationen wären insbesondere auch in Bezug auf die für das Kompetenzverständnis wichtige Frage nötig, inwieweit sich die verschiedenen Kompetenzbereiche (die in HarMoS unterschiedenen kommunikativen Sprachfertigkeiten, v. a. Leseverstehen und Schreiben) sowie andere Kategorisierungen (wie die nach Diskurstypen oder Leseprozessen) statistisch als empirische Dimensionen unterscheiden lassen. Kann also zum Beispiel, wer überdurchschnittlich gut lesen kann, in der Regel auch überdurchschnittlich gut schreiben – oder verteilen sich diese beiden Fähigkeiten bei unseren Zielgruppen jeweils anders, handelt es sich also um zwei Kompetenzdimensionen?

---

<sup>9</sup> Die Entscheidung für Aufgabenorientierung hat auch didaktische Gründe: Diese ist in der Fremdsprachendidaktik generell ein vielversprechender, zukunftsreicher Ansatz, der unterdessen relativ gut erforscht ist und in den letzten Jahren vermehrt auch im deutschsprachigen Raum in programmatischen Beiträgen und Forschungsarbeiten Beachtung gefunden hat (vgl. Bygate, Skehan & Swain, 2001; Skehan & Foster, 2001; Bachman, 2002; Ellis, 2003 u. 2005; Nunan, 2004; Robinson, 2005; Bausch et al., 2006; Müller-Hartmann & Schocker-v. Dittfurth 2005; Portmann-Tselikas, 2006; Long 2007).

Entwickelt sich Wortschatzkompetenz parallel zur grammatischen Kompetenz, oder beherrschen die einen das eine besser und die anderen das andere? Können die einen Schülerinnen und Schüler leichter argumentierende Texte lesen, die anderen leichter instruierende? Auch zu den oben als Forschungsinteressen erwähnten Punkten gibt es erst ansatzweise einige Hinweise. Die komplexen Wechselwirkungen zwischen Personen- und Aufgabenfaktoren sind nicht leicht zu durchdringen, eine erklärende Interpretation der empirischen Ergebnisse nicht einfach zu bewerkstelligen. Die Untersuchungsanlage im vorliegenden Projekt setzt bestimmten Erkenntnismöglichkeiten auch ganz klar Grenzen. In diesem Zusammenhang ist es besonders bedauerlich, dass – anders als im Jahre 2006 geplant – keine qualitativen Untersuchungen zum Funktionieren der Aufgaben durchgeführt werden konnten. Die Schülerinnen und Schüler bleiben dadurch weitgehend eine *black box*, nur einige wenige Sozio-Daten liegen als potenzielle Personenfaktoren vor<sup>10</sup>, die im Zusammenspiel mit Aufgabenfaktoren und Merkmalen der Leistungen interpretiert werden könnten.

### 3 Die empirischen Untersuchungen in HarmoS

#### 3.1 Überblick

Das Konsortium Fremdsprachen beteiligte sich einerseits an der repräsentativen Hauptuntersuchung, die von der Projektleitung HarmoS organisiert und im Frühjahr 2007 durchgeführt wurde. Andererseits führte es selbstständig im Mai 2007 Feldtests zum Sprechen durch. An den Untersuchungen teilgenommen haben in beiden Fällen nur die Deutschschweiz und die Westschweiz, nicht die italienischsprachige Schweiz. Im Rahmen der Hauptuntersuchung wurden nur rein schriftliche Tests zum Lesen und Schreiben durchgeführt; versuchsweise wurde pro Sprache auch ein C-Test – eine Form von Lückentest – von 20 Minuten Dauer in die Untersuchung eingebracht. Für Hörverstehenstests stand nach den ergebnislosen Eventualplanungen des Jahres 2006 zu wenig Entwicklungszeit (und letztlich auch zu wenig Testzeit) zur Verfügung.

Die zweite Landessprache (Französisch bzw. Deutsch) wurde sowohl in den sechsten als auch in den neunten Klassen getestet, Englisch nur in den neunten, da die Sechstklässler/innen des Jahres 2007 in der Primarschule noch keinen Englischunterricht gehabt hatten.

In allen Kompetenzbereichen (inkl. C-Test) wurde ein Teil der Aufgaben nur in sechsten oder neunten Klassen eingesetzt, ein weiterer Teil auf beiden Stufen. Diese Aufgabenüberlappung war eine wichtige Voraussetzung für die Entwicklung einer durchgehenden (empirischen) Skala, die alle eingesetzten Aufgaben und alle teilnehmenden Schülerinnen und Schüler pro Zielsprache gemeinsam abbilden kann. Entwicklung, Einsatz und erfolgreiche Skalierung von stufenübergreifenden Aufgaben wurden dadurch erleichtert, dass nicht direkt curriculum- und lernstoffbezogen getestet wurde (*achievement testing*), sondern qualifikationsorientiert (*proficiency testing*), d. h. daraufhin, was Schüler in Bezug auf relevante handlungsbezogene Ziele zu leisten imstande sind.

Viele der Aufgaben wurden mit den minimal notwendigen Anpassungen in allen drei Zielsprachen eingesetzt. Die Vergleichbarkeit der Ergebnisse über die Zielsprachen hinweg ist allerdings recht schwierig zu bewerkstelligen bzw. zu belegen. Neben statistischen Methoden (v.a. beim Lesen) sind bisher auch Formen des Experten-Benchmarkings zum Zuge gekommen (z. B. bei den produktiven Leistungen). Auch zur Absicherung der darauf basierenden Resultate wäre weitere Forschung wünschbar.

Neben der HarmoS-internen Skalenbildung ist das (Rück-)Beziehen der Ergebnisse auf die mit Deskriptoren illustrierte europäische Referenzskala ein zentrales Anliegen. In den diesbezüglichen Arbeiten hat sich gezeigt, dass dieses Beziehen auf die Referenzskala je nach Kompetenzbereich und Sprache eher ein zusätzlicher Schritt, oder aber selbst ein wichtiges Mittel sein kann, um die Testergebnisse zu den verschiedenen Sprachen überhaupt miteinander in Beziehung zu setzen.

Das parallele Testen von drei Fremdsprachen bei unterschiedlichen Gruppen in unterschiedlichen Landesteilen macht das Projekt zwar sehr aufwändig, bietet aber wissenschaftlich und von der

---

<sup>10</sup> Einen Überblick über ein umfassenderes Validierungsdesign gibt Lenz 2006: 214-219.

Testvalidierung her gesehen ein willkommenes Korrektiv gegen voreilige Folgerungen aus Ergebnissen, die sich zum Beispiel nur bei den Tests in einer Sprache zeigen.

### 3.2 Umfang der Untersuchungen

#### Repräsentative Hauptuntersuchung April-Mai 2007

Die Fremdsprachenkompetenzen wurden in der Deutschschweiz und der Westschweiz in gleichem Umfang getestet. Die einzelnen Schülerinnen und Schüler bearbeiteten am Testtag jeweils vier so genannte Hefte (2 Hefte zu 30 min., 2 Hefte zu 20 min.) mit Aufgaben zur lokalen Schulsprache, zur zweiten Landessprache und/oder zu Englisch. Die Rotation der Hefte erfolgte auf der Ebene des einzelnen Schülers bzw. der einzelnen Schülerin (Ausnahme: der Hörverstehenstest zur lokalen Schulsprache, den jeweils eine ganze Klasse zu Beginn des Testhalbtags bearbeitete). Für die Aufgabenverteilung wurden einige wenige Restriktionen definiert. Beispielsweise mussten über die drei maximal möglichen Sprachen hinweg nie in mehr als zwei Heften Aufgaben zur selben kommunikativen Sprachfertigkeit gelöst werden (Lesen/C-Test, Schreiben). Oder eine Fremdsprache wurde immer während mindestens 30 min. bearbeitet, nie nur in einem einzelnen 20-Minuten-Heft.

Die Gesamttests wurden in einer Matrixstruktur auf die verschiedenen Schülerinnen und Schüler verteilt, sodass die einzelnen Schülerinnen und Schüler jeweils nur während 100 min. an den Testheften arbeiten mussten. Insgesamt konnten aber – dank geeigneter Verteilung (Rotation) der Testhefte sowie Software, die mit sehr lückenhaften Daten umgehen kann (Rasch-Software *ConQuest*) – Informationen gewonnen werden, die bei lückenlosen Designs eine sehr viel längere individuelle Testzeit bedingen würden. Die folgenden Tabellen geben einen Überblick über die Gesamttestzeiten (der virtuellen Gesamttests) sowie die Zahl der Hefte und der eingesetzten Aufgaben<sup>11</sup>:

#### Gesamt-Testzeiten 6. Klasse (zweite Landessprachen Französisch oder Deutsch)

	Französisch			Deutsch		
	Gesamtdauer	Anz. Aufg.	Anz. Hefte	Gesamtdauer	Anz. Aufg.	Anz. Hefte
<b>Leseverstehen</b>	140 min.	13 Aufg.	5 1/3 Hefte	140 min.	13 Aufg.	5 1/3 Hefte
<b>Schreiben</b>	120 min.	6 Aufg.	6 Hefte	120 min.	6 Aufg.	6 Hefte
<b>C-Test</b>	20 min.	3 Texte / 60 Lücken	2/3 Heft	20 min.	3 Texte / 60 Lücken	2/3 Heft
<b>Total</b>	280 min.			280 min.		

#### Gesamt-Testzeiten 9. Klassen (zweite Landessprachen Französisch oder Deutsch)

	Französisch			Deutsch		
	Gesamtdauer	Anz. Aufg.	Anz. Hefte	Gesamtdauer	Anz. Aufg.	Anz. Hefte
<b>Leseverstehen</b>	130 min.	10 Aufg.	5 Hefte	130 min.	10 Aufg.	5 Hefte
<b>Schreiben</b>	130 min.	6 Aufg.	6 Hefte	130 min.	6 Aufg.	6 Hefte
<b>C-Test</b>	20 min.	3 Texte / 60 Lücken	1 Heft	20 min.	3 Texte / 60 Lücken	1 Heft
<b>Total</b>	280 min.			280 min.		

#### Gesamt-Testzeiten 9. Klassen (Englisch)

<sup>11</sup> Beim Leseverstehen wird hier unter „Aufgabe“ die Kombination von Text(en) und Aufgabenstellungen (im Mittel etwa 5 Testitems) verstanden, die unter einer Situierung verbunden sind. Für diese Leseaufgaben wird eine Bearbeitungsdauer von 10 oder 15 Minuten veranschlagt. Beim Schreiben ist eine Aufgabe die Anweisung zum Schreiben eines einzelnen Texts. Die Bearbeitungsdauer beträgt 20 min., bei einer einzelnen Aufgabe in der neunten Klasse 30 min.

	Englisch Deutschschweiz			Englisch Westschweiz		
	Gesamtdauer	Anz. Aufg.	Anz. Hefte	Gesamtdauer	Anz. Aufg.	Anz. Hefte
<b>Leseverstehen</b>	110 min.	9 Aufg.	4 1/3 Hefte	110 min.	9 Aufg.	4 1/3 Hefte
<b>Schreiben</b>	150 min.	7 Aufg.	7 Hefte	150 min.	7 Aufg.	7 Hefte
<b>C-Test</b>	20 min.	3 Texte / 60 Lücken	2/3 Heft	20 min.	3 Texte / 60 Lücken	2/3 Heft
<b>Total</b>	280 min.			280 min.		

Die folgenden Tabellen vermitteln einen Überblick über die Anzahl der beteiligten Schülerinnen und Schüler sowie über den Rücklauf der Testhefte.

### Schülerbeteiligung

	D-CH+W-CH gesamt	D-CH		W-CH	
		6. Klasse	9. Klasse	6. Klasse	9. Klasse
<b>Anzahl verschiedener Schüler/-innen, die in den Fremdsprachen Hefte retourniert haben</b>	5412	980	1512	1160	1760

### Rücklauf an Testheften

	6. Klasse		9. Klasse	
	Anz. eingegangener Testhefte	Eingegangene Testhefte (%)	Anz. eingegangener Testhefte	Eingegangene Testhefte (%)
<b>Leseverstehen/C-Test 2. Landessprache (F bzw. D)</b>	1830	90.4%	1311	86.9%
<b>Leseverstehen/C-Test Englisch</b>			1221	86.5 %
<b>Schreiben 2. Landessprache</b>	1122	75% <sup>12</sup>	1101	81.8 %
<b>Schreiben Englisch</b>			1309	84.8%

Bei den Fremdsprachen liegt der Rücklauf damit etwas unter dem, was üblicherweise in grossen Untersuchungen zu erwarten ist. Der Rücklauf bei den gleichzeitig getesteten regionalen Schul-sprachen (D bzw. F) liegt bei über 90%. Die Differenz ist erstaunlich. Es liegen uns keine Informationen vor, die den geringeren Rücklauf bei den Fremdsprachen erklären könnten (aber s. Fussnote zum Schreiben).

### Feldtests zum Sprechen Mai 2007

Die Feldtests zum Sprechen wurden durch das Fremdsprachenkonsortium in Eigenregie organisiert. Vorrangiges Ziel dieser Untersuchungen war es, ein breites Spektrum von Schülerleistungen zu erfassen und auf Video zu dokumentieren, sodass eine gute Basis für die Diskussion und Illustration der Basisstandards entstehen würde<sup>13</sup>.

Die Auswahl der Schülerinnen und Schüler ist nicht repräsentativ, jedoch wurde darauf geachtet, dass eine gleichmässige Vertretung der Sprachregionen sowie eine möglichst breite Streuung über die Kantone und Klassenniveaus (Sek. A-C) zustande kam; allerdings erwies es sich als schwierig, in den Klassen des tiefsten Niveaus (C) Freiwillige zu finden. Eine ausgeglichene Beteiligung von

<sup>12</sup> Der Zahl der Arbeiten zum Schreiben in der 6. Klasse liegt nur deshalb tiefer, weil die Ergebnisse, die zu einer der sechs eingesetzten Aufgaben entstanden, aus der Analyse ausgeschlossen wurden. Diese Aufgabe konnte wider Erwarten von den Schülerinnen und Schülern nicht angemessen interpretiert werden.

<sup>13</sup> Es wurden nur solche Schülerinnen und Schüler gefilmt, deren Eltern ihre Zustimmung für die Verwendung der Aufnahmen im Rahmen der Standards-Diskussion gaben.

Knaben und Mädchen wurde erreicht. In der Regel wurden an einer Schule je drei Schülerinnen oder Schüler aus zwei verschiedenen Klassen gefilmt. Diese drei Schüler/innen wurden von der Lehrerin oder dem Lehrer immer so ausgewählt, dass sie jeweils das untere, mittlere und obere Drittel des Leistungsspektrums der Klasse (in Bezug auf das Sprechen) vertraten. Insgesamt nahmen 174 Schülerinnen und Schüler an der Untersuchung teil (F: 60; D: 54; E: 60).

Die folgende Tabelle vermittelt einen groben Überblick über die beteiligten Klassen:

	Französisch		Deutsch		Englisch	
	6. Klasse	9. Klasse	6. Klasse	9. Klasse	Dt.-Schweiz	Westschweiz
<b>Schüler pro Stufe</b>	30	30	30	24	30	30
<b>Anzahl Schu- len/Klassen</b>	5/10	5/10	5/8	4/8	5/10	5/10
<b>Klassentypen</b>	Primar	Sek I, A: 2 Sek I, B: 5 Sek I, C: 3	Primar	Sek I, A: 2 Sek I, B: 5 Sek I, C: 1	Sek I, A: 4 Sek I, B: 4 Sek I, C: 2	Sek I, A: 6 Sek I, B: 3 Sek I, C: 1

Aus den Tests zum Sprechen resultierten Aufnahmen mit einer Bruttodauer (inkl. Kurzvorbereitung vor den Aufgaben) von über 50 Stunden. Die erfassten Schülerleistungen illustrieren ein Niveauspektrum von A1.1 bis B1.2, einzelne Schüler/innen liegen sogar etwas darüber, obwohl Zweisprachige im engeren Verständnis des Wortes nicht einbezogen wurden.

In der ursprünglichen Projektplanung waren neben der Teilnahme an der repräsentativen Hauptuntersuchung und den Feldtests zum Sprechen ergänzende Untersuchungen in Spezial- oder Pilotklassen (z. B. mit CLIL-Unterricht oder früherem Beginn) vorgesehen, damit mehr Anhaltspunkte für die (in Zukunft erwartbaren) Kompetenzen von Schülerinnen und Schülern hätten gewonnen werden können. In Absprache mit der HarmoS-Projektleitung wurde dieser Projektteil aber nicht weiterverfolgt, und es wurden stattdessen ergänzende Untersuchungen zu einem Zeitpunkt nach der Umsetzung des Szenarios 2010/12 ins Auge gefasst.

### 3.3 Testinstrumente

#### 3.3.1 Grundsätzliche Überlegungen zur Überprüfung von Kompetenzen

Die Tests in HarmoS haben die Funktion, ein Kompetenzmodell zu operationalisieren, sodass Leistungen erfasst werden können, die sich auch wieder auf dieses Kompetenzmodell zurückführen lassen. Dass diese Rückführung legitim ist, muss im Rahmen einer Testvalidierung aufgezeigt werden, indem Aufgaben, Leistungen, Auswertung der Leistungen und Aussagen im Rahmen des Modells aufeinander bezogen werden.

Ein viel beachtetes Untersuchungs- und Validierungsmodell, das auch im Bereich des Sprachentestens immer wieder ins Spiel gebracht wird (Mislevy et al. 2002; Bachman 2002), ist das *Evidence-Centred Design* (ECD) (Mislevy, Steinberg, Almond & Russell 2002; Mislevy, Almond & Lukas 2004)<sup>14</sup>. Es war auch bei der Testentwicklung in HarmoS FS eine der konzeptuellen Grundlagen.

<sup>14</sup> Ebenfalls interessant in diesem Zusammenhang ist das Modell der *Four building blocks* von Mark Wilson (2004), das dem ECD sehr ähnlich ist.

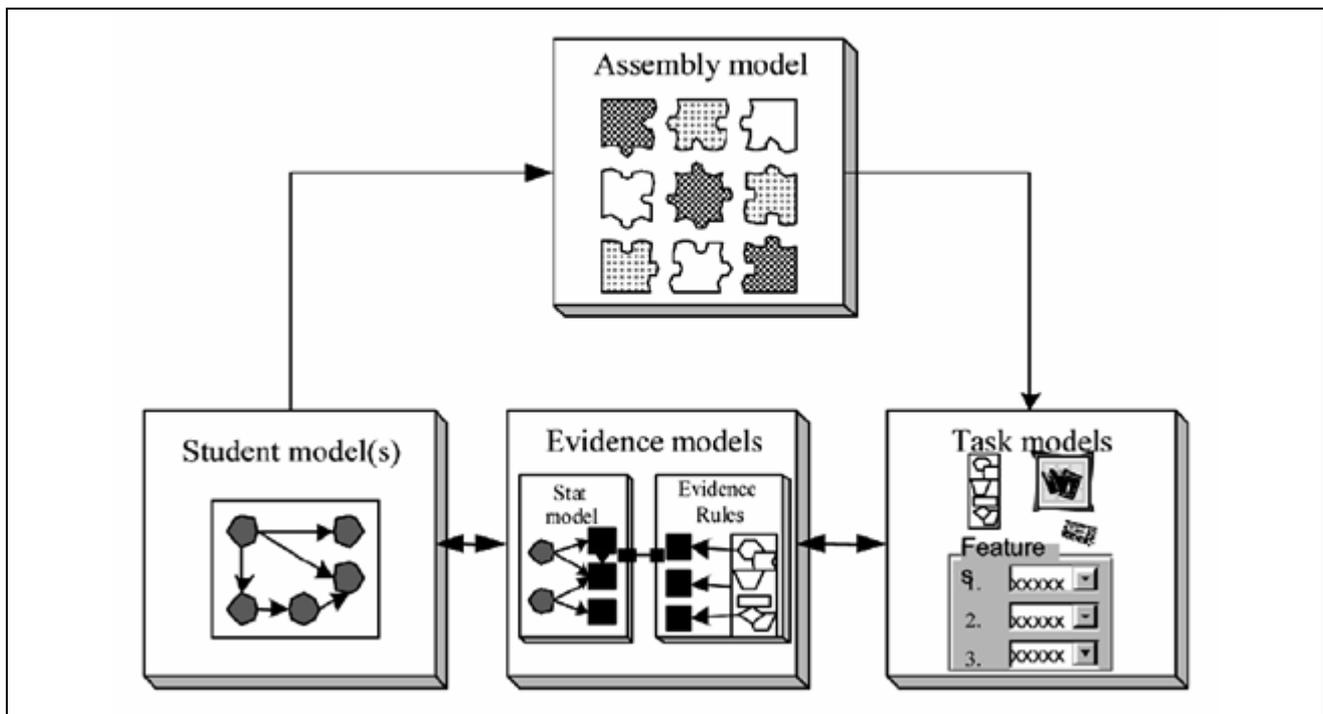


Abb. 4: Schema des *Evidence-Centred Design* (Mislevy et al. 2002: 481)

Das ECD-Modell kann folgendermassen interpretiert und genutzt werden: Im *student model* werden die relevanten Variablen (Wissen und/oder Können) als Kompetenzbereiche oder -aspekte spezifiziert<sup>15</sup>. Die Auswahl dieser Aspekte ist oft von externen Überlegungen bestimmt (Was will/muss man über die Population wissen?) oder, wie in HarmoS in Form des *Referenzrahmens*, durch ein bestimmtes Kompetenzverständnis. Bei den Variablen des *student model* handelt es sich um latente Variablen, die zu beobachtbaren Variablen in Beziehung gesetzt werden sollen:

Operationally, SM [student model] variables summarize patterns of values of observable variables, along lines we build into the evidence model structures [...], as evoked by features we build into tasks. (Mislevy et al. 2002: 481)

Es ist zu beachten, dass im *student model* bestimmte, als zentral (d.h. konstruktrelevant) angenommene Kompetenzen fokussiert werden. Es handelt sich also nicht um das Ensemble der Kompetenzen, über die jemand verfügt und zum Beispiel bei der Bearbeitung von (kommunikativen) Sprachverwendungsaufgaben innerhalb oder ausserhalb einer Testsituation einsetzt. Bei der Aufgabenkonstruktion und bei der Interpretation von Leistungen müssen aber Anforderungen im Bereich weiterer (auch konstruktirrelevanter) Kompetenzen mitberücksichtigt werden.<sup>16</sup>

Bei der Umsetzung des *student model* in Aufgaben wird versucht, in genügender Zahl und in der nötigen Vielfalt Gelegenheiten zu schaffen, bei denen die Kompetenzen, die im Konstrukt fokussiert werden, eingesetzt werden müssen. Gleichzeitig geht es darum, den Einsatz an konstruktirrelevanten Kompetenzen zu kontrollieren und einen verzerrenden Einfluss zu verhindern, besonders wenn es um Anforderungen geht, die spezifisch aus der Testsituation heraus entstehen können (z. B. durch komplexe Anweisungen, mangelnde Qualität eines Audiogeräts, nur einigen Schüler/innen bekannte thematische Inhalte). Eine grundsätzliche Schwierigkeit bei der Interpretation von Untersuchungs- (oder Test-)Ergebnissen besteht aber in jedem Fall darin, den relativen Anteil bestimmen zu können, den unterschiedliche Kompetenz-Ressourcen an den Ergebnissen haben, und so zu verwertbaren Aussagen über das Kompetenz-Konstrukt und die diesbezüglichen Kompetenzen der Getesteten zu kommen.

<sup>15</sup> Damit ist die *Form* der Beschreibung noch nicht festgelegt. Möglich sind Aussagen über abstrakte Fähigkeiten (z. B. „hohe Wortschatzkompetenz“) oder verallgemeinernde Aussagen über Handlungsfähigkeit („Kann sich auch bei ... meist angemessen ausdrücken“ – die Deskriptoren des Referenzrahmens gehören zu diesem Typ).

<sup>16</sup> Aus diesen beiden Sichtweisen auf Kompetenz resultierte das z. B. für die Interpretation von Leseverstehens-Items zentrale Problem der teilweisen Divergenz von geschätzter Schwierigkeit (orientiert sich schwergewichtig an sprachlich-kommunikativen Aspekten der Schwierigkeit bzw. Leistung) und empirischer Schwierigkeit (ist auf effektive Anforderungen/Leistungen im Test und damit auch auf konstruktferne und -irrelevante Merkmale bezogen).

Der zweite Ausgangspunkt für die Testentwicklung ist neben dem *student model* die *Task-Seite* mit *assembly model* und *task model*. Innerhalb eines „zweiseitigen“ Ansatzes, wie ihn das Fremdsprachenkonsortium in Anlehnung an Bachman & Palmer (1996: 12) verfolgt, geht es nicht nur darum, die zu überprüfenden Aspekte von Kompetenz (*student model*) in den Aufgaben insgesamt in genügendem Umfang zu implementieren, sondern auch darum, dies in Form von Aufgaben zu tun, die relevante Aspekte der Sprachverwendung repräsentieren, die überprüft werden soll (vgl. dazu Lenz 2006: 214ff.). In Testaufgaben ist eine „naturalistische“ Umsetzung von realweltlichen Aufgaben normalerweise nicht möglich. Tests tragen bestimmte Merkmale, die testspezifisch sind und die die Ergebnisse unweigerlich beeinflussen. Bei Tests zum Hör- oder Leseverstehen kommt beispielsweise die indirekte Überprüfung des Verständnisses mittels bestimmter Antwortformate hinzu (z. B. Multiple-Choice, Kurzantworten, evtl. Zusammenfassungen, je nachdem welche *task models* konkret gewählt werden). Bei Sprech- oder Schreibaufgaben ist es oft eine besondere Herausforderung für das Aufgabendesign, Aufgabenstellungen so zu formulieren, dass im Lernenden eine angemessene Handlungsabsicht entsteht – ein Problem, das bei realweltlichen Aufgaben oft gar nicht existiert, die Leistung im Test aber entscheidend beeinflussen kann. „Kompetenz“ kann grundsätzlich weder in Tests noch in realweltlichen Situationen direkt beobachtet werden, alle Aussagen darüber sind Inferenzleistungen. Damit Inferenzen aber möglich sind, muss überhaupt erst Wesentliches beobachtet werden können. Das ist nur möglich, wenn die Aufgaben zentrale Aspekte der angezielten Kompetenz(en) mobilisieren und gleichzeitig keine Störeffekte erzeugen, die so stark sind, dass sie den Blick auf die wesentliche (aber verborgene) Kompetenz verstellen oder verzerren (wie dies beispielsweise bei einem Hörverstehenstest der Fall sein kann, der das Verständnis durch eine anspruchsvolle Schreibleistung überprüft).

Die *evidence models* haben die Funktion, die relevanten Merkmale der Kombination von Aufgaben und den dazugehörigen Lernerleistungen zu erfassen und in Aussagen über die im Konstrukt bzw. *student model* fokussierten Kompetenzen zu überführen. Umgekehrt geben die *evidence models* auch wesentliche Anhaltspunkte dafür, wie Leistungen aussehen, die bestimmten Ausprägungen des Konstrukts in bestimmten Kontexten entsprechen – dies in Form von erwartbaren Testergebnissen und Merkmalen von Schülerleistungen. Diese zweiseitige Interpretationsleistung, die durch die *evidence models* geschieht, ist ein wichtiger Angelpunkt im ECD und eine permanente Herausforderung, besonders für ein Sprachentesten, das auf ganzheitliche Kommunikationsaufgaben auf der einen Seite sowie Aussagen über etwas so Komplexes wie Sprachverwendungskompetenz auf der anderen Seite nicht verzichten will<sup>17</sup>.

Das Verhältnis zwischen Kompetenzen, Aufgabeneigenschaften und beobachtbaren Variablen bzw. den verschiedenen Modulen des ECD-Modells erklären Mislevy et al. wie folgt:

Distinct SM variables are used to maintain belief about distinct knowledge/skills, and a claim is associated with particular patterns across them as they are called upon in tasks that stress competences in different ways. Thus, students' proficiency in a domain can be described in terms of which skills they possess (via SM variables), tasks can be described in terms of which skills they require (via TM [task model] variables) and the outcomes expected from any given match-up can be described in terms of values of observable variables. (Mislevy et al. 2002: 484)

Wie es auch das ECD-Modell erkennen lässt, sind die Testentwicklung und die Entwicklung und Validierung eines Kompetenzmodells interdependente Elemente eines rekursiven Prozesses; die Validierung von Modell und Erkenntnisinstrument ist im Prinzip nie abgeschlossen – trotzdem ist sie unabdingbar, vor allem bei Neuentwicklungen. Dabei gilt – im Sinne von Messick –, dass Validierungsmassnahmen ergriffen werden sollen, die in einem vertretbaren Verhältnis stehen zu den (realen) Konsequenzen, welche sich für die Betroffenen aus den Tests ergeben.

---

<sup>17</sup> In einem State-of-the-art-Artikel ziehen Alderson & Banerjee nach rund 20 Jahren der Entwicklung von „kommunikativen Tests“ eine eher skeptische Bilanz:

Strategies, and presumably traits [d. h. die individuellen kognitiven Dispositionen], can vary across persons and tasks, even when the same scores are achieved. The same test score may represent different abilities, or different combinations of abilities, or different interactions between traits and contexts, and it is currently impossible to say exactly what a score might mean. This we might term The Black Hole of language testing. (Alderson & Banerjee 2002, 100f.)

### 3.3.2 Zur Testentwicklung in HarmoS

Während das ECD-Modell bei der Testentwicklung die methodische Grundlage bildete, war das Kompetenzmodell des *Referenzrahmens* in inhaltlicher Hinsicht grundsätzlich der Ausgangspunkt für die Entwicklung der HarmoS-Tests zum Leseverstehen und zum Schreiben (repräsentative Untersuchung) sowie zum interaktiven und produktiven Sprechen (Feldtests). Das Kompetenzmodell des *Referenzrahmens* ist in verschiedenen Dokumenten konkretisiert, im *Gemeinsamen europäischen Referenzrahmen* selbst und in Dokumenten, die im Anschluss an den *Referenzrahmen* entstanden sind, darunter v.a.:

- das *Europäische Sprachenportfolio* und Arbeiten zum Sprachenportfolio;
- Ergebnisse des Deutschschweizer Projekts *IEF* (Verlagsprodukt *lingualevel*);
- auf dem *Referenzrahmen* aufbauende und ihn ergänzende Dokumente aus Projekten des Europarats (Handbuch für die Zuordnung von Sprachprüfungen; der so genannte *Dutch Grid* zur Beschreibung relevanter Aufgabenmerkmale; Dokumente der *Association of Language Testers in Europe*).

Ergänzend wurde theoretische und empirische Forschung zum Sprachenlernen und Sprachentesten, unter anderem auch zu PISA (*reading literacy*) beigezogen.

- Neben der Entscheidung, getrennt nach den kommunikativen Sprachfertigkeiten (Leseverstehen usw.) zu testen, diese also als getrennte Kompetenzbereiche zu behandeln, wurde frühzeitig auch entschieden, besonders auf eine genügende Konstruktdeckung im funktionalen Bereich zu achten. Zu diesem Zweck wurde in Anlehnung an verschiedene, jeweils leicht voneinander abweichende Klassifizierungen, eine funktionale Einteilung der Aufgaben nach den jeweils dominierenden Diskurstypen (Informieren/Beschreiben usw.) vorgegeben (vgl. dazu Kap. 2.2 oben).

Als dritter Kategorienbereich wurden die verschiedenen Kompetenz-Ressourcen des *Referenzrahmens*, d. h. die „allgemeinen Kompetenzen“ und die „kommunikativen Sprachkompetenzen“ (vgl. Kap. 2.1, oben) als konstitutive Konstruktionsgesichtspunkte festgelegt.

Je nach Kompetenzbereich kamen für die Aufgabenentwicklung evtl. weitere Systematisierungsgesichtspunkte dazu, welche eine zuverlässigere Konstruktdeckung sicherstellen sollten (vgl. dazu insbesondere die Zuordnung der Aufgaben zu Leseprozessen bei den Tests zum Leseverstehen).

Gesichtspunkte, die unterschieden werden, damit die Konstruktdeckung durch die Aufgaben besser gewährleistet ist, haben grundsätzlich das Potenzial, sich in der Analyse der empirischen Ergebnisse als empirisch unterscheidbare Kategorien herauszustellen, sonst müssten sie nicht separat ausgewiesen werden. Es muss also denkbar sein, dass in Abhängigkeit von diesen Kategorien jeweils andere Schülerinnen und Schüler besser bzw. weniger gut sind. Ob der empirische Nachweis von Dimensionalität entlang diesen Kategorien im konkreten Fall gelingt oder nicht, hängt allerdings auch wesentlich vom Testumfang und von der Datenqualität ab.

Als Grundlage für die konkrete Entwicklung der Testaufgaben und gegebenenfalls der Beurteilungsraster in den einzelnen Kompetenzbereichen (kommunikative Sprachfertigkeiten) wurden so genannte **construct maps** entworfen, in denen die genannten Kategorien niveaubezogen konkretisiert wurden. Die wichtigsten Quellen für diese Merkmale bildeten die skalierten Niveaubeschreibungen des *Referenzrahmens* und von *IEF/lingualevel*, die hinsichtlich der verschiedenen Gesichtspunkte ausgewertet wurden. In den folgenden Kapiteln zu den einzelnen Kompetenzbereichen wird konkreter darauf eingegangen.

Es wurden auch Entscheidungen bezüglich der Aufgabenformen und -formate getroffen:

- Als Testinstrumente werden zur Überprüfung der einzelnen Kompetenzbereiche ausschliesslich Kommunikationsaufgaben (*communication tasks*) eingesetzt.

Zusätzlich gibt es pro getestete Sprache auch einen 20-Minuten-Block mit einem C-Test, der sich nicht von vornherein auf einen einzelnen der getesteten Kompetenzbereiche bezieht. Der C-Test ist eine Form des Lückentests, die zum integrativen Testen von Sprachkompetenz eingesetzt wird. Dem C-Test wird nachgesagt, dass er geeignet sei, sehr effizient und recht zuverlässig entweder eine *globale* Aussage über die Kompetenzen eines Individuums in einer be-

stimmten Sprache zu machen oder aber eine Aussage über die Fähigkeiten im Umgang mit Schriftlichkeit, insbesondere im Schreiben. Der C-Test ist bisher vor allem bei Erwachsenen mit eher gut ausgebildeten literalen Fähigkeiten, wie Anwärtern für ein Hochschulstudium, untersucht worden<sup>18</sup>. Mit dem Einsatz eines (leicht reduzierten) C-Tests in HarMoS sollte ursprünglich dessen (diagnostisches) Potenzial im Zusammenhang mit jüngeren Lernenden ausgelotet werden. Entsprechende Analysen wurden aber vorläufig zurückgestellt.

- Die Kommunikationsaufgaben werden für die Schülerinnen und Schüler kurz situiert, d. h. es wird ein Bezug geschaffen zu plausiblen Situationen ausserhalb des Tests, in denen eine solche oder ähnliche Aufgabe möglicherweise bearbeitet werden muss. Dadurch soll eine Haltung aufgebaut werden, die für die Aufgabenbearbeitung motiviert und dem Schüler bei der Fokussierung auf das Erwartete hilft.
- Als Sprache für die Erklärungen, Anweisungen und Fragestellungen wird grundsätzlich die lokale Schulsprache verwendet, dies aus verschiedenen Gründen: Erstens muss das Aufgabenverständnis sichergestellt werden in einer Testsituation, in welcher der Schüler oder die Schülerin ganz auf sich gestellt ist; zweitens soll nur da Lesekompetenz getestet werden, wo es auch um Lesekompetenz geht, dort wiederum soll Lesekompetenz gezielt in Bezug auf die Texte und nicht in Bezug auf die Aufgabenformulierungen überprüft werden (z. B. in Multiple-Choice-Aufgaben); drittens soll verhindert werden, dass Aufgaben nur dank (u. U. unverstandenen) Wörtern aus den Aufgabenstellungen, richtig gelöst werden.
- Beim Sprechen und Schreiben werden sehr offene Aufgabenformate gewählt, d. h. keine Lückentexte oder Verfahren wie Prompts, die im Wortlaut vorgegeben sind und auf die nur ein sehr kleines Spektrum von Antworten möglich ist.
- Im Leseverstehenstest wird eine Reihe von Aufgabenformaten eingesetzt, dies auch in der Absicht, den potenziell verzerrenden Einfluss eines einzelnen Aufgabenformats zu neutralisieren bzw. mit Hilfe der Analyse der Ergebnisse überhaupt erkennen zu können. Generell wird aber darauf geachtet, dass der Korrekturaufwand relativ klein bleibt, dass also keine eigentlichen Ratings von längeren offenen Antworten nötig werden.

Die Anzahl der möglichen Aufgaben (bzw. Items) wurde konkret von der insgesamt für die Fremdsprachen vorhandenen Testzeit bestimmt. Durch Niveauschätzungen zu den Aufgaben wurde im Vorfeld darauf geachtet, dass für das gesamte Leistungsspektrum der zu testenden Schülerinnen und Schüler relevante Aufgaben vorhanden waren. Um eine durchgehende Kompetenzskala zu ermöglichen, wurde ein Teil der Aufgaben jedes Fertigungsbereichs sowohl in der sechsten als auch in der neunten Klasse eingesetzt. Dabei war bei den niveaumässig „breiteren“ Aufgabenstellungen zum Sprechen und Schreiben die Überlappung grösser als beim Leseverstehen.

Die Aufgaben wurden in einem mehrstufigen Verfahren entwickelt. In bereichsspezifischen Untergruppen wurden erste Entwürfe vorgeschlagen und anschliessend dem ganzen Konsortium zur Begutachtung vorgelegt. Nach weiteren Überarbeitungsschritten wurde unter den Aufgaben eine Auswahl getroffen, die dann unter sehr grossem Zeitdruck in Klassen erprobt wurde. Je nach Bereich erfolgte diese Erprobung in unterschiedlichem Umfang:

- Leseverstehen: D-CH: 7 Lehrer/innen - 8 Klassen; W-CH: 8 Lehrer/innen - 9 Klassen;
- Schreiben: D-CH: 16 Lehrer/innen - 18 Klassen; W-CH: Konsultation von 3 Fachdidaktiker/innen;
- Sprechen: verschiedene Schülerinnen und Schüler aus einer Deutschschweizer Primarschule und einer Deutschschweizer Sekundarschule; dazu weitere Erprobung der Aufgaben anlässlich des Interlokutorentainings unter Einbezug von 6 Deutschschweizer und 6 Westschweizer Schüler/innen.

Diesen Erprobungsschritten folgten gründliche Analysen und Überarbeitungen der Aufgaben.

---

<sup>18</sup> Die Website [www.c-test.de](http://www.c-test.de) gibt einen guten Überblick über Theorie und Praxis des C-Tests. In der Literaturliste finden sich zahlreiche wichtige Publikationen zum Thema, beispielsweise von R. Grotjahn.

Die Übersetzung der Aufgaben wurde zum Teil innerhalb des Konsortiums und zum Teil durch die EDK sichergestellt. Bei den Aufgaben zum Schreiben und Sprechen fielen fast ausschliesslich Übersetzungen von Deutsch auf Französisch an; Inputtexte wurden keine gegeben, die in die drei Zielsprachen hätten übersetzt werden müssen. Beim Leseverstehen fiel neben der Übersetzung von Anweisungen, Situierungen und Aufgabenstellungen auf Französisch auch die Übersetzung eines Teils der Lesetexte an. Den Ausgangspunkt der Aufgabenkonstruktion bildeten jeweils authentische Texte, die im Original entweder deutsch, französisch oder englisch waren. Je nachdem, für welche anderen Zielsprachen diese Texte eingesetzt wurden, mussten auch Übersetzungen vorgenommen werden. Eine besondere Herausforderung dabei war es, Texte mit vergleichbaren Eigenschaften zu schaffen.

Bei den Aufgaben, die in die repräsentative Untersuchung eingingen, folgte ein langwieriger Prozess in Zusammenarbeit mit Grafiker und Verlag; zudem entstanden zeitraubende Anlaufschwierigkeiten. Das Ergebnis war aber gut.

Die repräsentative Untersuchung selbst lag in der Verantwortung der EDK. Die Fremdsprachen waren nur in der Deutschschweiz und in der Westschweiz Teil der Untersuchung, nicht aber im Tessin. Die konkrete Durchführung der Tests wurde nicht vor Ort begleitet, sondern in die Verantwortung der beteiligten Schulen gegeben. Die Konsequenzen davon sind schwer abzuschätzen, aber es liegen zum Beispiel nur einige wenige Anhaltspunkte dafür vor, dass dieser Spielraum dafür genutzt worden wäre, um die Schülerinnen und Schüler konkret auf die Testaufgaben vorzubereiten.

Bei den Feldtests zum Sprechen begab sich jeweils ein Dreierteam vor Ort – zwei muttersprachliche Sprecher/innen der getesteten Sprache und zusätzlich jemand aus dem Konsortium, der die organisatorische und fachliche Verantwortung trug. Die Schülerbeiträge wurden alle in voller Länge gefilmt.

Die Ergebnisse der Untersuchung wurden nach Eingang durch temporäre Hilfskräfte korrigiert bzw. „geratet“ und kodiert. Für die statistische Analyse war bei den Ergebnissen aus den repräsentativen Tests die Methodologiegruppe zuständig. Die Ergebnisse der Feldtests zum Sprechen wurden, soweit überhaupt möglich, konsortiumsintern statistisch ausgewertet. Die Auswertung insbesondere des Teils Schreiben gestaltete sich als so aufwändig, dass erst im Rahmen eines Zusatzmandats erste Rohergebnisse vertieft analysiert werden konnten. Auch bezüglich der übrigen Teile konnten vertiefende Analysen zur Erklärung der empirischen Hauptbefunde noch nicht in der wünschbaren Breite realisiert werden. Gerade die Erkenntnisse in diesem Bereich sind aber wichtig, um Aussagen machen zu können über einen Kernbereich der Testvalidierung, nämlich die Beziehung zwischen Kompetenzmodell/Konstrukt und Test.

In den folgenden Kapiteln wird in unterschiedlichem Umfang auf die Untersuchungen zu den einzelnen Kompetenzbereichen eingegangen. Die Darstellungen folgen im Wesentlichen den Punkten, die in dieser Einführung angesprochen wurden: Kompetenzverständnis/Konstrukt; Aufgabenentwicklung, Datenerhebung; Korrektur/Rating; Analysen und Ergebnisse. Zum Schluss wird jeweils ein kurzer Ausblick auf noch zu leistende Arbeiten gegeben<sup>19</sup>.

### **3.4 Hörverstehen**

Zum Hörverstehen konnten im Rahmen von HarmoS keine Untersuchungen durchgeführt werden.<sup>20</sup> Das Hören ist aber ein zentraler Bereich der Fremdsprachenkompetenz, dem besonders auch beim frühen Sprachenlernen eine Schlüsselrolle zukommt. In der Sprachverwendung hat das Hörverstehen nicht nur als separate Fertigkeit (und als Hörsehverstehen) eine wichtige Funktion,

---

<sup>19</sup> Obschon sich die einzelnen beschriebenen Entwicklungs- und Validierungsschritte in den konzeptuellen Rahmen des *Evidence-centred Design* (s. 3.3.1) einordnen lassen, wird im Folgenden nicht mehr explizit darauf Bezug genommen.

<sup>20</sup> Zwar konnte der Aspekt „Hörverstehen in der direkten Interaktion“ in den Interviewgesprächen im Rahmen der Feldtests zum Sprechen mit beobachtet werden, aber von einer Analyse dieses Aspekts wurde aus Zeitgründen und auch wegen methodischer Bedenken abgesehen.

sondern auch in der mündlichen Interaktion und bei der Sprachmittlung. Deshalb hat das Hörverstehen einen festen Platz im Kompetenzmodell *Fremdsprachen* und deshalb wurden auch für diesen Kompetenzbereich Basisstandards vorgeschlagen.

Um das Hörverstehen wenigstens grob zu situieren, werden an dieser Stelle sehr kurz und summarisch ausgewählte Positionen in der Literatur referiert, die sich auf Parallelen und Unterschiede zwischen dem Hören und dem Lesen und zwischen erst- und fremdsprachlichem Hören beziehen. Im Anschluss daran werden die kommunikativen Aufgaben zum Hörverstehen aus dem Projekt *IEF/lingualevel* etwas genauer vorgestellt, weil die Vorschläge für Basisstandards mit Aufgaben aus diesem Projekt illustriert wurden.<sup>21</sup>

In der Literatur wird davon ausgegangen, dass es zwischen den Prozessen des Hörverstehens und des Leseverstehens (vgl. Kap. 3.5) weitergehende Parallelen gibt, dass aber den „Echtzeit-Prozessen“ zur Verarbeitung der sprachlichen Informationen beim Hören noch grössere Bedeutung zukommt als beim Lesen (u. a. Weir 2005, Rost 2002, Buck 2001, Solmecke 2000; vgl. auch Nold/Rossa 2006). Beim Hören ist die zu verarbeitende Information flüchtig und unterliegt nicht den gleichen Kontrollmöglichkeiten wie beim Lesen. Beim Lesen können schwierige Textstellen mehrmals oder besonders genau betrachtet oder aber übersprungen werden. Beim Hören ist das nicht möglich und es ergeben sich schnell Verstehensprobleme, wenn z. B. schnell oder undeutlich gesprochen wird oder wenn kontextuelles Wissen fehlt. – Beim Hörverstehen in der Fremdsprache sind „Echtzeit-Prozesse“ des Sprachverstehens noch anfälliger für Beeinträchtigungen als in der Erstsprache. Das hängt damit zusammen, dass die sprachlichen Ressourcen in der Fremdsprache oft noch lückenhaft und weniger schnell verfügbar sind als in der Erstsprache. Hinzu kommt, dass Strategien zur Anpassung des fremdsprachlichen Hörens an die Erfordernisse der Hörsituation, Hörtexte und Aufgabenstellungen nicht ohne weiteres aus der Erstsprache transferiert werden und in der Fremdsprache oft erst bewusst gemacht bzw. entwickelt werden müssen.

In *IEF/lingualevel* wird das Hörverstehen in enger Anlehnung an den *Europäischen Referenzrahmen* als „rezeptive kommunikative Aktivität“ aufgefasst, die mit verschiedenen Hör-Absichten verbunden sein kann. Solche Absichten sind: „global verstehen (erfahren, was insgesamt gemeint ist), selektiv verstehen (eine ganz bestimmte Information erhalten), detailliert verstehen (das Gesprochene in allen Einzelheiten verstehen), Schlussfolgerungen ziehen können usw.“ (Europarat 2001, S. 71). Mit geeigneten Aufgabenstellungen, Situierungen, Texten und Formaten wurde versucht, solche Absichten in Aufgaben umzusetzen. Die Hörabsichten wurden allerdings nicht systematisch variiert.

Die Aufgabenstellungen sind weit gehend aus den *IEF/lingualevel*-Kompetenzbeschreibungen für das Hören abgeleitet und sind möglichst so gestaltet, dass sie alltäglichen Hörkontexten entsprechen. Mit geeigneten Situierungen wird versucht, möglichst authentische Hörhaltungen zu erzeugen, die für die Schüler/innen relevant sind oder sein könnten. Beispiele solcher Situierungen sind: „Du hast eine Radiosendung mit einer jungen Sportlerin aufgenommen.“; „Du willst wissen, wie das Wetter an einigen Orten in der Westschweiz ist. Du hast bereits einmal den telefonischen Wetterdienst angerufen und Notizen gemacht. Jetzt rufst du noch einmal an und überprüfst deine Notizen.“; „Du bist zu einem kleinen Fest bei Julie, einer Bekannten aus Lausanne eingeladen. Du hörst ein paar kurze Gespräche.“; „Du hörst zufällig in einem Geschäft, wie ein junger Westschweizer ein Computerspiel bestellt.“

Ausgangspunkt für die Aufgaben waren wenn immer möglich authentische Hörtexte, die so vorgefunden oder dann entsprechend der Vorlage "nachgestellt" wurden. Alle Texte wurden von mutter-

---

<sup>21</sup> Das Projekt IEF – Instrumente für die Evaluation von Fremdsprachenkompetenzen – wurde von den Deutschschweizer Regional-EDKs NW EDK, EDK-Ost, BKZ (Koordination) beauftragt und finanziell unterstützt. Bearbeitet wurde das Projekt von Peter Lenz und Thomas Studer am Lern- und Forschungszentrum Fremdsprachen der Universität Freiburg/CH. Laufzeit des Projekts war 2002 bis 2005. Im Projekt wurden in Zusammenarbeit mit zahlreichen Unterrichtenden und deren Klassen Selbst- und Fremd-Beurteilungsinstrumente für den Französisch- und Englischunterricht ab der 5. und bis zur 9. Klasse entwickelt und erprobt, die für verschiedene Beurteilungszwecke gebraucht werden können. Die Instrumente umfassen Kompetenzbeschreibungen und Testaufgaben für die fünf Fertigkeiten *Hören*, *Lesen*, *interaktives Sprechen*, *zusammenhängendes Sprechen* und *Schreiben* sowie Beurteilungsraster und kommentierte Referenzleistungen für die produktiven Fertigkeiten. Diese Instrumente sind durch wissenschaftliche Methoden den Referenzniveaus des *Gemeinsamen europäischen Referenzrahmens für Sprachen* des Europarates zugeordnet. Publiziert wurden die Instrumente unter dem Namen *lingualevel* vom Berner Schulverlag (vgl. [www.lingualevel.ch](http://www.lingualevel.ch)). Allgemeiner Darstellungen des Projekts sind Lenz/Studer 2005; Studer/Lenz/Mettler 2004.

sprachlichen Sprechern gesprochen. Die Sprecher/innen stammen aus verschiedenen Regionen, in denen die Zielsprache gesprochen wird. Eine breitere Variation gibt es insbesondere für Englisch (USA, Irland, England, Indien). Bei Französisch dominiert die Westschweizer Varietät.

Die Texte beziehen sich primär auf private, schulische und öffentliche Bereiche der Sprachverwendung. Vertreten sind v.a.: Alltagsgespräche (*Face-to-face*-Interaktionen, Telefondialoge und Unterhaltungen zwischen mehreren Personen), Radio-Sendungen, Interviews und Durchsagen. Je nach Sprachverwendungsbereich tendiert der sprachliche Input eher zu konzeptuell mündlichen oder zu konzeptuell schriftlichen Texten. Konzeptuell mündliche Texte (z. B. Alltagsgespräche) weisen in einem stärkeren Mass spezifische Phänomene gesprochener Sprache wie z. B. Lautassimilationen, Ellipsen, Verzögerungsphänomene und Rezeptionssignale auf. Konzeptuell schriftliche Texte (z. B. Radionachrichten oder Durchsagen) sind mehr oder weniger stark an geschriebenen Texten orientiert und sind oft durch Merkmale genauerer Planung charakterisiert (u. a. klarer Textaufbau, sprachinterne Verweise, oft auch komplexere Syntax). Insgesamt wird in den Hörtexten ein breites Spektrum von Sprachfunktionen realisiert; beschreibende und informierende Texte kommen ebenso vor wie instruktive, narrative und argumentative Texte sowie Texte zur Kontaktherstellung und -erhaltung.

Die *IEF/lingualevel*-Aufgaben umfassen die folgenden Antwort-Formate: Richtig-Falsch, Kurzantwort (teils offen, teils durch Stichwörter oder Sätze gelenkt) sowie Multiple-Choice. Die Wahl dieser Formate ist teilweise auch ein Kompromiss zwischen testtheoretischen Ansprüchen und der Natürlichkeit des Ansatzes von *IEF/lingualevel*. So lässt sich z. B. gegen das Richtig-falsch-Format anführen, dass die Ratewahrscheinlichkeit recht hoch ist. Andererseits aber widerspiegelt eine der häufigsten *IEF/lingualevel*-Aufgabenstellungen bei Richtig-falsch-Items, nämlich das Überprüfen von Erwartungen oder Hypothesen (das in der realen Situation auf Vorwissen gestützt ist), eine authentische und durchaus übliche Hörhaltung. Deshalb wurde auch mit dem Richtig-falsch-Format gearbeitet.

Um die Gefahr der (kognitiven) Überforderung der Schüler durch die Hörverstehens-Aufgaben zu reduzieren, werden die Hörtexte entweder zweimal eingespielt (v.a. dann, wenn das in Nicht-Test-situationen auch möglich wäre, z. B. bei aufgenommenen Sendungen oder bei Durchsagen) und/oder es sind Pausen in die Hörtexte hineinmontiert (v.a. dann, wenn eine Kurzantwort aufgeschrieben werden muss). Auf diese Weise haben die Schüler/innen Zeit, die Aufgaben zu lösen und sich auf den nachfolgenden Hörtext einzustellen.

Sämtliche Aufgabenstellungen sind auf Deutsch (also in der lokalen Schulsprache der Deutschschweiz und nicht in der Zielsprache) formuliert. Dadurch soll erstens das Aufgabenverständnis sichergestellt werden, zweitens vermieden werden, dass Leseverstehen statt Hörverstehen getestet wird, und drittens ausgeschlossen werden, dass Antworten nur gegeben werden, weil im Hörtext Wörter aus der gelesenen Aufgabenformulierung wiedererkannt werden. Allgemein handelt es sich beim Einbezug der lokalen Schulsprache in die Testaufgaben um eine konsequente Massnahme zur Erhöhung der Validität der Aufgaben, keineswegs um eine Gegenbewegung gegen das durchaus positive Bestreben vieler Lehrpersonen und Didaktiker/innen, die Zielsprache im Unterricht auch zur Umgangssprache zu machen.

Die *IEF/lingualevel*-Aufgaben zum Hörverstehen wurden zwischen 2004 und 2005 entwickelt. Sie wurden sowohl qualitativ als auch mit statistischen Methoden validiert. Im Wesentlichen umfassten Aufgabenkonstruktion und -validierung die folgenden Schritte:

- Studium der Fachliteratur und Analyse vorhandener Aufgaben (darunter Aufgaben in Lehrwerken, Aufgabensammlungen und Zertifikatsprüfungen);
- Konstruktion von Aufgabenentwürfen auf der Basis der Deskriptorensammlung, die in der ersten Projektphase von *IEF* erarbeitet wurde<sup>22</sup>; Produktion von Hörtexten; Erstellen von Durchführungsbestimmungen und Lösungsschlüsseln;

---

<sup>22</sup> Teil *Hören* der strukturierten Deskriptoren-Datenbank, die insgesamt rund 250 zielgruppenrelevante, empirisch validierte Kompetenzbeschreibungen für die fünf auch in HarmoS unterschiedenen Sprachfertigkeiten umfasst. Die Deskriptoren sind den Feinniveaus des Europäischen Referenzrahmens zugeordnet. Zur Entwicklung der Deskriptoren siehe Lenz/Studer (2004).

- Qualitative Erprobungen der Aufgaben in Schulklassen; Überarbeitung der Entwürfe; Ergänzung des Aufgabenkorpus und der Lösungsschlüssel; Finalisierung der Durchführungsbestimmungen;
- Durchführung von Hörverstehenstests in insgesamt ca. 85 Klassen mit Schülern unterschiedlichen Alters und unterschiedlicher Niveaus. Viele der Lernenden bearbeiteten nicht nur Testaufgaben, sondern füllten zusätzlich einen Selbstbeurteilungsbogen aus, bestehend aus IEF-Kompetenzbeschreibungen, die bereits früher den Referenzniveaus zugeordnet worden waren.
- Kalibrierung der Daten zu den Tests und zur Selbstbeurteilung mit *Winsteps* bzw. *Facets*. Eliminierung problematischer Items. Auf der Basis der Selbstbeurteilungen Zuordnung der Aufgaben zu den Referenzniveaus.

Resultat dieser Entwicklungsarbeiten waren 75 empirisch validierte Testaufgaben (460 validierte Items), davon je ca. die Hälfte für Französisch und Englisch. Zwei Drittel dieser Aufgaben wurden in das Verlagsprodukt *lingualevel* übernommen, ein Drittel der Aufgaben wurde von den Kantonen vorläufig für besondere Evaluationsprojekte zurückbehalten.

### 3.5 Leseverstehen

Das fremdsprachliche Leseverstehen wird im Rahmen von HarmoS als kommunikative Sprachfertigkeit aufgefasst, an deren Erwerb der Fachunterricht in Fremdsprachen einen wesentlichen Anteil hat. Im Folgenden werden die Untersuchungen genauer vorgestellt, die zum Leseverstehen durchgeführt wurden. Mit diesen Untersuchungen sollte eine empirische Grundlage für die Setzung und Illustrierung von Basisstandards für das Lesen geschaffen und ein Beitrag zur fertigkeitsspezifischen Konkretisierung und Validierung des Kompetenzmodells *Fremdsprachen* geleistet werden.

#### 3.5.1 Konzeptuelle Grundlagen und Konstrukt

Das Konstrukt *fremdsprachliche Lesekompetenz* basiert besonders auf drei Konzepten, die wesentliche und für die Zielgruppen relevante Aspekte des Lesens beschreiben (zum grundsätzlich selektiven Charakter eines Konstrukts vgl. die Überlegungen zum *student model* unter 3.3.1). Diese Konzepte sind: a) Prozesse des Lesens, b) ein funktionaler Literalitätsbegriff und, damit verbunden, Handlungsaspekte des Lesens sowie c) Texttypen. a) und b) tragen den Besonderheiten des Kompetenzbereichs *Leseverstehen* Rechnung, bei c) handelt es sich um eine für das Lesen adaptierte Variante der Diskurstypen des Kompetenzmodells *Fremdsprachen*, nach denen sich alle unterschiedenen Kompetenzbereiche sinnvoll gliedern lassen. Ebenfalls in einer für das Lesen spezifizierten Variante werden im Konstrukt *Leseverstehen* die Kompetenzressourcen des Kompetenzmodells *Fremdsprachen* berücksichtigt (vgl. dazu unten unter c).

##### a) Prozesse des Lesens

Lesen ist eine komplexe und dynamische Fertigkeit, die auf das Verstehen schriftlicher Texte gerichtet ist. Beim Lesen greifen eine Vielzahl von sprachlichen und kognitiven Prozessen ineinander und interagieren mit verschiedenen Arten von Wissen. Unterschieden werden oft Prozesse, die bei den Sprachformen ansetzen und die beim geübten Leser weitgehend automatisiert ablaufen (u. a. Worterkennung, Zugriff auf das mentale Lexikon, syntaktische Analyse) und Prozesse zur (Re-)Konstruktion von Bedeutungen sowie zum Erschliessen von Inhalten und Zusammenhängen, die von kognitiven Schemata gesteuert werden und die in der Regel mehr Aufmerksamkeit beanspruchen (Alderson 2000, Wolff 2002; vgl. auch Nold & Rossa 2006). Strategien kommt beim Lesen u. a. die wichtige Aufgabe zu, diese Prozesse in Bezug auf die Erfordernisse einer Lesesituation zu regulieren (Phakiti 2007). Das bedeutet auch, dass der Anteil verschiedener Prozesse am Verständnis eines Textes (oft „mentales Modell“ des Textes genannt) unterschiedlich gross sein kann. Ehlers (2006, 35) bezeichnet die strategische Fähigkeit, das Leseverhalten den Anforderungen einer Situation, einer Aufgabe und des Textes anpassen zu können, als „kognitive Flexibilität“ und diskutiert in dem Zusammenhang die in der Praxis gut bekannten „Leseformen“ wie „selektierendes“ und „detailliertes Lesen“.

Im Grundsatz werden diese Prozesse für das Lesen in der Erstsprache *und* in den Fremdsprachen postuliert. Beim Lesen in der Fremdsprache werden die Prozesse aber von verschiedenen Faktoren beeinflusst, insbesondere von den Kenntnissen einer Fremdsprache (v.a. Wortschatz und Grammatik) und von der generellen Lesefertigkeit, die meist über die Erstsprache aufgebaut wird. Als typisch für das Lesen in der Fremdsprache werden geringere Lesegeschwindigkeit und oft

auch reduzierte Inferenzbildung angesehen, was in der Regel mit noch wenig automatisierten Prozessen auf den „unteren“ Verarbeitungsebenen (z. B. Worterkennung) und eingeschränkten Wortschatzkenntnissen sowie mit nur teilweise verfügbaren Wissensbeständen in Zusammenhang gebracht wird (z. B. Ehlers 2006).

Vor dem Hintergrund der allgemeinen Kompetenzdefinition in HarmoS (vgl. 3.1) ist für das Konstrukt *fremdsprachliche Lesekompetenz* von besonderem Interesse, ob es den Schüler/innen gelingt, das Lesen an die Anforderungen der Aufgaben und Texte anzupassen: Können sie z. B. den einen Text sorgfältig lesen, z. B. um Erklärungen oder Anweisungen genau zu verstehen, und einen anderen „überfliegen“, um das Thema des Textes zu erkennen? Hinsichtlich dieser Frage wird auf der Basis von Urquhart & Weir 1998, Weir 2005 und Sigott 2007 grob unterschieden zwischen *sorgfältig-genauem Lesen* („careful reading“) einerseits sowie *erkundendem und selektivem Lesen* („expeditious reading“) andererseits.<sup>23</sup> Sorgfältig-genaues Lesen zielt auf genaues Verstehen bzw. auf Verstehen im Detail und ist nicht selektiv. Erkundendes und selektives Lesen umfasst mehrere Arten des „schnellen“ Lesens, v.a. „skimming“ („Überfliegen“ des Textes oder von Textpassagen, z. B. um das Thema des Textes zu erkennen) und „scanning“ (Durchsuchen eines Textes, um Einzelinformationen und Details aufzufinden).

#### b) Funktionale Literalität und Handlungsaspekte des Lesens

Die Arbeitsgruppe Lesen des Fremdsprachkonsortiums HarmoS nimmt diese „Prozesse“ des Lesens – mit dem Fokus auf sorgfältig-genauem sowie erkundendem und selektivem Lesen – als Ausgangspunkt, stellt sie aber in den grösseren Rahmen einer sog. „reading literacy“. In diesem Rahmen kommen dem sozialen Kontext des Lesens und besonders den Lesenden als sozialen Akteuren grössere Bedeutung zu (u. a. Urquhart & Weir 1998). Da Literalität in der Regel mehr meint als Lesen (besonders auch Schreiben, vgl. Kap. 3.6, manchmal aber auch noch „numeracy“), ist die begriffliche Engführung *reading literacy* nötig. Mit Blick auf die *Basisstandards* versteht die Arbeitsgruppe unter „literacy“ eine Form von funktionaler Literalität, d. h. im Zentrum stehen der schriftliche Alltag der Schüler/innen und die Frage, wie die Schüler/innen in den verschiedenen Handlungsbereichen („domains“) ihres schriftlichen Alltags zurecht kommen (sollten). So gefasste Literalität hat zwei grosse Vorteile: Zum einen kann Lesen im Einklang mit dem Kompetenzmodell des Konsortiums Fremdsprachen als kommunikative *Sprachverwendung* (und nicht „nur“ als psycholinguistischer Prozess) konzipiert werden, zum andern wird ein Anschluss an das Reading-Literacy-Konzept von PISA hergestellt. Dieses Konzept wurde zwar für die Erstsprache entwickelt, ist aber für das Sprachenlernen allgemein relevant und kann deshalb in manchen Punkten auch zur Differenzierung des Konstrukts *fremdsprachliche Lesekompetenz* genutzt werden. Die zentrale PISA-Definition für das Lesen lautet: „Reading literacy is understanding, using and reflecting on written texts, in order to achieve one’s goal, to develop one’s knowledge and potential and to participate in society“. (OECD, 2003, 108)

In enger Anlehnung an PISA werden zur weiteren Differenzierung des Konstrukts drei Handlungsaspekte des Lesens unterschieden<sup>24</sup>: *Informationen entnehmen*, *Interpretieren* und *(Sich) in Beziehung setzen* oder auch „*Evaluiieren*“ (OECD 2003, 2002). Einige Stichworte dazu: Beim Aspekt *Informationsentnahme* geht es darum, in einem Text Informationen zu finden und zu verstehen. Dabei wird der Text in der Regel überflogen und durchsucht, um relevante Informationen zu lokalisieren und auszuwählen. Beim Aspekt *Interpretieren* steht einerseits ein „breites“ bzw. allgemeine-

---

<sup>23</sup> In der Fachliteratur wird eine unterschiedliche Anzahl solcher ‚Prozesse‘ angenommen und die Terminologie ist sehr uneinheitlich (häufiger gebraucht werden: „operations“, „skills“, „strategies“, „comprehending strategies“, „subskills“, „Leseformen“, „Lesearten“). Nach Phakiti (2007, 6) würde es sich um eine kognitive Strategie handeln. Da der Ausdruck ‚Strategie‘ z.T. stärker mit bewusstem Handeln in Verbindung gebracht wird und weil der Bewusstseinsstatus der „Leseformen“ bei den Zielgruppen nicht untersucht ist, sprechen wir hier nicht von Strategien, sondern behelfsmässig von Prozessen (und manchmal auch, mit Blick auf didaktische Traditionen, von Lesearten). Prozesse gelten in der Psycholinguistik in der Regel als dem Bewusstsein prinzipiell zugänglich, können aber auch unterhalb der Schwelle bewusster Aufmerksamkeit ablaufen. – Die Untersuchung von nur zwei (Gruppen von) Leseprozessen oder –arten ist ein Kompromiss zwischen theoretisch motivierter Differenziertheit und praktischer Operationalisierbarkeit des Konstrukts. Auch im Teilprojekt Lesen zu den österreichischen Bildungsstandards werden zwei Leseprozesse unterschieden (Sigott 2007).

<sup>24</sup> PISA selbst spricht von „Prozessen“ oder „Aspekten“ des Lesens.

res Verständnis des gesamten Textes im Vordergrund („form a broad general understanding“; z. B. Erkennen von Bezügen im Text, um den übergreifenden Gedanken zu verstehen) und andererseits ein genaueres oder umfassenderes Verständnis des Gelesenen, wozu Zusammenhänge zwischen Textteilen oder Texten nachvollzogen und/oder erschlossen werden müssen („develop an interpretation“). Während beim Interpretieren textinterne Informationen verarbeitet werden, geht es beim Aspekt (*Sich*) in Beziehung setzen bzw. „Evaluieren“ von Texten darum, für das Verstehen textexternes Wissen beizuziehen, d. h. die Informationen in einem Lesetext mit Kenntnissen aus anderen Quellen (z. B. andere Texte oder eigenes Weltwissen) zu vergleichen und zu verknüpfen.<sup>25</sup>

### c) Texttypen

Um das Konstrukt *fremdsprachliche Lesekompetenz* unter den Vorzeichen einer funktionalen Literalität hinreichend abzudecken (vgl. Kap. 3.2), wurde ein breites Textsortenspektrum berücksichtigt. Beispielsweise kommen im Korpus Medientexte ebenso vor wie Texte aus dem schulischen und aus dem privaten Bereich, erzählende Texte ebenso wie argumentative und zusammenhängende Sachtexte ebenso wie „diskontinuierliche Texte“. Ausgehend von den Diskurstypen des Kompetenzmodells Fremdsprachen sowie von Projekten mit vergleichbaren Fragestellungen (darunter der „Dutch CEFR Grid Reading/Listening“, vgl. <http://www.lancs.ac.uk/fss/projects/grid/>) wurden die sonst so genannten „Textsorten“ auf der Basis von linguistischer Literatur (v.a. Stede 2007) spezifischer gefasst als *Texttypen*.<sup>26</sup>

Unterschieden wurden fünf Texttypen. Unterscheidungskriterien waren sprachliche Merkmale der Texte und Textpassagen (u. a. typische Satzarten und typische morphosyntaktische Merkmale). Die fünf Typen sind: *deskriptiv* und *expositorisch* (diese beiden Typen wurden aus praktischen Gründen zusammengefasst), *narrativ*, *argumentativ* und *instruktiv*. Hinzu kommen *diskontinuierliche Texte*, d. h. nicht im üblichen Sinn zusammenhängende Texte wie z. B. Verzeichnisse oder Tabellen. Diskontinuierliche Texte wurden einbezogen, weil sie unter dem Gesichtspunkt einer funktionalen Literalität für die Schüler/innen bedeutsam sind. Diese Texte werden aber als separate Gruppe behandelt, weil sie semiotisch vielfältiger sind als die „klassischen“ Texttypen und deshalb teilweise anders beschrieben werden müssen. Offen ist auch, inwieweit am Verstehen diskontinuierlicher und zusammenhängender Texte verschiedene Verarbeitungsprozesse oder aber die gleichen mit verschiedenem Gewicht beteiligt sind.

In den folgenden drei Abbildungen ist die Verteilung der Aufgaben nach den Merkmalen Handlungsaspekt und Leseprozess (Abb. 5) sowie Texttyp (Abb. 6) dargestellt. Alle Verteilungen sind auf das Total von 198 Items bezogen (N=198); das sind alle Items, die erfolgreich skaliert werden konnten (ca. vier Items pro Aufgabe).

---

<sup>25</sup> Neben dieser Art des Nachdenkens über Textinhalte setzt PISA als weitere Kategorie eine Reflexion über die Form resp. die Gestalt eines Textes an. Dieser Aspekt wird im Rahmen von HarMoS nicht weiterverfolgt, weil er für die Zielgruppen in den Fremdsprachenfächern (noch) weniger relevant scheint.

<sup>26</sup> Nach Stede (ebd., 47) können sich Texttypen – verstanden als „funktional interpretierbare Bündel linguistischer Merkmale“ – „auf mehr oder weniger lange Textabschnitte beziehen“, wohingegen Textfunktionen auf „das (vom Leser nachvollziehbare) Ziel des Gesamttextes“ bezogen sind. Diese Terminologie trägt dem Umstand Rechnung, dass viele Texte Passagen mit unterschiedlichen Funktionen (erzählende, beschreibende usw.) aufweisen. Für das Kodieren der Lesetexte hat diese Terminologie den Vorteil, dass sich den itemrelevanten Textstellen des gleichen Textes verschiedene *Texttypen* zuweisen lassen.

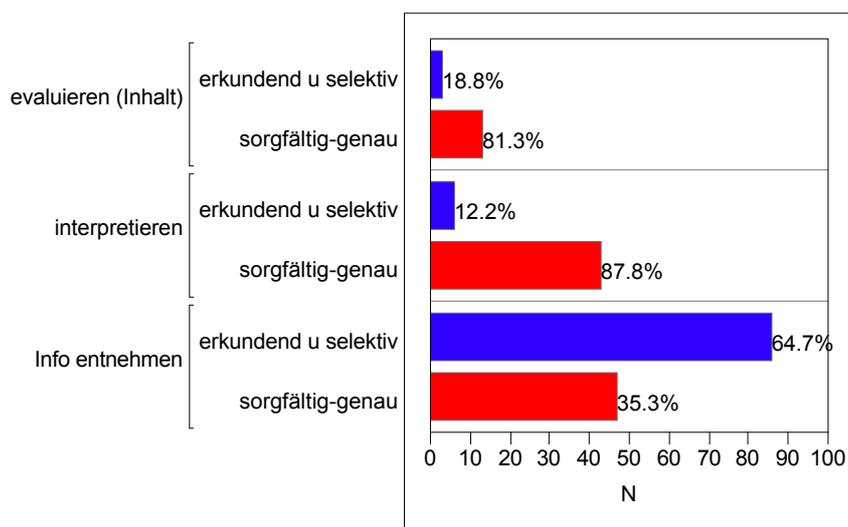


Abb. 5: Verteilung der Items nach Handlungsaspekten und Leseprozessen

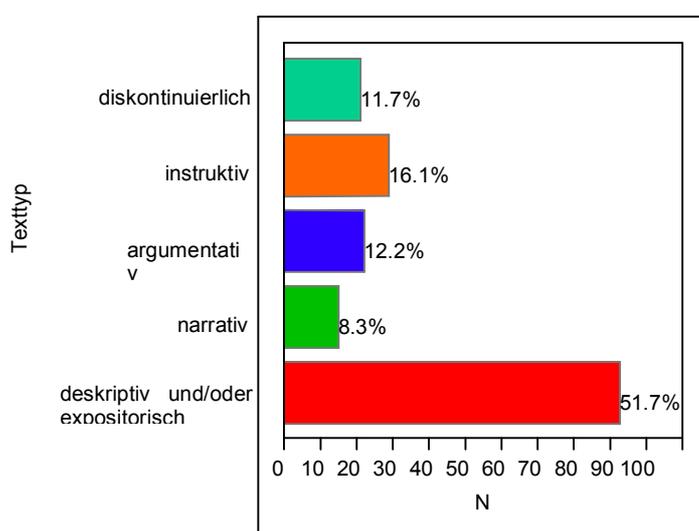


Abb. 6: Verteilung der Items nach Texttypen

Die Abbildungen zeigen, dass bei mehr als der Hälfte der Items eine Form der Informationsentnahme kombiniert mit erkundendem und selektivem Lesen überprüft wurde. Bei ca. einem Viertel der Items stehen das Interpretieren und sorgfältig-genaues Lesen im Vordergrund. Vergleichsweise wenige Items betreffen den Aspekt des (Sich-)In-Beziehung-Setzens bzw. Evaluierens. Ein grosser Teil der Lesetexte und Textstellen hat beschreibenden Charakter, aber auch instruktive und argumentative Texte sind gut vertreten. Diskontinuierliche Texte sind mit gut 10% vertreten, die narrativen Texte bilden die kleinste Gruppe.

Insgesamt umfasst das Konstrukt *fremdsprachliche Lesekompetenz* also die folgenden Komponenten: Zwei Leseprozesse oder -arten (sorgfältig-genaues Lesen sowie erkundendes und selektives Lesen), drei Handlungsaspekte (Informationen entnehmen, Interpretieren und (Sich) in Beziehung setzen) sowie fünf Texttypen (deskriptiv und expositorisch, narrativ, argumentativ und instruktiv) und die Gruppe der diskontinuierlichen Texte. Diese Komponenten fokussieren ausgewählte Aspekte des Lesens, die für das Lesen als *Sprachverwendungskompetenz* allgemein und für den Lese-Erwerb der Zielgruppen in den Fremdsprachen besonders relevant sind. In diesen Komponenten sind die Kompetenz-Ressourcen des Kompetenzmodells *Fremdsprachen* bereits in einer „ganzheitlichen“ Weise integriert. Jeder Lesetext stellt Anforderungen an die *kommunikativen Sprachkompetenzen* der Lesenden (u. a. in Form von einfachen oder schwierigen Strukturen und in Form von häufigeren oder selteneren Wörtern), und beim Lösen einer Aufgabe zum Leseverste-

hen sind immer auch *allgemeine Kompetenzen* beteiligt (z. B. in Form von Wissen über das Thema der Aufgabe). In der hier berichteten Untersuchung werden ausgewählte Aspekte der *linguistischen Kompetenz* (als bedeutende Komponente der *kommunikativen Sprachkompetenz*) genauer betrachtet, und zwar aus der Perspektive der Texte und Aufgaben. Vorgestellt werden diese Aspekte unten im Zusammenhang mit Überlegungen zur Schwierigkeit von Items, weil davon ausgegangen wird, dass sprachliche Merkmale die Schwierigkeit von Items in Verbindung mit anderen Merkmalen wie z. B. bestimmten Lesearten, Handlungsaspekten und Texttypen beeinflussen können (siehe das Stichwort „item maps“ im nächsten Abschnitt). Demgegenüber wird nicht versucht, die *allgemeinen Kompetenzen* zu operationalisieren, sondern diese wenigstens teilweise zu kontrollieren, z. B. durch Abklärung von thematischen Interessen in den durchgeführten Pretests oder auch durch Berücksichtigung verschiedener Antwortformate.

Allerdings wird man aufgrund der Literatur vorsichtig bis skeptisch sein in Bezug auf die (unter 3.3.2 allgemein diskutierte) Frage, ob die hier angesetzten Komponenten des Konstrukts *fremdsprachliche Lesekompetenz* auch auf empirischem Weg unterschieden werden können (vgl. Alderson 2000, 95; Weir 2005, 87).

Weir (2005, 89) setzt den Akzent in der Diskussion um die Unterteilbarkeit bzw. Dimensionalität des Leseverstehens etwas anders: Wichtiger als die Grundsatzdebatte sei ein sorgfältiger Umgang mit den einzelnen Items. „Whatever theoretical position we take, we inevitably measure reading skills/strategies (individually or in combination) as soon as individual items are written on a passage. The debate would seem to be more a question of the strength of any claims concerning which skills/strategies are being tested by which items ..., than a strong claim for treating reading as a unified skill.“

Der Stand der Diskussion um die Dimensionalität des Leseverstehens veranlasste das Fremdsprachenkonsortium zu zwei methodologischen Entscheidungen:

(1) Da unklar ist, ob Konstruktkomponenten empirisch unterschieden werden können, sind nicht-reduktive Tests zur Abklärung einer allgemeineren Leseverstehenskompetenz der sicherere Weg (vgl. auch Alderson 2000, 206f.), besonders, wenn es sich um eine breit angelegte Untersuchung wie bei HarMoS handelt. Tests, die auf isolierte Aspekte des Leseverstehens zielen, könnten z. B. in kleineren, explorativen Studien sinnvoll sein.

(2) Statistische Analysen zur Dimensionalität des Konstrukts sollten auf der Grundlage genau beschriebener Items erfolgen. Im vorliegenden Teilprojekt werden die Items in Bezug auf die fokussierten Komponenten des Konstrukts einschliesslich ausgewählter Aspekte linguistischer Kompetenz beschrieben. Primärer Zweck dieser Beschreibungen ist die Interpretation der empirischen Itemschwierigkeit.

### 3.5.2 Instrumente

Konkretisiert wurde das Konstrukt *fremdsprachliche Lesekompetenz* zunächst durch eine sog. „**construct map**“ (vgl. 3.3.2). Im Teilprojekt Leseverstehen hatte dieses Arbeitsinstrument zunächst die Form von gestuften Merkmalsmatrizen, später auch von Kompetenzbeschreibungen. Quellen der Beschreibungen waren einerseits die vorgestellten Komponenten des Konstrukts und andererseits umfassendere Analysen von Testaufgaben und Deskriptoren aus anderen Projekten. Analysiert wurden u. a. Aufgaben zum Leseverstehen aus den deutschen und den österreichischen Projekten zu Bildungsstandards<sup>27</sup> sowie Aufgaben und Deskriptoren aus dem Deutschschweizer Projekt IEF/lingualevel. In diesen Quellen wurden potentiell schwierigkeitsbestimmende (konstruktrelevante ebenso wie konstruktirrelevante) Merkmale identifiziert. Die konstruktrelevanten Merkmale wurden breiteren, provisorischen Niveaus zugeordnet. In einer frühen Projektphase, die der Aufgabenkonstruktion voranging, sah die *construct map* für das Lesen zum Beispiel so aus (Ausschnitt für ein tiefes Niveau „N1“):

---

<sup>27</sup> Bildungsstandards der Kultusministerkonferenz für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss (9 und 10. Jahrgangsstufe); Beschluss vom 4.12.2003. Bildungsstandards des Österreichischen Bildungsministeriums für Englisch für die 8. Schulstufe; Erprobungsphase bis 2007. Zentrale Lernstandserhebungen für Englisch in der Jahrgangsstufe 9, Nordrhein-Westfalen, Ministerium für Schule, Jugend und Kinder, 2004.

N1 (IEF-LV-Items)	Charakterisierung von Responses				
	Textsorte; Textmerkmale	Types of reading	weiter konstrukt-relevant	Antwort-Format	Konstrukt-irrelevant?
LV22, Le17 LV43, Le03 LV104, Le01 LV23, Le21 LV105, Le17	Zahlen, Wortfolgen, einfache Sätze, ganz kurze Texte Einladungskarte; einfacher Fragebogen, (Verbots-, Warn-)Schilder; kurze, einfache Mitteilungen, Inserate, Plakate	Wiedererkennen, Scanning, careful-local: „local“ dominiert	Angaben zur Person; sehr vertraute Ereignisse und sehr bekannte Themen; sehr häufige, einfache Wörter und Wendungen	ganz einfacher Info-Transfer; matching (v.a. mit Bild)	- Analogie zu Beispiel-Item? (LV22, Item 5) - Weltwissen? (LV104, Item 2) - Bildunterstütz.? (LV23., Item 3: sehr hohe Trefferquote, Item 2: extrem schwer) - Einfluss Distraktor? (LV23, Item 5)

Abb. 7: Ausschnitt aus der *construct map* für das Leseverstehen (frühe Arbeitsphase; mit „responses“ sind hier Rekonstruktionen von Item-Merkmalen gemeint, die das Verhalten der Testteilnehmenden beeinflussen haben könnten; LV22 etc. = validierte Leseverstehensitems aus IEF/ lingualevel, Le17 etc. = validierte Deskriptoren aus demselben Projekt)

Genutzt wurde die *construct map* für die Aufgabenkonstruktion (iterativer Prozess, d. h. erste Aufgabenentwürfe wurden auch zur Differenzierung der *construct map* gebraucht) und als Ressource für die definitiven Niveaubeschreibungen.

Die Entwicklung der **Testaufgaben** wurde von der Arbeitsgruppe Leseverstehen geleistet (Verena Tunger, Daniel Elmiger, Thomas Studer). Diese Arbeit lässt sich in drei Schritten zusammenfassen. Schritt 1: Konstruktion von Aufgabenentwürfen, Schritt 2: Übersetzungen, Schritt 3: Erprobungen und Revision. Für diese drei Schritte standen nur ca. vier Monate Zeit zur Verfügung.

#### Schritt 1: Konstruktion von Aufgabenentwürfen

Im ersten Schritt wurden auf der Basis der *construct map*, von Fachliteratur (z. B. McKay 2006) und Erfahrungen bei der Testkonstruktion in andern Projekten insgesamt 21 Aufgabenentwürfe erarbeitet, je sieben für Deutsch, Englisch und Französisch als Fremdsprachen.

Geachtet wurde bei den Entwürfen v.a. auch auf eine einigermaßen ausgewogene Verteilung der beiden Leseprozesse und der drei Handlungsaspekte sowie der unterschiedenen Texttypen des Konstrukts. Eine ausbalancierte Umsetzung der Komponentenstruktur des Konstrukts in Aufgaben wurde aber nicht angestrebt, u. a. deshalb nicht, weil es zwischen den Konstruktkomponenten sinnvollere und weniger sinnvolle Kombinationen gibt, beispielsweise zwischen Leseprozessen und Texttypen (Weir 2005, 69): Scanning etwa lässt sich gut bezogen auf deskriptive Texte untersuchen, weil beschreibende Texte in der Regel mehr faktische Details aufweisen, auf die das Lesen gelenkt werden kann, als es üblicherweise bei argumentativen Texten der Fall ist. Umgekehrt ist ein Geflecht von satzübergreifenden Zusammenhängen („Makropropositionen“) eher ein typisches Merkmal von argumentativen Texten als von deskriptiven. Deshalb lässt sich erkundendes Lesen oft adäquater am Beispiel von meinungsbetonten als von beschreibenden Texten abklären. Spezifische Beziehungen gibt es auch zwischen Prozessen und Handlungsaspekten, darunter sehr enge (z. B. offensichtliche Verwandtschaft von selektivem Lesen und Informationsentnahme), aber auch nur lose (Interpretieren kann erkundendes und/oder selektives Lesen voraussetzen).

Alle Lesetexte sind authentisch oder nah-authentisch (z. B. leicht gekürzt). Hauptsächlich Quellen für die Texte waren Zeitschriften und Magazine für junge Leser/innen sowie das Internet (Foren, Chat u. a.), aber auch Dokumente und Prospekte aus dem schulischen und aus dem öffentlichen Bereich, die für die Zielgruppen relevant sind oder sein könnten. Durch die Berücksichtigung dieser

verschiedenen Quellen wurde auch eine gewisse Themenvielfalt angestrebt. Thematische Aspekte wurden aber nicht systematisch kontrolliert.

Die Testaufgaben weisen in der Regel einen Handlungsrahmen resp. eine zielgruppenspezifische Situierung auf, z.B.: „Tu veux réaménager ta chambre. Sur internet, tu es tombé-e sur un forum de discussion ... .“ „Ihr plant zusammen mit der Englischlehrerin eine Exkursion ... .“, „Du arbeitest an einem Vortrag ... und hast in einer Jugendzeitschrift ... .“ Situierungen dienen in Testaufgaben v.a. auch dazu, den in Nicht-Testsituationen oft evidenten Zweck des Lesens plausibel zu machen. Auch soll eine gewisse Lesemotivation geschaffen werden, die unter Nicht-Test-Bedingungen gegeben ist. Für Alderson (2000, 83) ist die Nähe zu Leseabsichten in realen Lebenssituationen ein wichtiges Kriterium der Validität eines Tests.

Besondere Sorgfalt verlangt bei der Aufgabenkonstruktion die Wahl des Antwortformats. Neuere Forschung geht davon aus, dass das Antwortformat die Überprüfung des Konstrukts erheblich beeinflussen kann. In der Schusslinie steht v.a. auch das Multiple-Choice-Format (z. B. Rupp/Ferne/Choi 2006, Kobayahi 2002). Problematisch an diesem Format ist nicht nur die Ratewahrscheinlichkeit, sondern v.a. auch der Einfluss derjenigen Antwortvorgaben, die als „Ablenker“ (Distraktoren) dienen. Distraktoren stehen kaum in Zusammenhang mit dem Lesen in Nicht-Test-situationen. Der Einfluss der Distraktoren kann für verschiedene Testteilnehmende unterschiedlich gross sein, ist aber grundsätzlich schwer zu kontrollieren. Auf der andern Seite sind offene Formate mit dem Problem konfrontiert, dass Antworten interpretiert und kategorisiert werden müssen, was trotz Lösungsschlüssel und Ratertraining zu erheblichen Unterschieden in den Bewertungen führen kann. Um allenfalls verzerrende Effekte wenigstens grob kontrollieren zu können, wurde mit drei verschiedenen Formaten gearbeitet: Neben Multiple-Choice-Formaten (oft drei Antwortmöglichkeiten) wurden Vergleichs- und Zuordnungsformate (Multiple-Matching), offene Fragen bzw. Kurzantworten und Kombinationen dieser drei Formate (z. B. Multiple-Matching gefolgt von Kurzantwort) eingesetzt. Auch Alderson (2000, 206) empfiehlt eine Kombination von Formaten, auch Kombinationen von Formaten in der gleichen Aufgabe.

Die Situierungen und die genauen Instruktionen zur Lösung der Aufgaben sowie die Testfragen (Items) selber wurden in der jeweiligen lokalen Schulsprache verfasst, nicht in der Fremdsprache. Damit sollte erstens sichergestellt werden, dass die Schüler/innen verstehen, was sie tun müssen, und zweitens, dass Lesekompetenz in Bezug auf die Texte getestet wird – und z. B. nicht in Bezug auf die Antwortalternativen in einer Multiple-Choice-Aufgabe, die von den Testkonstrukteuren formuliert wurden. Drittens sollte vermieden werden, dass Antworten nur gegeben werden, weil gleiche Wörter in Text und Aufgabenformulierung vorkommen. Dies wäre nur ein Indiz für Wort(wieder)erkennung, aber noch kein sicherer Hinweis auf Verstehen.

## Schritt 2: Übersetzungen

Eine Auswahl der 21 Aufgabenentwürfe („Originale“) wurde von Übersetzer/innen in die anderen Sprachen übertragen. Die Übersetzungen sollten sinngemäss (nicht wörtlich) sein, den Charakter des Textes beibehalten und den Text punkto sprachliche Anforderungen möglichst nicht erheblich verändern. Die Übersetzungen wurden von Mitgliedern des Konsortiums kontrolliert, v.a. um zu garantieren, dass die Items auch in den Übersetzungen eindeutig beantwortbar sind und im Schwierigkeitsgrad nicht erheblich von den Originalen abweichen (z. B. wegen ausgeprägteren idiomatischen Sprachgebrauchs eines Übersetzers). Mit diesem Vorgehen wurden „strukturgleiche“ Aufgaben erstellt, die in allen Zielsprachen eingesetzt werden konnten. Das Korpus mit den Übersetzungen umfasste schliesslich 46 kommunikative Testaufgaben. Sie decken zusammen das Spektrum von Niveaus ab, auf denen sich die Zielgruppen bewegen.

## Schritt 3: Erprobungen und Revision

Die Aufgaben wurden in insgesamt 17 Klassen erprobt. An der Erprobung beteiligt waren neun Klassen in der Deutschschweiz und acht Klassen in der Westschweiz (zwei Klassen des 6. und vier des 9. Schuljahres für Französisch, je drei 6. und 9. Klassen für Deutsch und fünf 9. Klassen für Englisch).

Erprobt wurden die Aufgaben unter quantitativen und qualitativen Gesichtspunkten. Zum einen wurde jede Aufgabe von mindestens zwei und durchschnittlich von drei Klassen gelöst. Auf dieser

Grundlage konnten schlecht funktionierende Items identifiziert werden (Items mit schlechter Modellpassung resp. schlechten „Fit-“ und mangelhaften Diskriminations-Werten). Zum andern wurden bei den Lehrpersonen und bei den Schüler/innen mittels kurzen Fragebogen Einschätzungen der Aufgaben und Items erhoben. Die Lehrpersonen beurteilten die Aufgaben in Bezug auf Verständlichkeit, Attraktivität, Komplexität und vorgesehene Durchführungszeit. Ausserdem vermerkten sie allfällige Probleme der Lernenden beim Bearbeiten der Aufgaben. Die Schüler/innen kommentierten die Aufgaben hinsichtlich Verständlichkeit und Schwierigkeit.

In qualitativer Hinsicht bekamen die Aufgaben insgesamt gute Noten. Aufgrund der quantitativen Befunde musste aber etwa ein Drittel der Items revidiert werden. Zwei Aufgaben wurden ganz ausgeschieden. Drei Aufgaben mussten nachkonstruiert werden, um den unteren Niveaubereich besser abzudecken. Insgesamt besteht das definitive Aufgabenkorpus aus 47 Testaufgaben (davon sind 22 Originale und 25 Übersetzungen).

Die folgende Abbildung zeigt die Verteilung der Aufgaben nach Revision auf die 6. und 9. Klassen und nach den drei Sprachen. Aufgabenbeispiele resp. illustrative Aufgabenserien für alle drei Sprachen sind im Vernehmlassungsdokument zu den Basisstandards abgedruckt.

9. Klasse Englisch (D-CH = W-CH)													
Pet Fish 15'	Stoneheart 15'		Friends 10'	Papilio 10'		C-Test 10'	Websites 10'		Argent 10'	Vélo 10'		Bee 15'	Chat 15'

9. Klasse Französisch (D-CH)							9. Klasse Deutsch (W-CH)						
Jaël 15'	Moyen Age 15'	Fou 15'	Chat 15'	Popstar 15'	Stoneheart 15'	Stoneheart 15'	Chat 15'	Bücher 15'	MoyenAge 15'	Fussballer 15'	Popstar 15'		
			Zimmer 10'	Jug-zent 10'	Argent 10'	Vélo 10'	Argent 10'	Vélo 10'	Zimmer 10'	Jug-zent 10'			

6. Klasse Französisch (D-CH)						6. Klasse Deutsch (W-CH)									
C-Test 10'	Ki'rechte 10'	Zimmer 10'	Friends 10'	Argent 10'	Vélo 10'	Argent 10'	Vélo 10'	Zimmer 10'	Friends 10'	Ki'rechte 10'	C-Test 10'				
Anything 10'	Ballo 10'	Steck-br 10'	Websites 10'	Winter-sp 10'	Eragon 10'	Fou 15'	Fussballer 15'	Steck-br 10'	Winter-sp 10'	Eragon 10'	Anything 10'	Websites 10'	Ballo 10'	Bee 15'	Fussballer 15'

Abb. 8: Verteilung der Aufgaben zum Leseverstehen in der nationalen Validierungsuntersuchung. - Dunklere Farben: Lesetexte in Originalsprache; hellere Farben: Texte in Übersetzungen. Grün: Ursprung Englisch; Rot: Ursprung Französisch; Blau: Ursprung Deutsch; Gelb: zwei Ursprungssprachen (nachkonstruierte Aufgaben). Gleiche Aufgabennamen bedeuten strukturgleiche Aufgaben in den jeweils anderen Sprachen. 10' resp. 15' sind die Bearbeitungszeiten der Aufgaben in Minuten.

Zur genaueren Beschreibung der Items wurden sog. *item maps* erstellt. *Item maps* bestehen aus Merkmalen, von denen man annimmt, dass sie die Schwierigkeit der Items beeinflussen. Die spezifischen Ausprägungen dieser Merkmale können kodiert und mittels statistischer Verfahren mit der empirisch ermittelten Schwierigkeit der Items in Beziehung gesetzt werden. Solche Verfahren sollen dazu beitragen, die empirische Schwierigkeit von Items besser interpretieren (und, darauf aufbauend, in Zukunft besser voraussagen) zu können.

In der Untersuchung zum Leseverstehen wurden insgesamt 11 potentiell schwierigkeitsbestimmende Merkmale ausgewählt<sup>28</sup>, darunter die drei fokussierten Komponenten des Konstrukts *fremdsprachliche Lesekompetenz*, das Antwortformat der Testaufgaben sowie sieben weitere Merkmale. Die weiteren Merkmale charakterisieren die Sprache der Lesetexte und besonders der Textpassagen genauer, auf die sich die Items beziehen. Damit werden auch ausgewählte linguisti-

<sup>28</sup> Grundlagen für die Auswahl waren v.a. der oben zitierten „Dutch CEFR Grid Reading/Listening“, die PISA-Studien sowie die Fachliteratur, darunter insbesondere auch die sehr gute Übersicht von Grotjahn (2000).

sche Kompetenzen genauer beleuchtet, auf die zugegriffen werden muss, um die Aufgabenstellungen zu bearbeiten und die Texte zu.

Alle 11 Merkmale wurden operational definiert. Die Definitionen wurden in einem Kodier-Leitfaden festgehalten. Jedes einzelne Item der total 47 Testaufgaben wurde nach den 11 Merkmalen kodiert.

Abbildung 9 zeigt einen kleinen Ausschnitt aus den erstellten *item maps*; Abb. 10 und 11 illustrieren einen Ausschnitt aus dem Kodier-Leitfaden (Abb. 10 synoptisch und Abb. 11 exemplarisch am Beispiel des Merkmals „Leseprozess“ bzw. „Leseart“); im nachfolgenden Text wird knapp resümiert, auf welche Art und Weise die Merkmalsausprägungen festgestellt wurden (Expertenschätzung einerseits und quantitative Erfassung andererseits).

Item	Aufgabe	Handlungsaspekt	Antwortformat	Texttyp	Leseprozess	Anzahl Wörter	lexikalische Dichte	Lesbarkeit	Wortfrequenz	Textmenge	Grammatik	Lexik
FL311f	Steckbriefe_dF31-1	1	3	1	2	2	1	1	3	2	1	2
FL273f	Jugendzentrum_dF27-3	1	2	1	2	3	1	1	3	1	1	1
FL244f	Taschengeld_dF24-4	2	1	3	1	1	1	1	4	2	2	1

Abb. 9: Ausschnitt aus den *item maps*. Ausgewählt sind drei Items von Testaufgaben für Französisch als Fremdsprache („Steckbriefe“, „Jugendzentrum“ und „Taschengeld“). Diese Aufgaben sind im Vernehmlassungsdokument zu den Basisstandards ausgewiesen und kurz charakterisiert ausgewiesen. Die hervorgehobenen Merkmale (grau hinterlegte Zellen) werden in Abb. 10 etwas genauer dargestellt.

Merkmal	Umschreibung	Merkmalsausprägungen/ Kodes	Charakter der Kodes und Datenniveau	Besonderes; Interpretation der Kodes
<b>Format</b>	Antwortformat der Aufgabe	1= MCQ (Multiple Choice Question) 2= MM (Multiple Matching) 3= offen/kurz (Kurzantwort) 4=Kombi (Kombination von zwei Formaten, z.B. 2 und 3)	Expertenurteil <i>nominal</i>	Gleiche Kodes für alle Items einer Aufgabe
<b>Texttyp</b>	Itembezogene funktionale Haupt-Ausrichtung des Lesetextes	1= deskriptiv und/oder expositorisch 2= narrativ 3= argumentativ 4= instruktiv 5= diskontinuierlich 7= kein (genau spezifizierbarer) Text zur richtigen Lösung des Items	Expertenurteil <i>nominal</i>	Kodes können für Items innerhalb einer Aufgabe verschieden sein
<b>Prozess</b>	Itembezogener dominanter Leseprozess (bzw. Leseart)	1= sorgfältig-genaueres Lesen (careful reading) 2= erkundendes und selektives Lesen (expeditious reading) 7= kein (genau spezifizierbarer) Text zur richtigen Lösung des Items	Expertenurteil <i>nominal</i>	Kodes können für Items innerhalb einer Aufgabe verschieden sein
<b>Lesbarkeit</b> spezifiziert für DaF, FLE und EFL	Lesbarkeitsformeln: Gunning-Fog, Readability, Coleman-Liau	1= erstes Quartil 2= zweites Quartil 3= drittes Quartil 4= viertes Quartil	Quantitativ; Summe der Durchschnittswerte von drei Lesbarkeitsformeln; Kategorisierung nach Quartilen <i>ordinal</i>	Gleiche Werte für alle Items einer Aufgabe  The higher, the harder to read
<b>Lexik</b> spezifiziert für DaF, FLE und EFL	Itemrelevanter Anspruch der Lexik des Textes	1= einfach(er)e Lexik 2= komplex(er)e Lexik 7= kein (genau spezifizierbarer) Text zur richtigen Lösung des Items	Expertenurteil <i>ordinal</i>	Werte können für Items innerhalb einer Aufgabe verschieden sein  Je höher der Kode, desto höher der Anspruch, desto schwieriger

Abb. 10: Synoptische Darstellung eines Ausschnitts aus dem Kodier-Leitfaden (Kurz-Charakteristik der Merkmale „Antwortformat“, „Texttyp“, „Leseprozess“, „Lesbarkeit“ und „Lexik“; die ersten drei Merkmale wurden sprachübergreifend kodiert, die letzten beiden sprachspezifisch); Abb. 11 zeigt am Beispiel des Merkmals „Leseprozess“, wie diese Itemmerkmale operationalisiert wurden.

<b>Merkmal Leseprozess (Leseart)</b> Kodes und Operationalisierungen	<b>Identifizierung des Merkmals Leseprozess (Leseart)</b>
<p><b>1= Sorgfältig-genaues Lesen (careful reading)</b></p> <ul style="list-style-type: none"> <li>- Ziel ist genaues Verstehen und Verstehen im Detail.</li> <li>- Typische Beispiele sind: das Identifizieren, Verstehen und Vergleichen von Fakten, Definitionen oder Meinungen; das Verstehen von Hauptgedanken (läuft in der Regel über das Erfassen der Textorganisation und die Gewichtung von Informationen) und/oder von wichtigen Details (z. B. stützenden Argumenten); das Erschließen von impliziten Informationen aufgrund von expliziten Informationen (propositionale Inferenzen).</li> <li>- Das Lesen kann auf Textelemente, auf eine Textstelle oder auf eine grössere Textmenge gerichtet sein.</li> </ul> <p><i>Achtung: Sorgfältig-genaues Lesen kann selektives Lesen (vgl. Kode 2) voraussetzen (auch innerhalb von 1-Punkte-Items). Wenn das der Fall ist, wird dennoch der Kode 1 vergeben.</i></p> <p><b>2= erkundendes und selektives Lesen (expeditious reading)</b></p> <ul style="list-style-type: none"> <li>- Ziel ist eher rasches, z. T. grobes Verstehen.</li> </ul>	<p>Dieses Merkmal erfasst den Leseprozess (die Leseart), den ein kompetenter Leser beim Lesen des Textes / der Texte wahrscheinlich anwenden würde, um ein Item richtig zu lösen.</p> <p>Leitfrage: Was ist der jeweils erfolgversprechendste Weg / die beste Art, den itemrelevanten Text zu finden und zu verstehen?</p> <p><b>Konkretisierung der Leitfrage für 1 Pkt.:</b> Welcher Lese-Prozess (welche Leseart) ist von der Konstruktion her intendiert und folglich „angemessen“,</p> <ul style="list-style-type: none"> <li>- um 1-Punkte-Items richtig lösen zu können (d. h. den einen Punkt zu bekommen)?</li> <li>- um 2-Punkte-Items teilweise richtig lösen zu können (d. h. einen von zwei möglichen Punkten zu bekommen)?</li> </ul> <p>Wo zutreffend: <b>Konkretisierung der Leitfrage für 2. Pkt.:</b> Welcher Lese-Prozess (welche Leseart) ist von der</p>

Abb. 11: Operationalisierung des Merkmals „Leseprozess“ oder „Leseart“ im Kodier-Leitfaden (Auszug)

Die Arbeiten zur Kodierung der Items wurden parallel zur Aufgabenentwicklung an die Hand genommen, erstreckten sich aber über einen viel längeren Zeitraum.

Die unterschiedenen Leseprozesse, Handlungsaspekte und Texttypen sowie das Aufgabenformat wurden durch Expertenurteil eingeschätzt (Konsensfindung in der Arbeitsgruppe).

Vier der sieben weiteren Merkmale wurden quantitativ erfasst. Für jeden Lesetext einzeln wurden berechnet: die Anzahl Wörter, die lexikalische Dichte („lexical density“, eine Art von *type-token ratio*), die Lesbarkeit (Durchschnittswerte von drei Lesbarkeitsformeln: Gunning-Fog, Readability, Coleman-Liau) und die Wortfrequenz (prozentualer Anteil der Wörter eines Textes an den häufigsten 1000 Wörtern gemäss Frequenzwörterbüchern)<sup>29</sup>. Diese Kennzahlen können je nach Sprachversion der Lesetexte (Originale vs. Übersetzungen) unterschiedlich sein. Deshalb war die Bezugsgrundlage für alle diese Berechnungen nicht das Total aller Texte des Aufgabenkorpus, sondern das Total der in einer Sprache eingesetzten Lesetexte, also zum Beispiel das Total der Wörter aller Lesetexte in Deutsch als Fremdsprache (DaF). Berechnet wurde dann, ob z. B. ein einzelner DaF-Text relativ zum Total aller DaF-Texte viele oder wenige Wörter umfasst.

Die restlichen drei Merkmale wurden durch Expertenurteile erfasst (in der Regel Konsensfindung in der Arbeitsgruppe). Eingeschätzt wurden: die Textmenge, die grammatische Schwierigkeit und die lexikalische Schwierigkeit. Diese Einschätzungen beziehen sich aber nicht auf den gesamten Lesetext (wie bei den quantitativen Merkmalen), sondern spezifischer auf diejenigen Textstellen, die für die Beantwortung der Items von den Experten als relevant erachtet wurden.

Mit den kodierten Items wurde eine Grundlage für die Erklärung der Itemschwierigkeit und damit auch für eine (Teil-)Validierung des Konstrukts geschaffen.

### 3.5.3 Datenerhebung

In der nationalen Validierungsuntersuchung wurden die total 47 Aufgaben zum Leseverstehen auf 32 verschiedene Testhefte verteilt. Jedes Heft bestand aus zwei bis drei Aufgaben und wurde von 150 bis 250 Schüler/innen bearbeitet (für genauere Angaben zur Datenerhebung vgl. die Tabellen

<sup>29</sup> Quellen für diese Berechnung waren: für die Anzahl Wörter, die lexikalische Dichte und die Lesbarkeit: die Internetressource „topicalizer“ (<http://www.topicalizer.com>) sowie, zur Kontrolle, die Software LingoFox (<http://www.lingofox.de>); für die Wortfrequenz: die Angaben aus dem Wortschatzprojekt an der Universität Leipzig (<http://wortschatz.uni-leipzig.de/html/sitemap.html>) sowie, zu Vergleichszwecken für Englisch und Französisch, die kanadische Internetressource „lxtutor“ (<http://www.lxtutor.ca>).

in 3.2). Eine Aufgabe umfasste durchschnittlich fünf Items. Die vorgesehenen Bearbeitungszeiten pro Aufgabe betragen 10 resp. 15 Minuten. Alle Aufgaben mussten ohne Hilfsmittel gelöst werden. Zehn Aufgaben wurden sowohl in der 6. als auch in der 9. Klasse eingesetzt, darunter z. B. die Aufgaben „Vélo“, „Argent“ und „Zimmer“ (vgl. Abb. 8; Abb. 8 zeigt auch die Überlappungen zwischen den Sprachen; z. B. wurden die Aufgaben „Vélo“ und „Argent“, die für Französisch als Fremdsprache konstruiert wurden, als strukturgleiche Aufgabenübersetzungen auch für Deutsch und Englisch als Fremdsprachen eingesetzt).

#### **3.5.4 Korrektur**

Für die Korrektur und die elektronische Erfassung der bearbeiteten Hefte wurden vier Studierende angestellt. Jede Person übernahm die Verantwortung für acht Testhefte, was einem Volumen von über 1000 einzelnen Schülerheften entspricht. Die Studierenden benötigten für diese Arbeit zwischen 35 und 50 Stunden.

Korrigiert wurden die Testhefte mit einem Lösungsschlüssel (Beispiele solcher Schlüssel sind zusammen mit den Testaufgaben zur Illustration der Standards im Vernehmlassungsdokument abgedruckt.) Die Korrektur gestaltete sich je nach Antwortformat einer Testaufgabe unterschiedlich. Bei geschlossenen Formaten wurden die Schülerantworten nicht bewertet, sondern „registriert“, d. h. erfasst wurde die Position einer Antwort (z. B. wurde die Ziffer „1-0-0-0-0“ notiert, wenn nur das erste Feld in einer Tabelle mit fünf Spalten angekreuzt war), unabhängig davon, ob diese Antwort richtig oder falsch war. Mit diesem Verfahren bleibt die Korrektur besser nachvollziehbar und das Antwortverhalten der Schüler/innen kann auch noch nachträglich genauer analysiert werden (Möglichkeit von Distraktoren-Analysen). Bei offenen Fragen wurden die Schülerantworten gemäss einem Richtig-falsch-Schema bewertet.

Erfasst wurden die registrierten und bewerteten Schülerantworten von der gleichen Person mit Hilfe von Excel-Masken. Von einer systematischen Doppel-Kodierung wurde u. a. aus finanziellen Gründen abgesehen. Als qualitätssichernde Massnahme wurden die Eingabemöglichkeiten in diese Masken aufgaben- und itemspezifisch beschränkt. So konnte die Gefahr von Erfassungsfehlern beträchtlich reduziert werden.

Um insgesamt eine möglichst zuverlässige Korrektur der Daten zu erreichen, wurden verschiedene Massnahmen getroffen: Schulung der Korrektorinnen an einem Einführungstag, Einrichtung einer Hotline zur Begleitung der individuellen Korrekturarbeit, stichprobenartige Kontrollen durch einen Zweitkorrektor besonders bei der Datenabgabe. Zur weiteren Absicherung wurde am Beispiel von Testheften für Englisch ein kleines Experiment durchgeführt. Mit dem Experiment wurde exemplarisch überprüft, ob die Korrektorinnen die Kriterien und Codes des Lösungsschlüssels einheitlich anwenden (Überprüfung der Intercoder-Reliabilität). Die Resultate dieses Experiments waren sehr positiv.<sup>30</sup>

Insgesamt kann man davon ausgehen, dass die Daten zum Leseverstehen zuverlässig korrigiert und erfasst wurden.

#### **3.5.5 Ergebnisse und Analyse der Ergebnisse**

An dieser Stelle werden ausgewählte Resultate berichtet. Die Resultate beziehen sich a) auf die Testaufgaben und -items sowie auf die Skalenkonstruktion, b) auf die Frage der Dimensionalität des Konstrukts *fremdsprachliche Lesekompetenz* und c) auf die Verteilung der fremdsprachlichen Lesekompetenz in der Population. Unter b) werden auch Konsequenzen von Resultaten für Aspekte von Testspezifikationen angesprochen.

a) Insgesamt konnten 198 Items zum Leseverstehen erfolgreich skaliert werden, nämlich 86 für Deutsch, 82 für Französisch und 30 für Englisch. Die 198 Items verteilen sich auf 47 Aufgaben. Das sind etwa vier Items pro Aufgabe. Ursprünglich umfasste eine Aufgabe durchschnittlich fünf Items. Der Itemverlust bei der Skalierung betrug damit insgesamt etwa 20%. Die statistischen

---

<sup>30</sup> Zwei Korrektorinnen korrigierten unabhängig voneinander 100 Items (zehn Schülerhefte mit zwei Aufgaben à je fünf Items). Es handelte sich um Aufgaben mit offenen Formaten. Es zeigte sich eine hohe Übereinstimmung der Bewertungen: Kappa 0.9369, Standardfehler 0.0206.

Auswahlkriterien bei der Skalenkonstruktion waren: Modell-Konformität oder –„Passung“ der Items (Raschmodell; Infit < 1.2) und hinreichende Trennschärfe der Items (Diskrimination > 0.3) sowie, bei einem Teil der Aufgaben, gleiches Funktionieren der Items in der Deutschschweiz und in der Westschweiz einerseits und in den 6. und 9. Klassen andererseits. Mit diesen Kriterien wurden die folgenden Skalen konstruiert: zuerst je eine Skala für Deutsch, Französisch und Englisch, dann eine gemeinsame Skala für Deutsch und Französisch (aber ohne Englisch) und schliesslich die definitive Skala, die alle drei Sprachen umfasst. Bei der Skalenkonstruktion zeigte sich, dass viele Englisch-Items in der Deutschschweiz und in der Westschweiz jeweils anders funktionierten und deshalb ausgeschieden werden mussten. Im Unterschied dazu funktionierten viele Items in übersetzten Aufgaben für Deutsch (in der Westschweiz) und Französisch (in der Deutschschweiz) in gleicher Weise, was zu einem geringeren Itemverlust in diesen beiden Sprachen führte. Überraschend ist nicht nur der grosse Itemverlust bei Englisch (gleiche Aufgaben in den beiden Landesteilen!), sondern auch die weit gehende Parallelität von Items für Französisch und Deutsch („nur“ strukturgleiche Aufgabenübersetzungen). Diese Resultate sind schwer zu erklären; im Fall von Englisch könnten sie mit den unterschiedlichen Erstsprachen der Schüler/innen in der Deutschschweiz und in der Westschweiz und auch mit dem unterschiedlichen Stellenwert des Englischen in den beiden Landesteilen zu tun haben; aber das sind vorläufige Hypothesen, denen genauer nachgegangen werden sollte.

Die Itemskalen wurden durch Expertenurteil auf die Niveaus des *Europäischen Referenzrahmens* resp. auf die in *IEF/lingualevel* differenzierten Feinniveaus bezogen (Konsensfindung von drei in Bezug auf den *Referenzrahmen* sehr erfahrenen Personen, die auch an verschiedenen Benchmarking-Konferenzen teilgenommen haben). Ausgegangen wurde dabei von Aufgaben und Items, deren empirische Schwierigkeit sich (weitgehend) mit der geschätzten Schwierigkeit deckt. Diese Items übernahmen nach der Zuordnung zur Referenzskala die Funktion von Ankeritems. Die Zuordnung der weiteren Items geschah weitgehend rechnerisch, wobei die *cut-offs* zwischen den Niveaus unter Berücksichtigung in etwa gleicher Abstände auch mit Augenmass gesetzt wurden.

Auf diesem Weg ist es letztlich gelungen, eine einheitliche, d. h. für alle drei Sprachen und für beide untersuchten Schulstufen gemeinsam geltende Skala zu konstruieren, auf der jedes Item einem Feinniveau der europäischen Referenzskala zugeordnet ist. Ausserdem wurden auch die HarmoS-Niveaus für die Klassen 8 und 11 auf diese Skala abgebildet. Die Skala umfasst Items für die europäischen Niveaus A1.1 bis B2 bzw. für die HarmoS-Niveaus I<sub>8</sub> bis V<sub>11</sub>, wobei das unterste Niveau nur mit ganz wenigen Items belegt ist.

Um die empirische Schwierigkeit der Items, die teilweise plausibel, teilweise aber auch schwer nachvollziehbar ist, besser zu verstehen, wurden einerseits in der Arbeitsgruppe Thesen zur Divergenz von empirischer und geschätzter Item-Schwierigkeit diskutiert<sup>31</sup>, andererseits wurden die 11 potentiell schwierigkeitsbestimmenden Merkmale, nach denen jedes Item kodiert wurde (vgl. 3.5.2, „item maps“) zusammen mit den empirischen Skalen verschiedenen statistischen Analysen unterzogen. Eingesetzt wurden Clustering-Verfahren, „Partitioning“ und die Multiple Korrespondenzanalyse. Leitfrage dabei war, welchen Anteil die unterschiedenen Merkmale (z. B. Textmenge, Wortfrequenz, grammatische Schwierigkeit) einzeln und zusammen an den empirisch ermittelten Itemschwierigkeiten haben. Die Ergebnisse dieser Analysen deuten auf eine gewisse Prominenz wortschatzbezogener Merkmale hin. Sie zeigen aber auch, dass alle einbezogenen Merkmale zu-

---

<sup>31</sup> Die Divergenzen gehen in beide Richtungen: Es gibt Items, die empirisch schwieriger sind als geschätzt, und solche, bei denen das Umgekehrte zutrifft. Im ersten Fall beziehen sich einige Thesen auf die Antwortformate im weiteren Sinn (z. B. erhöhte Schwierigkeit durch Anzahl Antwortalternativen in Multiple-Matching-Aufgaben oder Art der Distraktoren in Multiple-Choice-Aufgaben). Zu den intervenierenden Faktoren bei Items, die empirisch leichter sind als geschätzt, könnten u. a. Aspekte allgemeiner Kompetenzen ebenso gehören wie sprachliche Merkmale der Texte und der Items selber.

sammen nur etwas weniger als die Hälfte der Itemschwierigkeit „erklären“ können ( $R^2 = 0.46^{32}$ ), wobei dieses Ergebnis durchaus im Rahmen der Resultate von vergleichbaren Projekten ist<sup>33</sup>.

b) Bei den Analysen zur Dimensionalität des Konstrukts *fremdsprachliche Lesekompetenz* geht es um die Frage, ob die unterschiedenen Kompetenzaspekte (Leseprozesse, Handlungsaspekte und Texttypen) auch empirisch unterschieden werden können. Empirische Unterscheidbarkeit würde dafür sprechen, dass die Aspekte bei verschiedenen Schüler/innen unterschiedlich funktionieren. Aufgrund der durchgeführten Analysen für Französisch und Deutsch (Berechnung latenter Korrelationen direkt bei der Skalierung der Items mit ConQuest) lässt sich diese Frage nur vorläufig beantworten. Am klarsten sind die Befunde für die Leseprozesse: *Sorgfältig-genaues Lesen* einerseits sowie *erkundendes und selektives Lesen* andererseits könnten tatsächlich zwei verschiedene Dimensionen sein (latente Korrelation von 0.72<sup>34</sup>). Bei den Handlungsaspekten gibt es zwar Indizien, die auf eine gewisse Selbstständigkeit des Aspekts (*Sich*) in *Beziehung setzen/Evaluieren* hindeuten, aber dieser Aspekt ist nur durch wenige Items vertreten. Kaum unterscheidbar als Dimensionen scheinen dagegen die Aspekte *Informationen entnehmen* und *Interpretieren* zu sein, wobei dieser Befund auch mit Operationalisierungsproblemen zu tun haben könnte (z. T. Abgrenzungsprobleme zwischen den Aspekten *Informationsentnahme* und *Interpretation*)<sup>35</sup>. Bei den Texttypen schliesslich ergibt sich kein einheitliches Bild<sup>36</sup>. – Bilanzierend lässt sich dazu festhalten, dass es einige empirische Anhaltspunkte dafür gibt, dass *sorgfältig-genaues Lesen* einerseits sowie *erkundendes und selektives Lesen* andererseits bei verschiedenen Schüler/innen unterschiedlich funktionieren (vgl. dazu auch Sigott 2007). Ein guter genauer Leser könnte also nicht unbedingt auch Aufgaben zum erkundenden und selektiven Lesen gut lösen – und umgekehrt. Für entsprechende Aussagen zu den Handlungsaspekten und den Texttypen sind die empirischen Indizien dagegen zu wenig deutlich und/oder zu heterogen.

Wie erwartet scheint auch das Antwortformat der Testaufgaben einen grösseren Einfluss auf das Konstrukt zu haben. Die Resultate der statistischen Analysen deuten darauf hin, dass die drei Formate *Multiple-Choice*, *Multiple-Matching* und *Kurzantwort* empirisch unterscheidbare Dimensionen sein könnten.<sup>37</sup> Falls dieser Befund in weiteren Untersuchungen bestätigt würde, müsste man davon ausgehen, dass eine Schülerin, die zum Beispiel gut mit Multiple-Choice-Aufgaben umgehen kann, nicht unbedingt auch bei Multiple-Matching-Aufgaben oder bei Aufgaben mit Kurz-

---

<sup>32</sup> Dieser Wert wurde mit dem so genannten *partitioning* (und der Statistiksoftware JMP) berechnet, einem mehrstufigen, rekursiven Verfahren, das im Unterschied zum *Bottom-up*-Ansatz des hierarchischen Clustering *top-down* gerichtet ist und bei dem nach den Kriterien gesucht wird, die eine optimale Gruppierung der Daten leisten. Auch dieses Verfahren zeigt die Wichtigkeit wortschatzbezogener Kriterien, aber auch der Antwortformate.

<sup>33</sup> Beispielsweise erklärt das Regressionsmodell im DESI-Teilprojekt zum Leseverstehen in Englisch als Fremdsprache mit  $R^2 = 0.45$  knapp die Hälfte der beobachteten Unterschiede zwischen den Itemschwierigkeiten („multiple lineare Regressionsanalyse“; Nold/Rossa 2006, 205). Allerdings sind die Resultate von DESI und HarmoS nur bedingt miteinander vergleichbar, weil es zwischen diesen Projekten eine Reihe von Unterschieden konzeptueller und methodischer Art gibt.

<sup>34</sup> Latente Korrelationen sind messfehlerbereinigt und deshalb als ‚oberer‘ Wert anzusehen.

<sup>35</sup> Korrelation von 0.78 zwischen (*Sich*) in *Beziehung setzen/Evaluieren* und *Informationen entnehmen*; Korrelation von 0.77 zwischen (*Sich*) in *Beziehung setzen/Evaluieren* und *Interpretieren*. Demgegenüber ist die Korrelation zwischen *Informationen entnehmen* und *Interpretieren* höher (0.87), was gegen die Unterscheidbarkeit dieser beiden Aspekte spricht.

<sup>36</sup> Zwar gibt es gewisse Hinweise darauf, dass *deskriptive* (inklusive *expositorische*) und *instruktive* Texttypen einerseits sowie *narrative* und *argumentative* Texttypen andererseits die Itemschwierigkeit in ähnlicher Weise beeinflussen (Korrelationen von 0.80 und höher innerhalb der beiden ‚Gruppen‘ von Texttypen vs. tiefere Korrelationen zwischen den ‚Gruppen‘, z. B. 0.66. zwischen *deskriptiven* und *narrativen* Texttypen). Es zeigt sich aber z. B. auch eine relativ hohe Korrelation von 0.81 zwischen diskontinuierlichen Texten und deskriptiven Texttypen. Insgesamt lässt sich aufgrund dieser Analysen eher nicht von Texttypen als Dimensionen des Lesens sprechen.

<sup>37</sup> Die Werte für die latenten Korrelationen sind, zwischen *Multiple-Choice* und *Multiple Matching*: 0.67, zwischen *Multiple-Choice* und *Kurzantwort*: 0.62 und zwischen *Multiple-Choice* und *kombinierten Formaten* 0.73. Die Korrelationen zwischen den drei ‚Einzel‘-Formaten (MC, MM, Kurzantwort) und den kombinierten Formaten sind besonders schwer zu interpretieren, weil unter den kombinierten Formaten verschiedene Formats-Kombinationen subsumiert sind.

antworten gute Leistungen zeigt. Es ist allerdings diskutabel, ob es sich bei Aufgabenformaten ausschliesslich um Faktoren handelt, welche die Messung der Kompetenzen in Bezug auf das Konstrukt stören, denn beispielsweise könnte das Multiple-Matching-Format besser geeignet sein, suchendes und vergleichendes Lesen anzustossen als das unspezifische Multiple-Choice-Format<sup>38</sup>. Jedenfalls könnte es im Sinne der Testfairness gerade dann angezeigt sein, in einem Test verschiedene Formate zu verwenden, wenn Formate einen leistungsdifferenzierenden Einfluss haben, denn durch verschiedene Formate würden Testteilnehmende, die mit einem bestimmten Antwort-Format Probleme haben, weniger stark benachteiligt, als wenn nur *ein* Format verwendet würde. Es sei denn, es liesse sich zeigen, dass bestimmte Formate anderen schlicht überlegen sind, weil sie den Zugang zu den im Konstrukt fokussierten Kompetenzen weniger verstellen. – Festzuhalten ist, dass das Antwortformat die Leistungen beim Lesen in erheblichem Ausmass beeinflussen kann. Dies sollte in künftigen Testkonzeptionen, besonders auch im Rahmen des geplanten Bildungsmonitorings, aber z. B. auch bei der Revision oder Neuentwicklung von Zertifikaten, noch stärker beachtet werden und möglichst auch Gegenstand von einschlägigen Voruntersuchungen sein.

Einschränkend für alle Befunde zur Dimensionalität des Leseverstehens gilt, dass die diesbezüglichen Analysen im Rahmen von HarmoS nicht weit genug verfolgt werden konnten (v.a. wurde nicht nach Alter und Sprache der Zielgruppen differenziert).

c) In diesem Abschnitt wird zuerst ein tabellarischer Überblick über die Ergebnisse gegeben, welche die Sechst- und Neuntklässler/innen in den Lesetests erreicht haben. Anschliessend werden die Ergebnisse kurz erläutert. Bei der Darstellung der Ergebnisse wird für alle drei Fremdsprachen eine einheitliche Skala verwendet. Die Punktwerte stammen aus der Skalierung der Testdaten. Der Skalenmittelwert von 500 Punkten entspricht der mittleren Leistung all derjenigen Schüler/innen, die Deutsch- oder Französisch-Leseaufgaben bearbeitet haben. Die Standardabweichung der entsprechenden Verteilung entspricht 100 Punkten. Die empirische Englischskala wurde nachträglich an diese Skala angepasst. Im Zusammenhang mit dem Wert für den Basisstandard wird in den Tabellen auch der Bezug zur europäischen Referenzskala angegeben.

---

<sup>38</sup> Allerdings bleibt in quantitativen Untersuchungen grundsätzlich offen, ob, was von der Aufgabenstellung z. B. als erkundendes Lesen intendiert ist, von den Testteilnehmenden auch intentionsgemäss realisiert wird. Näheren Aufschluss zu solchen Fragen könnten qualitative Verfahren unter Einschluss von Protokollen des lauten Denkens bringen.

<b>Lesen</b>	
<b>Deutschschweiz</b>	<b>Französisch</b>
<b>6. Klasse</b>	
<b>Basisstandard 1. FS (A2.1)</b>	<b>465</b>
<b>Basisstandard 2. FS (A1.2)</b>	<b>400</b>
<b>6. Kl. alle</b>	<b>480</b>
Der Basisstandard für die 1. FS würde heute von 55% der Sechstklässler/innen erreicht.	
Der Basisstandard für die 2. FS würde heute von 80% der Sechstklässler/innen erreicht.	
6. Kl. Mädchen	492
6. Kl. Knaben	471
<b>9. Klasse</b>	
<b>Basisstandard 1./2. FS (A2.2)</b>	<b>530</b>
<b>9. Kl. alle</b>	<b>523</b>
Der Basisstandard würde heute von 50% der Neuntklässler/innen erreicht.	
9. Kl. Mädchen	537
9. Kl. Knaben	514
9. Kl. Grundansprüche	454
9. Kl. erweiterte Ansprüche	537
9. Kl. hohe Ansprüche	599

<b>Lesen</b>	
<b>Westschweiz</b>	<b>Deutsch</b>
<b>6. Klasse</b>	
<b>Basisstandard 1. FS (A2.1)</b>	<b>465</b>
<b>6. Kl. alle</b>	<b>487</b>
Der Basisstandard würde heute von 60% der Sechstklässler/innen erreicht.	
6. Kl. Mädchen	496
6. Kl. Knaben	480
<b>9. Klasse</b>	
<b>Basisstandard 1./2. FS (A2.2)</b>	<b>530</b>
<b>9. Kl. alle</b>	<b>534</b>
Der Basisstandard würde heute von 50% der Neuntklässler/innen erreicht.	
9. Kl. Mädchen	540
9. Kl. Knaben	525
9. Kl. Grundansprüche	470
9. Kl. ohne Niveautrennung	512
9. Kl. erweiterte Ansprüche	530
9. Kl. hohe Ansprüche	575

<b>Lesen</b>	
<b>Deutschschweiz und Westschweiz</b>	<b>Englisch</b>
<b>9. Klasse Basisstandard 1./2. FS (A2.2)</b>	<b>530</b>
<b>9. Kl. alle</b>	<b>512</b>
9. Kl. Mädchen	527
9. Kl. Knaben	497
<b>DCH: 9. Kl. alle</b>	<b>515</b>
Der Basisstandard würde heute von 45% der Neuntklässler/innen erreicht.	
DCH: 9. Kl. Mädchen	531
DCH: 9. Kl. Knaben	498
DCH: 9. Kl. Grundansprüche	451
DCH: 9. Kl. erweiterte Ansprüche	524
DCH: 9. Kl. hohe Ansprüche	593
<b>WCH: 9. Kl. alle</b>	<b>503</b>
Der Basisstandard würde heute von 40% der Neuntklässler/innen erreicht.	
WCH: 9. Kl. Mädchen	513
WCH: 9. Kl. Knaben	491
WCH: 9. Kl. Grundansprüche	414
WCH: 9. Kl. ohne Niveautrennung	477
WCH: 9. Kl. erweiterte Ansprüche	497
WCH: 9. Kl. hohe Ansprüche	558

Im Folgenden werden einige der Resultate kurz erläutert und gewichtet.<sup>39</sup> Dank der gemeinsamen Skala ist es möglich, die Testergebnisse über die beiden beteiligten Landesteile und alle drei Fremdsprachen hinweg miteinander zu vergleichen.

Ein erster Vergleich zeigt, dass die Lesekompetenz der Westschweizer Schüler/innen in Deutsch etwas besser ausgeprägt ist als diejenige der Deutschschweizer Schüler/innen in Französisch, aber der Unterschied ist nicht signifikant.<sup>40</sup>

Für **Französisch in der Deutschschweiz** ergibt sich folgendes Bild: In den 9. Klassen ist die Lesekompetenz, wie nicht anders zu erwarten, weiter entwickelt als in den 6. Klassen.<sup>41</sup> Geschlechtsspezifische Unterschiede sind vorhanden, und zwar immer in der erwarteten Richtung, also mit Vorteilen für die Schülerinnen, aber signifikant sind diese Unterschiede nur in der 6. Klasse. In der 9. Klasse sind die Unterschiede nicht (mehr) signifikant.<sup>42</sup>

<sup>39</sup> Vergleiche zwischen Teilpopulationen wurden mit t-Tests und Varianzanalysen berechnet. Die Signifikanz der Resultate wird durch „\*“ angezeigt, wobei \*=signifikant ( $p < 0.05$ ), \*\*=sehr signifikant ( $p < 0.01$ ) und \*\*\*=hochsignifikant ( $p < 0.001$ ). Die Zahlenangaben in den Histogrammen entsprechen nicht der Anzahl Schüler/innen, die in der Stichprobe vertreten waren, sondern es handelt sich um Werte, die auf die Schülerpopulation hochgerechnet wurden. Bei der Bestimmung der Signifikanzen wurde die Qualität der Stichprobe in Form eines geschätzten Designeffekts von 2 berücksichtigt.

<sup>40</sup> Mittelwert Westschweiz: 505, Mittelwert Deutschschweiz: 499,  $t = 1.016$ .

<sup>41</sup> Mittelwertsunterschied zwischen 6. und 9. Klasse 42,  $t = 5.593^{***}$

<sup>42</sup> Mittelwertsunterschiede zwischen Schülerinnen und Schülern, 6. Klasse: 21,  $t = 2.323^*$ ; 9. Klasse: 23,  $t = 1.675$

In der folgenden Abbildung ist die Verteilung der Lesekompetenzen in der 9. Klasse nach Schultypen dargestellt:

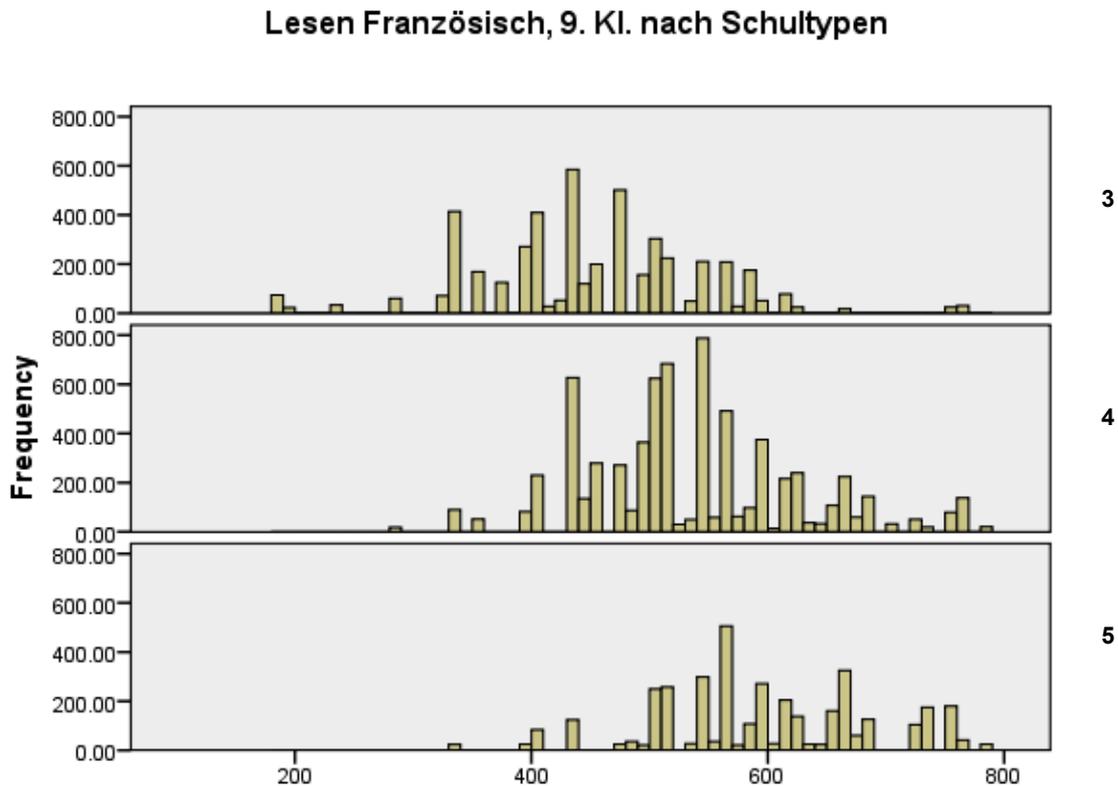


Abb. 12: Lesekompetenzen in Französisch als Fremdsprache in der 9. Klasse (Deutschschweiz) nach Schultypen; 3= Grundansprüche, 4= erweiterte Ansprüche, 5= hohe Ansprüche

Abb. 12 zeigt, dass besonders die Unterschiede zwischen Klassen mit Grundansprüchen einerseits und Klassen mit erweiterten und hohen Ansprüchen andererseits sehr gross sind.<sup>43</sup> Der Mittelwert in den 9. Klassen mit Grundansprüchen liegt tiefer als der Mittelwert der 6. Klassen (454 Punkte gegenüber 480 Punkten).

Die entsprechenden Befunde für **Deutsch als Fremdsprache in der Westschweiz** sehen so aus: Wiederum gibt es den erwarteten Unterschied zwischen 6. und 9. Klassen.<sup>44</sup> Auch die geschlechtsspezifischen Unterschiede sind bei Deutsch in der Westschweiz sehr ähnlich ausgeprägt wie bei Französisch in der Deutschschweiz. Wiederum ist die Lesekompetenz bei den Schüler<sup>innen</sup> höher als bei den Schülern, aber (knapp) signifikant sind diese Unterschiede nur in der 6. Klasse. In der 9. Klasse sind die Unterschiede nicht (mehr) signifikant.<sup>45</sup>

Die folgende Abbildung zeigt wiederum die Verteilung der Lesekompetenzen in der 9. Klasse nach Schultypen:

<sup>43</sup> Die Mittelwertsdifferenz beträgt 105 Punkte,  $t= 7.894^{***}$ ; eine Varianzanalyse unter Einbezug aller drei Klassentypen bestätigt den grossen Unterschied:  $F= 42.114^{***}$ ,  $df=2$

<sup>44</sup> Mittelwertsunterschied zwischen 6. und 9. Klasse 46,  $t=6.670^{***}$

<sup>45</sup> Mittelwertsunterschiede zwischen Schülerinnen und Schülern, 6. Klasse: 16,  $t=1.963^*$ ; 9. Klasse: 15,  $t=1.422$

## Lesen Deutsch, 9. Kl. nach Schultypen

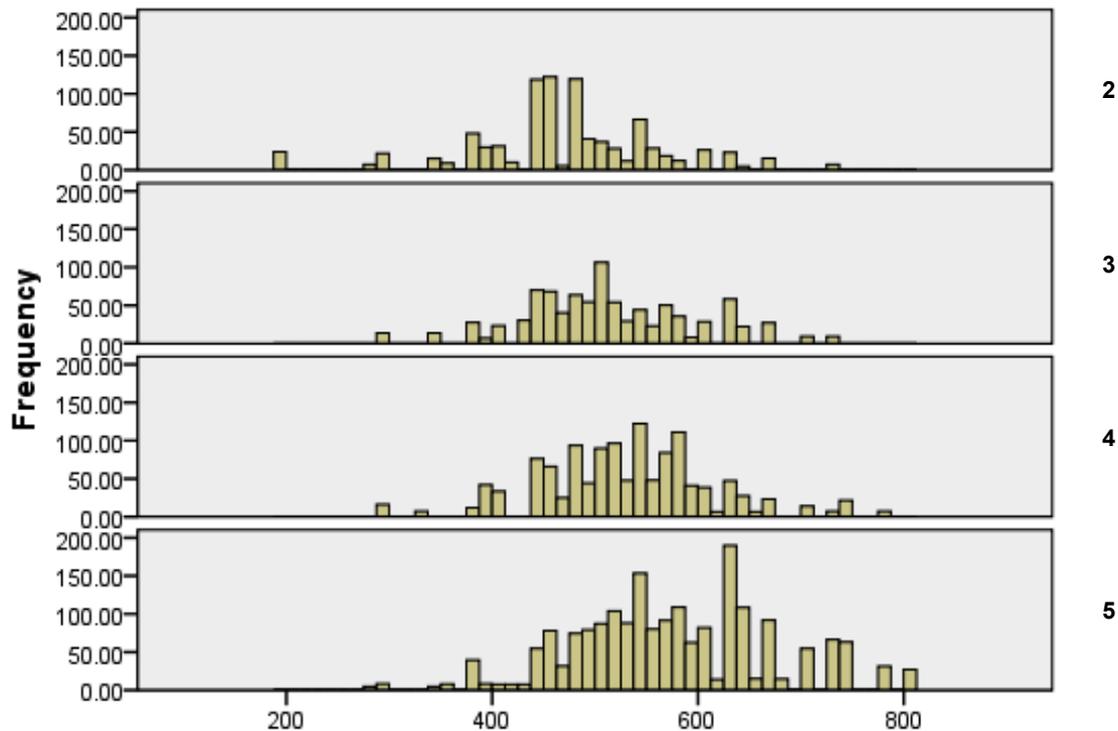


Abb. 13: Lesekompetenzen in Deutsch als Fremdsprache in der neunten Klasse (Westschweiz) nach Schultypen; 2= exigences élémentaires, 3= sans separation, 4= exigences élevées, 5= exigences hautes

Abb. 13 zeigt, dass die Unterschiede zwischen den verschiedenen Schultypen in der Westschweiz differenziert einzuschätzen sind. Signifikant sind die Unterschiede zwischen den beiden unteren Typen („exigences élémentaires“ und „sans separation“) sowie zwischen den beiden oberen Typen („exigences élevées“ und „exigences hautes“). Nicht signifikant sind die Unterschiede dagegen zwischen den beiden mittleren Schultypen, also zwischen integrierten Klassen und Klassen mit erhöhten Ansprüchen („sans separation“ und „exigences élevées“).<sup>46</sup> Auch in der Westschweiz fällt auf, dass die Lesekompetenz in den 6. Klassen im Durchschnitt besser ausgeprägt ist als im untersten Niveau der 9. Klassen (487 Punkte gegenüber 470 Punkten).

In **Englisch als Fremdsprache** erreichen die 9. Klassen insgesamt einen Mittelwert von 512 Punkten. Verglichen mit den Mittelwerten für Deutsch (533 Punkte) und Französisch (523 Punkte) in den 9. Klassen ist das etwas weniger. Dieser Unterschied ist gering, wenn man bedenkt, dass die durchschnittliche Lernzeit für Englisch zum Zeitpunkt der Erhebung viel kürzer war als für die Landessprachen.<sup>47</sup> In der Deutschschweiz liegt der Mittelwert für Englisch leicht höher als in der Westschweiz, aber der Unterschied ist nicht signifikant.<sup>48</sup> Einen geschlechtsspezifischen Unterschied in der Englischkompetenz gibt es in beiden Landesteilen, aber der Unterschied zwischen

<sup>46</sup> Mittelwertsunterschied zwischen „exigences élémentaires“ und „sans separation“: 42,  $t=2.666^{**}$ ; Mittelwertsunterschied zwischen „exigences élevées“ und „exigences hautes“: 45,  $t=3.607^{***}$ ; Mittelwertsunterschied zwischen „sans separation“ und „exigences élevées“ 18,  $t=1.275$ . Insgesamt, d. h. unter Einbezug aller vier Klassentypen, ist der Unterschied hochsignifikant ( $F=20.01^{***}$ ,  $df=3$ ).

<sup>47</sup> Allerdings müssen Vergleiche zwischen Deutsch und Französisch einerseits und Englisch andererseits allgemein vorsichtig interpretiert werden, weil die Englisch-Skala nicht direkt mit statistischen Mitteln an die Deutsch- und Französisch-Skala angebunden, sondern nachträglich durch ein qualitatives Verfahren zugeordnet wurde.

<sup>48</sup> Mittelwertsunterschied 13,  $t=1.177$

Schülerinnen und Schülern ist nur in der Deutschschweiz signifikant.<sup>49</sup> Markant sind im Englischen die Unterschiede nach Schultypen, sowohl in der Deutschschweiz als auch in der Westschweiz. In der Deutschschweiz ist nicht nur der Unterschied zwischen dem unteren Typ (Grundansprüche) und den oberen Typen (erweiterte und hohe Ansprüche) bedeutsam, sondern auch der Unterschied zwischen erweiterten und hohen Ansprüchen.<sup>50</sup> In der Westschweiz ist die Situation für Englisch gleich wie für Deutsch: Signifikant sind die Unterschiede zwischen den beiden unteren Typen („exigences élémentaires“ und „sans separation“) sowie zwischen den beiden oberen Typen („exigences élevées“ und „exigences hautes“). Nicht signifikant sind die Unterschiede dagegen zwischen den beiden mittleren Schultypen („sans separation“ und „exigences élevées“).<sup>51</sup>

Die populationsbezogenen Resultate lassen sich wie folgt bilanzieren: Die Lesekompetenzen in den Fremdsprachen sind in der Deutschschweiz und in der Westschweiz ähnlich verteilt. Sehr grosse Unterschiede zeigen sich, in beiden Landesteilen und sowohl bei den Landessprachen als auch bei Englisch, in Bezug auf die verschiedenen Schultypen. Dabei fällt auf, dass die Lesekompetenz in den 6. Klassen im Durchschnitt besser ausgeprägt ist als im untersten Niveau der 9. Klassen. Dieser Befund ist besonders für die Deutschschweiz sehr bedeutsam, weil dort ein grösserer Anteil der Neuntklässler/innen in Klassen mit Grundansprüchen eingeteilt ist. Die geschlechtsspezifischen Unterschiede sind besonders in den 6. Klassen ausgeprägt, scheinen sich aber bis zur 9. Klasse auszugleichen.

### 3.5.6 Bilanz und Ausblick

Es konnte eine einheitliche Skala für das Leseverstehen in Deutsch, Französisch und Englisch als Fremdsprachen konstruiert werden, auf der die Lesekompetenzen, über welche die Schüler/innen der 6. und 9. Klassen zum Erhebungszeitpunkt verfügten, situiert werden konnten. Die Skala umfasst knapp 200 Items (aus rund 50 Testaufgaben), die alle einem Feinniveau der europäischen Referenzskala zugeordnet sind und die durch Merkmale genauer beschrieben sind, die auf das Kompetenzmodell *Fremdsprachen* und spezifischer auf das Konstrukt *fremdsprachliche Lesekompetenz* bezogen sind. Dieses Material kann für verschiedene Zwecke gebraucht werden, insbesondere auch dazu, die Erwartungen an die Lesekompetenzen künftiger Schüler/innen der Klassen 8 und 11 aussagekräftig, entwicklungsorientiert und verständlich zu illustrieren (vgl. die Illustration des Kompetenzbereichs Leseverstehen und die Aufgabenbeispiele zum Lesen im Vernehmlassungsdokument zu den Basisstandards. Damit wurde ein Hauptziel der Auftragsarbeit erreicht.

Die statistische Analyse der Daten erbrachte einige wichtige Resultate sowohl zum Konstrukt *fremdsprachliche Lesekompetenz* als auch zur Verteilung der Lesekompetenzen in der Population.

Die konstruktbezogenen Analysen können dahingehend interpretiert werden, dass die beiden unterschiedenen (Gruppen von) Leseprozessen oder -arten – sorgfältig-genaueres Lesen versus erkundendes und selektives Lesen – eigenständige Dimensionen des Lesens sind. Demgegenüber müssten die empirischen Indizien bei den unterschiedenen Handlungsaspekten und Texttypen deutlicher ausgeprägt und homogener sein, bevor man von Dimensionen sprechen könnte. Mit den Analysen zur Dimensionalität konnte auch ein Beitrag zur empirischen Validierung des Konstrukts geleistet werden, obwohl sie noch differenziert werden müssten und weiterer Bestätigung bedürften; erst dann könnten z. B. verlässliche didaktische Konsequenzen abgeleitet werden.

Die populationsbezogenen Analysen zeigen besonders auch sehr markante Unterschiede in den Lesekompetenzen in Abhängigkeit vom Schultyp. Wenn – wie es zum Konzept der *Basisstandards* gehört – in Zukunft auch die Schüler/innen in den Klassen mit Grundansprüchen die Basisstan-

---

<sup>49</sup> Mittelwertsunterschied Englisch in der Deutschschweiz: 33,  $t=2.430^*$ ; Mittelwertsunterschied Englisch in der Westschweiz: 21,  $t=1.743$

<sup>50</sup> Mittelwertsunterschied zwischen Grundansprüchen einerseits sowie erweiterten und hohen Ansprüchen andererseits 99,  $t=7.502^{***}$ ; Mittelwertsunterschied zwischen erhöhten und hohen Ansprüchen: 69,  $t=4.684^{***}$

<sup>51</sup> Mittelwertsunterschied zwischen „exigences élémentaires“ und „sans separation“: 63,  $t=3.501^{**}$ ; Mittelwertsunterschied zwischen „exigences élevées“ und „exigences hautes“: 61,  $t=4.378^{***}$ ; Mittelwertsunterschied zwischen „sans separation“ und „exigences élevées“ 19,  $t=1.198$ .

dards erreichen sollen, bedarf es enormer Anstrengungen aller am Sprachenlehren und -lernen Beteiligten.

Weiter haben die Analysen einige Einsichten in das Funktionieren von Tests und die Bedingungen von Testaufgaben ermöglicht. Zu den wichtigeren Befunden in diesem Bereich gehört, dass Antwortformate wie Multiple-Choice, Multiple-Matching oder Kurzantwort einen erheblichen leistungsdifferenzierenden Einfluss haben können. Dieser Befund kann insbesondere auch Anlass zur Überprüfung bestehender und zur Differenzierung künftiger Testanlagen sein. Da zur Feststellung von Lesekompetenzen nicht auf Testaufgaben mit bestimmten Antwortformaten verzichtet werden kann, gehört es zu den Herausforderungen künftiger Tests, den Einfluss der Formate zu erforschen, offen zu legen und so gut wie möglich zu kontrollieren.

Schliesslich haben sich im Verlauf der Untersuchungen zum Leseverstehen eine ganze Reihe von lohnenden Anschlussarbeiten abgezeichnet. Zu den Arbeiten, die sich auf die erhobenen Daten stützen könnten, gehören besonders auch vergleichende Analysen

- von guten und schlechten Leser/innen;
- von Lesekompetenzen von Schüler/innen, die in der Validierungsuntersuchung Lesehefte in zwei Fremdsprachen bearbeitet haben;
- von fremdsprachlichen Lese- und Schreibkompetenzen sowie von Lesekompetenzen und C-Test-Resultaten;
- von Lesekompetenzen in der Fremdsprache und in der lokalen Schulsprache.

Weitere Forschungs-Desiderata wären:

- Elaborierung und Ergänzung des Merkmalkatalogs, mit dem die Testitems kodiert wurden: Vermuten kann man, dass die „Varianzaufklärung“ auf der Basis der unterschiedenen Merkmale höher wäre, wenn zusätzliche bzw. andere Merkmale in den Katalog aufgenommen würden, darunter z. B. das Thema der Texte und die Textorganisation und/oder inhaltsbezogene Verständlichkeitsschätzungen der Lesetexte (z. B. auf der Grundlage des sog. Hamburger Modells).
- Entwicklung und Erprobung von ergänzenden Testaufgaben besonders zu denjenigen Kompetenzaspekten, die im Korpus eher untervertreten sind. Das betrifft z. B. Aufgaben zum Handlungsaspekt (*Sich*) *in Beziehung setzen* oder Aufgaben mit narrativen Texten. Resultate aus entsprechenden Untersuchungen könnten zur besseren Einschätzung vorliegender Indizien zur „Unterteilbarkeit“ des Leseverstehens beitragen.
- Qualitative Untersuchungen zur Frage, wie die Zielgruppen ganz konkret mit Aufgaben zum Leseverstehen umgehen. Solche Studien, wie sie im Rahmen von HarmoS geplant waren, aber nicht durchgeführt werden konnten, könnten (auch, aber nicht nur als Ergänzung zu den quantitativen Analysen zur Dimensionalität des Leseverstehens) genaueren Aufschluss darüber geben, ob Lernende ihr Leseverhalten den Erfordernissen der Situation, der Texte und der Aufgaben anpassen und ob sie das in der Weise tun, wie es von der Aufgabenstellung her intendiert ist (z. B. selegierendes versus genaues Lesen). Geeignete Methoden für solche Studien wären u. a. Formen der Beobachtung und des lauten Denkens.<sup>52</sup>

### 3.5.7 Zusammenfassung

Das fremdsprachliche Leseverstehen wurde im Rahmen von HarmoS als kommunikative Sprachfertigkeit aufgefasst und anlässlich der nationalen Validierungsstudie empirisch untersucht. In dieser Untersuchung sollten die Lesekompetenzen von Schüler/innen in den heutigen 6. und 9. Klassen in Deutsch, Französisch und Englisch als Fremdsprachen erfasst werden. Damit sollte einerseits eine empirische Grundlage für die Setzung und Illustrierung von Basisstandards für das

---

<sup>52</sup> Ein erster Schritt in diese Richtung wurde im Rahmen eines Seminars am Departement für Mehrsprachigkeits- und Fremdsprachenforschung der Universität Freiburg unternommen. In diesem Seminar wurden im Rahmen von kleinen empirischen Projekten Schüler/innen zu ihren Lösungswegen beim Bearbeiten von HarmoS-Testaufgaben zum Leseverstehen befragt. Erste Resultate aus diesen Befragungen weisen darauf hin, dass es durchaus Belege für ‚erwartungskonformes‘ Verhalten gibt (z. B. dann, wenn bei Aufgaben zum selektiven Lesen in der Tat zuerst die Frage bzw. das Item gelesen und dann der Text nach der Antwort durchsucht wird). Es gibt aber auch Anzeichen dafür, dass dieses Verhalten sehr störanfällig ist und von ganz verschiedenen Faktoren beeinflusst werden kann, darunter auf das Antwortformat bezogene Aspekte ebenso wie sprachliche Merkmale der Texte, Items und Distraktoren.

Leseverstehen geschaffen und andererseits ein Beitrag zur fertigkeitsspezifischen Konkretisierung und Validierung des Kompetenzmodells *Fremdsprachen* geleistet werden.

Genauer untersucht wurden ausgewählte Aspekte des Lesens, die für den Lese-Erwerb der Zielgruppen in den Fremdsprachen und für den fremdsprachlichen Fachunterricht im Lesen bedeutsam sind. Wichtige Fragestellungen waren: Gelingt es den Schüler/innen, das Lesen an die Anforderungen von Situationen, Aufgaben und Texten anzupassen, d. h. können sie z. B. *einen* Text sorgfältig und genau und einen andern Text selektiv lesen? Wie gut können die Schüler/innen verschiedenen Texten Informationen entnehmen, wie gut können sie Texte interpretieren und wie gut über Textinhalte reflektieren? Welchen Einfluss haben Textfunktionen bzw. Textsorten sowie qualitative Aspekte von Lesetexten (z. B. Schwierigkeit von Lexik und Grammatik, Wortfrequenz) auf das Leseverstehen der Schüler/innen?

Untersucht wurden diese Fragestellungen anhand eines Korpus von rund 50 kommunikativen Testaufgaben.

Zu den wichtigsten Resultaten gehören: Es konnte eine einheitliche Skala für das Leseverstehen in Deutsch, Französisch und Englisch als Fremdsprachen konstruiert werden. Die Skala besteht aus rund 200 Testitems, die alle den Niveaus des *Europäischen Referenzrahmens* resp. den in *IEF/lingualevel* unterschiedenen Feinniveaus zugeordnet sind. Auf dieser Skala konnten die Lesekompetenzen, über welche die Schüler/innen der 6. und 9. Klassen zum Erhebungszeitpunkt verfügten, situiert werden. Damit wurde ein Hauptziel der Untersuchung erreicht.

Ausserdem konnte ein Beitrag zur empirischen Validierung des Konstrukts *fremdsprachliches Leseverstehen* geleistet werden. Zwar bedürften viele Einzelbefunde in diesem Bereich weiterer Analysen und zusätzlicher synoptischer Interpretationen, aber es hat sich abgezeichnet, dass die beiden Leseprozesse oder -arten – sorgfältig-genaueres Lesen einerseits und erkundendes und selektives Lesen andererseits – empirisch unterscheidbare Dimensionen sind und somit mit Recht als selbstständige Konstruktkomponenten angenommen wurden (bei den *Handlungsaspekten* und *Texttypen* ist dies vermutlich eher nicht der Fall).

Bezüglich der Verteilung der Lesekompetenzen in der Population lässt sich festhalten, dass die Lesekompetenzen der Schüler/innen in der Westschweiz und in der Deutschschweiz in etwa gleich ausgeprägt sind. Sehr grosse Leistungsunterschiede zeigen sich aber in den 9. Klassen in Abhängigkeit vom Schultyp, und zwar in beiden Landesteilen und sowohl bei den Landessprachen als auch bei Englisch.

Schliesslich haben sich Befunde zur leistungsdifferenzierenden Wirkung von Antwortformaten ergeben, die für künftige Test- und Testaufgabenpezifikationen genutzt werden können.

## **3.6 Schreiben**

### **3.6.1 Ausgangspunkt**

Das Schreiben hat heute in den Lehrwerken und in der Unterrichtspraxis der Primarschule als eigener Kompetenzbereich keine grosse Bedeutung. Es wird eher „aufgeschrieben“ (Wortschatz, Übungen) als geschrieben mit dem Ziel eines kommunikativen Handelns. Auf der Sekundarstufe bleibt diese unterstützende Funktion des Schreibens erhalten, aber es werden auch – in der Regel kleinere – Texte mit kommunikativer Funktion geschrieben.

In zukünftigen Unterrichtsszenarien, in denen Elemente der Austauschpädagogik sowie der Einbezug von Sachinhalten (z. B. in CLIL) eine grössere Rolle spielen, hat das kommunikative Schreiben einen festen Platz.

In HarmoS wurde vor diesem Hintergrund entschieden, sowohl in der sechsten als auch in der neunten Klasse den Kompetenzbereich Schreiben zu testen. Ein wichtiger Gesichtspunkt war dabei die Aussicht, mit Hilfe von Schreibaufgaben einiges über die, im (Fremdsprachen-)Unterricht zwar kaum gezielt entwickelte, aber vermutlich doch bis zu einem gewissen Grad vorhandene schriftliche Mitteilungsfähigkeit der Schülerinnen und Schüler erfahren zu können. Es konnte davon ausgegangen werden, dass Kompetenzen im Bereich des handlungsorientierten Schreibens aus dem Schulsprachen-Unterricht oder auch aus Alltagserfahrungen in den fremdsprachlichen Kontext transferiert werden können. Die Schreibentwicklung der grossen Mehrheit der getesteten Schüler/innen (besonders natürlich der Sechstklässler/innen) dürfte allerdings auch in der L1 we-

der unter kognitiven noch unter funktionalen Gesichtspunkten (Textsortenkompetenz!) als abgeschlossen gelten<sup>53</sup>.

Den Hintergrund der handlungsorientierten HarmoS-Tests bildet ein sozio-kognitives Modell des Schreibens, das sich einerseits für sprachlich-kognitive Kompetenzaspekte interessiert und andererseits für die Fähigkeit der Lernenden, Kontextfaktoren, die sich aus Aufgabenstellung und Aufgabensetting ergeben, angemessen in ihr Schreiben einzubeziehen (vgl. Shaw/Weir 2007, 62). Dementsprechend haben alle Aufgaben eine klare Situierung und einen klaren Handlungsfokus, den es zu berücksichtigen galt. Beurteilt wurden sowohl die funktional-inhaltliche Aufgabenerfüllung als auch die sprachlich-kommunikative Gestaltung.

Die schriftliche Kommunikationsfähigkeit der Schülerinnen und Schüler wurde mittels insgesamt neun unterschiedlichen Aufgaben überprüft. Die Aufgaben waren so angelegt, dass sie die weiter oben genannten verschiedenen Diskurstypen berücksichtigten. Die im *Referenzrahmen* an sich angelegte Unterscheidung von Schriftlicher Interaktion und Schriftlicher Produktion wurde, anders als beim Sprechen (s. u.), in HarmoS nicht in den Vordergrund gestellt; in den Aufgaben kommen aber Aufgaben zu beiden Kommunikationsmodi vor.

Die Schreibtests konnten komplementär zu den (rezeptiven) Tests zum Leseverstehen im Rahmen der repräsentativen Untersuchung im Frühjahr 2007 durchgeführt werden, während es bei den Tests zum Sprechen aus praktischen Gründen undenkbar gewesen wäre, eine repräsentative Anzahl von Schüler/innen einzubeziehen.

### 3.6.2 Instrumente (Aufgaben)

Als inhaltliche Grundlage für die Entwicklung von Aufgaben und Beurteilungsinstrumenten – die ja beide als komplementäre Mittel für die Umsetzung eines Konstrukts gesehen werden müssen – dienten insbesondere der *Referenzrahmen* und die bereits früher genannten Quellen aus dessen Umfeld, insbesondere die Kompetenzbeschreibungen, Beurteilungskriterien und Kommunikationsaufgaben aus dem Deutschschweizer Projekt *IEF/lingualevel* sowie Hilfsmittel zur Entwicklung und Beschreibung von Schreibaufgaben, die von ALTE (ALTE o.J. a; b) entwickelt und in das *Manual*-Projekt des Europarates (Council of Europe 2003) eingebracht wurden. Dazu kam Fachliteratur zur Schreibkompetenz und zum Testen und Beurteilen des Schreibens<sup>54</sup>.

Als Planungsgrundlage für die Merkmale von Aufgaben und die dazugehörigen Beurteilungsinstrumente entstanden mehrere Generationen von *construct maps* für den internen Gebrauch. Darin wurden insbesondere Merkmale festgehalten, die aus handlungsorientierten Kompetenzbeschreibungen und sprachressourcenbezogenen Beurteilungsrastern stammten; dazu kamen systematische Gesichtspunkte, welche die Konstruktdeckung sicherstellen sollten, besonders die Repräsentation zentral wichtiger Diskurstypen im Aufgabensatz.

Die folgende Tabelle vermittelt einen Überblick über die eingesetzten Aufgaben. Dabei werden Angaben gemacht zu den Diskurstypen, auf welche die Aufgaben(teile) zielen, sowie zur Verteilung auf die Klassenstufen.

---

<sup>53</sup> Vgl. zur Schreibkompetenz und zum Schreibkonstrukt besonders Baer et al. 1997; Bereiter/Scardamania 1987; Grabe/Kaplan 1996; Harsch 2005; Hayes/Flower 1980 und Shaw/Weir 2007.

<sup>54</sup> Besonders zu erwähnen sind die in der letzten Fussnote genannten Werke; spezifisch zum Testen und Beurteilen: Beck/Klieme (2007); Cushing Weigle 2002; Harsch (2006); McKay (2006); Wolfe-Quintero / Inagaki & Kim (1998).

Diskurstyp	Schreibaufgaben								
	602 Glückwünsche	801 Kurzporträt	805 Termin für einen Besuch Tomek	807 Typischer Schultag	808 Erlebnis mit Haustier	811 Meinung zu Hausaufgaben	906 Informationen zu Schokolade	909 Unvergesslicher Tag	912 Handyverbot
Informieren/ Beschreiben	X	X		X		X			
Auffordern/ Veranlassen			X				X		
Erzählen/ Berichten	X				X			X	
Meinung äussern/ Argumentieren						X	X	X	X
Unterhalten von Beziehung	X		X				X		
<b>Einsatz in Klassen:</b>									
Deutsch, 6. Klasse									
Deutsch, 9. Klasse									
Französisch, 6. Kl.									
Französisch, 9. Kl.									
Englisch, 9. Kl., D-CH und W-CH									

In der Verteilung der Aufgaben auf die Klassen bestehen nur kleine Unterschiede zwischen den Sprachen. Viele Ergebnisse können folglich in vertiefenden Untersuchungen direkt verglichen werden.

Je nach Aufgabe verlangte eine erfolgreiche Bearbeitung die Realisierung von einem oder mehreren der unterschiedenen Diskurstypen. In der Aufgabe 909 „Unvergesslicher Tag“ beispielsweise verlangten die Leitpunkte, dass einerseits über einen unvergesslichen Tag *berichtet* wurde und andererseits *erklärt (begründet, argumentiert)* wurde, warum dieser Tag speziell war.

Die Präsentation der Schreibaufgaben in den Testheften folgt in den Grundzügen dem Aufbau der Aufgaben in allen getesteten Kompetenzbereichen. Konkret sah dies bei den Schreibaufgaben (hier exemplarisch aufgezeigt anhand von Aufgabe 906) wie folgt aus:

1. Allgemeine Arbeitsanweisung mit Hinweis auf die zur Verfügung stehende Zeit. Die 20 Minuten Bearbeitungszeit (30 Minuten im Fall der Aufgabe 912) waren knapp bemessen, deshalb auch der Hinweis, dass sich die Schüler/innen nicht zu sehr vor Fehlern fürchten sollten. Dieser Hinweis wurde erst nach einigem Abwägen eingefügt, weil er falsch interpretiert werden könnte (Fehler sind egal); schliesslich überwog aber das Interesse, die Schülerinnen und Schüler zum Schreiben zu motivieren.

### Was musst du tun?

Dieses Heft enthält eine Schreibaufgabe. Für die gesamte Aufgabe hast du 20 Minuten zur Verfügung.

In dieser Zeit musst du:

- die Aufgabe genau durchlesen;
- einen Text schreiben, den du abgeben kannst.

**Hinweis:** Wichtig ist vor allem, dass du schreibst, was in der Aufgabe verlangt wird, und dass man deinen Text versteht. Fürchte dich also nicht zu sehr davor, Fehler zu machen.

Abb. 14: Allgemeine Arbeitsanweisung zur Aufgabe 906

## 2. Situierung der Aufgabe mit einem Bezug zu einer möglichen, realen Aufgabe:

### Informationen über Schokolade

**Situation:** Du hast dich entschieden, im Französischunterricht einen kleinen Vortrag über Schokolade zu halten. Deshalb möchtest du die bekannte Firma *Cailler* um Informationen bitten. Im Internet hast du eine Seite mit einem Kontaktformular gefunden. Du schreibst in dieses Formular und schickst es an *Cailler*.

Abb. 15: Situierung der Aufgabe 906

3. **Leitpunkte:** Den Schülerinnen und Schülern wurde durch Leitpunkte die inhaltliche Strukturierung des Textes weitgehend abgenommen. Verschiedene Gründe sprechen für diese Engführung: Die kurze Zeit, die zur Verfügung stand, hätte für viele Schülerinnen und Schüler nicht für eine grundlegende Planung und das Verfassen eines Textes ausgereicht, in der dann diese Planung umgesetzt worden wäre. Wenn Inhalte (und Funktionen) fest vorgegeben sind, wird die Vergleichbarkeit der Arbeiten grösser. Vermeidungsstrategien können weniger leicht angewandt werden, sodass Defizite im Bereich der sprachlichen (Kompetenz-)Ressourcen öfter sichtbar werden. Als möglicher Nachteil könnte gesehen werden, dass die Fähigkeit, aufgrund eigener Planung Texte zu schaffen, die einer vorgegebenen allgemeinen Situierung angemessen sind, weniger zur Geltung kommen kann<sup>55</sup>. Weiter ist zu erwarten, dass eher Texte entstehen, die gewissermassen seriell die Leitpunkte abarbeiten. Die vermuteten Vorteile wurden aber höher gewichtet als die potenziellen Nachteile.

**Aufgabe:** Schreibe eine Anfrage an die Schokoladefabrik *Cailler*.

Zu den folgenden Punkten musst du in deiner Anfrage auf Französisch etwas schreiben:

- Wer du bist
- Warum du schreibst (dein Vortrag, dein Interesse am Thema Schokolade)
- Was du für deinen Vortrag über Schokolade wissen oder haben möchtest (erwähne 3 Punkte)
- Wohin dir die Firma *Cailler* Material schicken kann (deine Post-Adresse)
- Dich bedanken.

Abb. 16: Die Leitpunkte zur Aufgabe 906

<sup>55</sup> Das knappe Zeitbudget, die Engführung durch relativ viele Leitpunkte und bis zu einem Grad auch das Verbot, Nachschlagewerke zu benutzen können als Einschränkung der kognitiven Validität (Shaw/Weir 2007, 34ff.) der Schreibe gesehen werden, weil einerseits Planungs- und andererseits Überarbeitungskompetenzen nur eingeschränkt zur Geltung kommen können.

Wie das Beispiel zeigt, werden die Schreibaufgaben nicht in der Zielsprache, sondern ganz in der lokalen Schulsprache präsentiert, dies im Unterschied zu dem, was in den bekannten internationalen Prüfungen üblich ist oder (bisher) in deutschen und österreichischen Tests zur Überprüfung der Bildungsstandards praktiziert wird. Es gibt verschiedene Argumente, die für diese Entscheidung sprechen: 1) Bei solchen handlungsorientierten Testaufgaben spielt die Aufgabenstellungen eine wichtige Rolle, es geht nicht etwa darum, irgendeinen Text zu schreiben, sondern eine Kommunikationsaufgabe mit vorgegebenen Zielen zu erfüllen. Situierung und Aufgabenstellung ersetzen das, was bei authentischen Schreibanslässen oft einfach gegeben und evident ist: den Handlungskontext und die Schreibabsicht. 2) Die Schreibkompetenz soll nicht in Abhängigkeit von der Leseverstehenskompetenz getestet werden, und umgekehrt sollen in Situierung und Aufgabenstellung keine sprachlichen Mittel vorgegeben werden, die beim Schreiben genutzt werden könnten; dies umso mehr, als die Schüler/innen je nach Lehrperson aus dem Unterricht einen ganz unterschiedlichen Umgang mit vorhandenen sprachlichen Mitteln gewohnt sind (von Verbot bis Ermutigung zur Nutzung). 3) Anders als in einer normalen Unterrichtssituation müssen die Schülerinnen und Schüler mit den HarMoS-Testaufgaben ganz allein zurechtkommen. Zudem muss angesichts der engen Zeitvorgaben das Aufgabenverständnis möglichst wenig Zeit beanspruchen. Anders als HarMoS können internationale Diplomprüfungen in der Regel davon ausgehen, dass die Kandidat/innen mit den möglichen Aufgabenformaten und der Anweisungssprache vertraut sind, weil sie sich spezifisch vorbereitet haben.

Die **konkrete Aufgabenentwicklung** erfolgte in mehreren Schritten. Ein Kernteam (Ruth Keller und Peter Lenz) verfasste auf der Grundlage der *construct map* eine Reihe von Aufgaben, die in mehreren Zyklen dem Konsortium vorgelegt wurden. Dabei wurden die Aufgaben fortlaufend modifiziert, einige davon auch eliminiert.

Schliesslich wurden in 18 Deutschschweizer Französisch- und Englischklassen Vorerprobungen durchgeführt. Zusätzlich wurden drei Fremdsprachendidaktiker/innen aus der Westschweiz konsultiert. Aufgrund der so entstandenen Schülertexte, der Auswertung der Fragebögen, welche Lehrer/innen und Schüler/innen der Erprobungsklassen auszufüllen hatten, sowie der Anregungen der Fachdidaktiker/innen wurden die Aufgaben weiter optimiert.

Zum Schluss wurden die Aufgaben auch auf Französisch übersetzt und schliesslich in den Satz gegeben.

Es zeigte sich später, dass die Aufgabe 602 „Glückwünsche zum Geburtstag“ trotz all dieser Massnahmen in der Untersuchung selbst von den Sechstklässler/innen so uneinheitlich verstanden wurde, dass die daraus entstandenen Schülertexte nicht sinnvoll in die HarMoS-Untersuchung zum Schreiben einbezogen werden konnten.

### **3.6.3 Datenerhebung**

In der von der EDK organisierten Validierungsuntersuchung mussten die Schülerinnen und Schüler maximal zwei Texte in einer Fremdsprache schreiben, und zwar entweder in der gleichen oder in zwei verschiedenen Fremdsprachen. Jede Aufgabe wurde pro Sprache und Gruppe, in der sie eingesetzt wurde, von rund 150 bis 250 Schülerinnen und Schülern bearbeitet. Da die Testhefte auf der Ebene der Individuen rotiert wurden, ergaben sich sehr viele verschiedene Aufgabenkombinationen innerhalb jeder einzelnen Sprache und auch über die am selben Halbtage getesteten drei Sprachen (L1 und zwei Fremdsprachen) hinweg. Die Bearbeitungszeit betrug bei acht von neun Schreibaufgaben in den Fremdsprachen 20 Minuten, bei der Aufgabe 912 „Handyverbot“ 30 Minuten. Hilfsmittel waren nicht erlaubt. Wegen der Anlage der Untersuchung (die Verantwortung für die Durchführung wurde den Lehrpersonen übergeben, und die Konsortien hatten keinen Zugang zu den beteiligten Klassen), ist sehr wenig bekannt über den Verlauf des Testvormittags und über die Reaktionen der Schüler/innen und Lehrer/innen auf die Testmodalitäten und die Aufgabenstellungen. Neben den Schülerprodukten liegen lediglich die wenigen Soziodaten vor, die mithilfe eines kurzen Schülerfragebogens erhoben wurden.

### 3.6.4 Rating

Im Anschluss an die Hauptuntersuchung mussten die Texte nach Kriterien beurteilt, im Fachjargon „geratet“, werden. Dabei wurde zwischen einem **holistischen** und einem **analytischen Rating** unterschieden. Die beiden Ratings wurden zeitlich verschoben und unabhängig voneinander durchgeführt. Beim holistischen Rating der Schülertexte bestand der Auftrag an die Rater darin, unter Berücksichtigung von einigen vorgegebenen Gesichtspunkten, für jeden Text zu einer einzigen, globalen Bewertung zu kommen. Für das analytische Rating wurden diese und weitere Gesichtspunkte aufgeschlüsselt und konkretisiert und dann von den Ratern getrennt erfasst.

Angesichts der zeitlichen und personellen Ressourcen konnten nicht ganz alle Schülertexte geratet werden. Wenn überdurchschnittlich viele Texte einer Subpopulation zu ein und derselben Aufgabe vorlagen<sup>56</sup>, wurde vor Beginn des holistischen Ratings eine relativ kleine Zahl von Texten nach dem Zufallsverfahren ausgeschieden. Die Arbeiten zur Aufgabe 602 „Glueckwunsch“ wurden nach einer ersten Sichtung gar nicht einbezogen. Die Arbeiten zu den vier Aufgaben 801 „Porträt“, 807 „Typischer Schultag“, 808 „Erlebnis mit einem Haustier“ und 811 „Meinung zu Hausaufgaben“ wurden in allen Sprachversionen auch *analytisch*, mithilfe eines längeren Kriterienkatalogs (s. u.) geratet. Eine analytische Beurteilung war allerdings erst ab einer Textlänge von rund 25 Wörtern möglich. Dies führte vor allem bei den Texten der Sechstklässler zum Ausschluss eines mehr oder weniger grossen Teils der Texte (je nach Aufgabe). Pro Aufgabe/Zielsprache liegen zwischen 100 und über 200 analytisch geratete Texte vor. Je nach Aufgabe und Zielsprache stammen diese Texte aus 6. und 9. Klassen oder auch nur aus einer der beiden Stufen.

Das holistische Rating kann als Rückgrat der Untersuchung gesehen werden. Die Datenbasis ist so strukturiert, dass mittels einer (Mehrfacetten-)Rasch-Analyse alle Schülerinnen und Schüler auf einer einzigen, „fairen“, Skala abgebildet werden können, die der unterschiedlichen Aufgabenschwierigkeit und der Strenge der einzelnen Beurteiler Rechnung trägt. Mit dem analytischen Rating sollten differenziertere Einsichten ermöglicht werden, u. a. mit dem Ziel, die Ratinginstrumente für künftige Untersuchungen zu optimieren; weiter sollte eine Datenbasis geschaffen werden, um in vertiefenden Untersuchungen in den Kompetenzprofilen der Lernenden charakteristische Muster zu identifizieren.

Beim **holistischen Rating** wurde jeweils eine Skala mit einer Abstufung von 0 bis 5 eingesetzt, wobei bei der Beurteilung „0“ nur dann vergeben wurde, wenn nichts geschrieben wurde, oder jedenfalls nichts im Sinne der Aufgabe. Die Stufen 1 bis 5 sind zwar aufgabenspezifisch, aber für alle Aufgaben möglichst analog beschrieben. Die verschiedenen Gesichtspunkte, die in den Stufenbeschreibungen angesprochen werden, betreffen einerseits funktional-pragmatische, andererseits sprachliche Aspekte. Die Stufe 2 der Ratingskala zu Aufgabe 807 ist beispielsweise so beschrieben:

#### 2 Ein Text auf dieser Stufe zeigt weitgehend die folgenden Merkmale:

- ist ein kurzer, teilweise bruchstückhafter Text, der eine beschränkte kommunikative Wirkung erzielt
- ist eine elementare und/oder teilweise inkonsistente Beschreibung, in der einige Aspekte erwähnt werden
- die Aussagen sind einfach formuliert und folgen oft derselben einfachen Satzstruktur
- weist ein begrenztes Spektrum an elementaren sprachlichen Mitteln auf, enthält mehrere irritierende Fehler

Aufgabe der Rater/innen war es zu bestimmen, welche der Stufenbeschreibungen (0, 1, 2 ...5) *am ehesten* auf den jeweils vorliegenden Schülertext zutrifft. Für die Beurteilung der Texte in den verschiedenen Zielsprachen wurden pro Aufgabe dieselben Skalen verwendet, sodass insgesamt acht, nur in einigen aufgabenspezifischen Konkretisierungen voneinander abweichende, holistische Skalen zum Einsatz kamen.

---

<sup>56</sup> Die ungleiche Anzahl von Schülertexten je nach Aufgabe und Subpopulation ist vorwiegend ein Ergebnis der Itemrotation: Wenn mehr Schüler/innen in der insgesamt vorgegebenen Testzeit weniger Aufgaben bearbeiten mussten, entstanden mehr Schülerarbeiten zur selben Aufgabe. So hatten die Sechstklässler in der Westschweiz am Sprachen-Testhalbttag insgesamt ein kleineres Aufgabenspektrum zu bearbeiten als die übrigen Gruppen.

Da die Beurteilungsstufen 1 bis 5 in Bezug auf die Anforderungen der jeweiligen Aufgabe beschrieben sind, müsste ein und derselbe Schüler bei einer leichten Aufgabe grundsätzlich eine höhere Bewertung erhalten als bei einer schwierigen, sofern es ihm gelingt, seine Kompetenzen bei beiden Anlässen in gleichem Masse zu mobilisieren. Bei diesem Vorgehen handelt es sich somit nicht um eine Zuordnung zu einer allen Aufgaben gemeinsamen Referenzskala, beispielsweise den Niveaus des *Europäischen Referenzrahmens*. Diese Zuordnung erfolgt erst in einem zweiten Schritt auf der Grundlage der empirisch ermittelten Skalierung aller Schülertexte zu den verschiedenen Aufgaben. Der Hauptgrund für dieses indirekte Vorgehen ist der, dass auf diese Weise – und dank des Umstands, dass ein Teil der Schüler/innen jeweils zwei Schreibaufgaben bearbeitet hat – mittels Rasch-Skalierung der Faktor Aufgabenschwierigkeit quantifiziert werden kann. Menschlichen Ratern würde es vermutlich nicht mit der wünschbaren Sicherheit gelingen, den Einfluss der Aufgabenschwierigkeit auf die Leistungen zu schätzen und in den Beurteilungen als relativierendes Element angemessen zu berücksichtigen. Empirisch abgesicherte Aussagen zur Aufgabenschwierigkeit machen zu können, ist ein wichtiger Schritt hinsichtlich der Konstruktvalidierung und eine wichtige Vorarbeit für zukünftige Aufgabenspezifikationen.

Pro Aufgabe standen für das holistische Rating in der Regel vier Rater für jede der drei Zielsprachen zur Verfügung. Da die Kodierung länger dauerte als zuerst angenommen, konnten nicht alle Rater bis zum Schluss mitmachen. Es kamen dafür insgesamt vier neue Rater dazu. Bei jeder Aufgabe wurde pro Zielsprache und pro Teilpopulation (6. oder 9. Klasse bzw. DCH oder WCH) jeweils ein Teil der Texte von allen beteiligten Ratern beurteilt; angestrebt wurden 15 von allen geratete Texte pro Kombination Zielsprache/Population, d. h. unter anderem, dass bei Aufgaben, die sowohl in der sechsten als auch in der neunten Klasse bearbeitet worden waren, ca. 30 von allen geratete Texte zusammenkamen. Das mehrfache Raten von Leistungen durch unterschiedliche Beurteiler erlaubt die Schätzung der Interrater-Reliabilität, die sich wiederum zusammensetzt aus der (internen) Konsistenz des einzelnen Beurteilers und der Beurteilerstrenge. Die Analysen haben gezeigt, dass die Beurteiler zum Teil recht unterschiedlich streng beurteilten. Da die Beurteilerstrenge als so genannte Facette in die statistische Analyse mit der Rasch-Software ConQuest aufgenommen und in der Skalierung weitgehend neutralisiert werden kann, ist dies nur in Extremfällen (an den Skalenenden, wenn beispielsweise 1 und 2 auf 1 nivelliert werden) ein Problem. Die Beurteilerkonsistenz dagegen muss schon im Beurteilungsprozess gesichert werden, weil dazu keine nachträglichen Korrekturen möglich sind.

Diesbezügliche Qualitätssicherungsmassnahmen waren fester Bestandteil des Ratingprozesses: Am Anfang, und dann immer wieder, wenn Texte zu neuen Aufgaben in Angriff genommen wurden, gab es eine Phase des Ratertrainings. Die Hauptziele davon waren, die Rater zu einem Skalenverständnis zu führen, das der Intention der Verfasser entspricht, und die sichere Anwendung der Kriterien auf unterschiedliche Schülertexte vorzubereiten. In einem ersten Schritt wurden die Erfahrungen der Beurteilenden jeweils auch dazu genutzt, die Skalen weiter zu verbessern, v. a. Formulierungen zu desambiguieren. Damit sich die Skaleninterpretationen nach anfänglicher Übereinstimmung nicht allzu rasch wieder auseinanderentwickeln konnten, wurden jeweils nach einer kurzen Zeit des individuellen Beurteilens und dann ungefähr nach der Hälfte der Arbeit wieder gemeinsam zu beurteilende Texte ins Spiel gebracht, deren Ratings dann in den sprachspezifischen Gruppen gemeinsam besprochen wurden. Die Übereinstimmung zwischen den Beurteilenden wurde in den ersten Tagen auch immer wieder mit kleinen statistischen Tests zur Interrater-Reliabilität (Cohens *Kappa* für jeweils zwei Rater) überprüft. Bei Auffälligkeiten bei bestimmten Ratern wurde jeweils das Gespräch gesucht<sup>57</sup>.

Zwischen den Sprachen wurden abgesehen von einem gemeinsamen Einstieg ganz zu Beginn der Rating-Arbeit und dann bei der Einführung jeder neuen Aufgabe (und Skala) weniger Massnahmen zur Abstimmung der Ratings ergriffen als innerhalb der Sprachen, dies insbesondere aus Gründen der Praktikabilität. Die wichtigste Verbindung stellten die identischen Beurteilungsskalen pro Auf-

---

<sup>57</sup> Eingriffe im Verlauf der Ratings sind nicht unproblematisch, weil z. B. die Gefahr besteht, dass Rater/innen aufgrund der Vergleiche in ihrem Urteil plötzlich strenger oder weniger streng werden. Diese Art von Inkonsistenz kann statistisch nicht direkt aufgefangen werden. Das Konsortium entschied sich für die erwähnte Art von Begleitung des Ratingprozesses, weil es ihm wichtiger und effizienter erschien, die Entstehung von divergierenden Sichtweisen auf die Lernertexte zu verhindern als (nur) gewisse Effekte nachträglich statistisch aufzufangen.

gabe und die Diskussionen der „Sprachgruppen“ mit denselben Verantwortlichen aus dem Konsortium dar.

Die Rater haben jeweils nur in einer Sprache Texte beurteilt. Aufgrund von Beobachtungen während der Rating-Sessionen haben wir Grund zur Annahme, dass die Rater im Durchschnitt pro Sprache nicht gleich streng waren. Deshalb darf die durchschnittliche Strenge der Rater pro Sprache nicht einfach gleichgesetzt werden, um Vergleichbarkeit zwischen den Sprachen zu schaffen. Diese kommt erst durch die Zuordnung der drei empirisch ermittelten sprachspezifischen Skalen zur europäischen Referenzskala zu Stande (Vorgehen siehe weiter unten).

Beim **analytischen** Rating wurde ähnlich vorgegangen wie beim holistischen. Allerdings musste, wie bereits erwähnt, wegen des gewaltigen Zeitbedarfs eine Aufgaben- bzw. Textauswahl getroffen werden. Zudem war zum Teil kein Mehrfachrating der Texte in befriedigendem, d. h. konsequent nutzbarem Umfang möglich. Aus diesen Gründen eignen sich die Ergebnisse des analytischen Ratings – wie von Anfang an vorgesehen – nur für ergänzende, vorwiegend explorative Untersuchungen, beispielsweise aber nicht für Aussagen über die Schülerpopulationen insgesamt. Die einzelnen Items der analytischen Beurteilungsinstrumente gruppieren sich um die folgenden Gesichtspunkte:

- 1) kommunikative Wirkung (global) und Relevanz des Textes
- 2) Beachtung und Elaboriertheit der Inhaltspunkte
- 3) pragmatische Aspekte auf der Textebene:
  - Textlänge
  - Textstrukturierung
  - Themenentwicklung
- 4) lexikalische Aspekte
  - lexikalisches Spektrum
  - lexikalische Korrektheit
  - linguistische Aspekte der Textstrukturierung (Kohäsion)
- 5) grammatische Aspekte
  - Verbalbereich
  - Satzbereich

Die Beurteilungsraster zu den vier auch analytisch gerateten Aufgaben sind, so weit wie möglich, identisch, mindestens aber analog gehalten. Neben sprachlichen Aspekten auf der pragmatischen und linguistischen Ebene wird auch der Inhalt (ausgehend von den Leitpunkten in den Aufgabenstellungen) detailliert erfasst, dies in der Absicht, etwas mehr Licht in das generell recht „sperrige“ Verhältnis von Inhalt und Textqualität zu bringen. Die Kriterien, die sich auf sprachliche Merkmale beziehen, sind auf einer vergleichsweise konkreten Ebene formuliert und betreffen zum Beispiel den Gebrauch der Tempora, die verwendeten Satzformen oder die verwendeten Kohäsionsmittel. Erwartet wurde, dass aufgrund der Ergebnisse z. B. die zum Teil wenig konkreten Formulierungen von Kriterien im *Referenzrahmen* („einfache grammatische Strukturen“) expliziter gefasst und für zukünftige Instrumente vorgeschlagen werden könnten. Sie sollten auch genutzt werden können, um in vertiefenden Untersuchungen bei den Schüler/innen ggf. ziel- und ausgangssprachenspezifische Kompetenzprofile zu identifizieren.

### **3.6.5 Ergebnisse der Untersuchung**

Die Skalierung der Daten sowohl aus den holistischen als auch aus den analytischen Ratings erwies sich als aufwändiger als ursprünglich angenommen. In Anbetracht der zeitlichen Vorgaben musste deshalb das an sich naheliegende Verfahren umgedreht werden: Die Auswahl von Schülertexten zur Illustration der Expertenvorschläge für Basisstandards konnte (vorerst) nicht aufgrund der empirischen Skalierungen erfolgen, sondern musste aufgrund von Expertenurteilen getroffen

werden. Dabei dienen die Niveaubeschreibungen im Dokument zu den HarmoS-Basisstandards, die auf verschiedenen *Referenzrahmen*-nahen Dokumenten basieren, sowie die *Benchmark*-Texte, die der Europarat zur Illustration der Referenzniveaus veröffentlicht hat, als Arbeitsgrundlage.

Erst später konnten dann die drei empirisch ermittelten Skalen für Deutsch, Französisch und Englisch als Fremdsprachen (je eine Skala) den europäischen Referenzniveaus zugeordnet werden. Alle drei Skalen basieren auf den holistischen Ratings. Für die Zuordnung zum *Referenzrahmen* wurde wiederum ein Verfahren gewählt, das auf Expertenkonsens beruht, aber ähnlich dem bekannten *Bookmark*-Standardsetting die empirische Skala *aller* Texte in einer Sprache als Ausgangspunkt nimmt. Konkret gaben sich die Experten die Aufgabe, auf der Skala mindestens zwei weit auseinanderliegende Punkte zu identifizieren, an denen Niveauübergänge stattfinden. Anhand der konkreten Beispiele zeigte sich, dass an den Übergängen zwischen den Niveaus A1.2 und A2.1 sowie B1.1 und B1.2 genügend geeignete Schülerarbeiten vorliegen, um mit einiger Sicherheit die entsprechenden *bookmarks* setzen zu können. Ausgehend von diesen Fixpunkten wurden anschließend unter Annahme gleicher Intervalle die übrigen Niveauübergänge rein rechnerisch bestimmt<sup>58</sup>. Die sich daraus ergebenden Niveauzuordnungen der einzelnen Arbeiten wurden anschließend anhand von Stichproben auf ihre Plausibilität hin überprüft. Durch den Bezug aller drei Skalen zur Referenzskala (europäische Referenzniveaus) sind die Ergebnisse, die in den drei (Fremd-)Sprachen erreicht wurden, nun direkt quantitativ vergleichbar. Trotzdem sollte bei jeglichen Vergleichen über die Sprachen hinweg die Unschärfe des gewählten *Equating*-Verfahrens im Auge behalten werden.

Wie schon beim Leseverstehen werden die Ergebnisse auf einer „HarmoS-Skala“ ausgedrückt, die sich an der empirischen Verteilung der Ergebnisse in Deutsch und Französisch (jeweils 6. und 9. Klasse) orientiert<sup>59</sup>. Diese Skala (s. unterste Tabellenzeile) entspricht den europäischen Referenzniveaus und den HarmoS-Stufen wie folgt:

<i>Europäischer Referenzrahmen</i>	A1.1	A1.2	A2.1	A2.2	B1.1	B1.2	B2 und höher
<b>HarmoS Klasse 8</b>	I <sub>8</sub>	II <sub>8</sub>	III <sub>8</sub>	IV <sub>8</sub>	V <sub>8</sub>		
<b>HarmoS Klasse 11</b>	I <sub>11</sub>		II <sub>11</sub>	III <sub>11</sub>	IV <sub>11</sub>	V <sub>11</sub>	
<b>HarmoS-Schreibskala (500/100)</b>	(477)	527	577	627	677	(727)	

Grau hervorgehoben sind die von den Experten vorgeschlagenen Mindeststandards für die Klassenstufe 8 (heutige 6. Klasse) und die Klassenstufe 11 (heutige 9. Klasse). Für die HarmoS-Schreibskala werden die Zahlenwerte für die Niveauübergänge angegeben, weil dies für die Mindeststandards relevant ist: Erwartet werden ein „volles A1.2“ für die Klasse 8 bzw. ein „volles A2.1“ für die Klasse 11. Die Mittelwerte für die verschiedenen Niveaus können bei Bedarf leicht berechnet werden.

Ähnlich wie im Abschnitt zum Leseverstehen werden im Folgenden einige Ergebnisse zur Verteilung der Schreibkompetenz in den Populationen dargestellt. Diese Ergebnisse werden auch zu den Expertenvorschlägen für Basisstandards in Beziehung gesetzt, damit ersichtlich wird, inwie-

<sup>58</sup> Beim *Bookmark-Standardsetting* wird nicht generell von gleichen Distanzen zwischen Niveauübergängen ausgegangen. Das Vorgehen lässt sich aber aufgrund der Eigenschaften der beiden Skalen rechtfertigen: Die europäische Referenzskala wurde durch *Rasch*-Skalierung ermittelt und ist damit eine lineare Intervallskala. Zudem umfassen die Niveaus auf dieser Skala im fraglichen Bereich jeweils gleich viele Logits (Einheiten in der *Rasch*-Skalierung). Die empirische Schreibskala ist eine Skala gleichen Typs, die zwar psychometrisch gesehen nur mehr oder weniger parallel zur Referenzskala verläuft und deren Einheiten nicht direkt verglichen werden können, die aber doch als so weit gleich genommen werden kann (und muss, wenn das Aufeinander-Beziehen legitim sein soll), dass die relativen Niveauabstände übernommen werden können.

<sup>59</sup> Dem mittleren Ergebnis, das in Deutsch und Französisch erreicht wurde, entsprechen 500 Punkte; die Standardabweichung der Verteilung beträgt 100 Punkte.

weit diese schon zum jetzigen Zeitpunkt erreicht würden, d. h. bevor sich die neuen Szenarien für den Fremdsprachenunterricht auswirken können.

<b>Schreiben</b> <b>Deutschschweiz</b>	<b>Französisch</b>
<b>6. Klasse</b> <b>Basisstandard 1./2. FS (A1.2)</b>	<b>527</b>
<b>6. Kl. alle</b> Der Basisstandard würde heute von 20% der Sechstklässler/innen erreicht.	<b>462</b>
6. Kl. Mädchen	471
6. Kl. Knaben	452
<b>9. Klasse</b> <b>Basisstandard 1./2. FS (A2.1)</b>	<b>577</b>
<b>9. Kl. alle</b> Der Basisstandard würde heute von 40% der Neuntklässler/innen erreicht.	<b>553</b>
9. Kl. Mädchen	568
9. Kl. Knaben	533
9. Kl. Grundansprüche	485
9. Kl. erweiterte Ansprüche	553
9. Kl. hohe Ansprüche	630

<b>Schreiben</b> <b>Westschweiz</b>	<b>Deutsch</b>
<b>6. Klasse</b> <b>Basisstandard 1./2. FS (A1.2)</b>	<b>527</b>
<b>6. Kl. alle</b> Der Basisstandard würde heute von 20% der Sechstklässler/innen erreicht.	<b>453</b>
6. Kl. Mädchen	466
6. Kl. Knaben	441
<b>9. Klasse</b> <b>Basisstandard 1./2. FS (A2.1)</b>	<b>577</b>
<b>9. Kl. alle</b> Der Basisstandard würde heute von 35% der Neuntklässler/innen erreicht.	<b>536</b>
9. Kl. Mädchen	548
9. Kl. Knaben	523
9. Kl. Grundansprüche	447
9. Kl. ohne Niveautrennung	525
9. Kl. erweiterte Ansprüche	543
9. Kl. hohe Ansprüche	579

<b>Schreiben</b>	
<b>Deutschschweiz und Westschweiz</b>	<b>Englisch</b>
<b>9. Klasse</b>	
<b>Basisstandard 1./2. FS (A2.1)</b>	<b>577</b>
<b>9. Kl. alle</b>	<b>578</b>
9. Kl. Mädchen	592
9. Kl. Knaben	562
<b>DCH: 9. Kl. alle</b>	<b>584</b>
Der Basisstandard würde heute von 50% der Neuntklässler/innen erreicht.	
DCH: 9. Kl. Mädchen	598
DCH: 9. Kl. Knaben	567
DCH: 9. Kl. Grundansprüche	532
DCH: 9. Kl. erweiterte Ansprüche	591
DCH: 9. Kl. hohe Ansprüche	644
<b>WCH: 9. Kl. alle</b>	<b>558</b>
Der Basisstandard würde heute von 40% der Neuntklässler/innen erreicht.	
WCH: 9. Kl. Mädchen	568
WCH: 9. Kl. Knaben	546
WCH: 9. Kl. Grundansprüche	478
WCH: 9. Kl. ohne Niveautrennung	525
WCH: 9. Kl. erweiterte Ansprüche	553
WCH: 9. Kl. hohe Ansprüche	597

Auf direkte Vergleiche *zwischen* den Sprachen (mit Angabe von Signifikanzen) wird an dieser Stelle verzichtet, weil die Verbindung zwischen den drei Skalen, wie erwähnt, auf einem Equating mittels Bezug zu den Referenzniveaus beruht. Die numerischen Ergebnisse bestätigen aber die Eindrücke, die während des Ratings entstanden sind, insbesondere den, dass die Neuntklässler/innen im Fach Englisch bessere Texte geschrieben haben als in den Landessprachen; besonders in der Deutschschweiz entstanden insgesamt bessere Texte in Englisch als in Französisch. Im Schreiben von englischen Texten schnitten die Deutschschweizer Jugendlichen auch signifikant besser ab als die Jugendlichen aus der Westschweiz. Wie schon beim Leseverstehen zeigten in allen Populationen und Subpopulationen die Mädchen wiederum klar bessere Leistungen als die Knaben<sup>60</sup>. Die Ergebnisse im Fertigungsbereich Schreiben sind auch insgesamt recht gut mit denjenigen Leseverstehen vergleichbar.

Generell, besonders aber im Fall der Sechstklässler/innen, verhält es sich so, dass die vorgeschlagenen Basisstandards heute nur von einer Minderheit der Schüler/innen erreicht würden, dies obschon die Standards für die erste Fremdsprache in Klasse 8 sowie generell in Klasse 11 beim Schreiben tiefer angesetzt sind als für die übrigen kommunikativen Sprachfertigkeiten.

<sup>60</sup> Nicht signifikant ist der beobachtete Abstand einzig im Fach Französisch in der sechsten Klasse. Bei den T-Tests zur Bestimmung der Signifikanz wurden die Daten aufgrund der Merkmale der Schülerstichproben gewichtet; es wurde auch ein Designeffekt von 2 angenommen, was sich in Form einer konservativeren Einschätzung der Signifikanzen ausgewirkt hat.

Das im Vergleich zu den vorgeschlagenen Standards bescheidene Ergebnis kann durchaus ein Spiegel der recht geringen Schreibkompetenzen der Schülerinnen und Schüler sein. Vor allem bei den Sechstklässler/innen lassen sich die im Mittel sehr bescheidenen Ergebnisse zum Teil sicher dadurch erklären, dass der schriftliche Ausdruck in kommunikativer Absicht bisher in der Regel nicht zu den (vorrangigen) Lernzielen zählt. Das Ergebnis kann aber teilweise auch ein Effekt sein der anspruchsvollen Bedingungen, unter denen die Schülertexte entstanden sind. So lässt sich anhand der zahlreichen leer oder fast leer abgegebenen Blätter annehmen, dass vor allem viele der jüngeren Schülerinnen und Schüler angesichts der knapp bemessenen Zeit (20 Minuten pro Text) schlicht resignierten. Die anspruchsvollen Untersuchungsbedingungen dürften also einen sogenannten *Floor*-Effekt gezeitigt haben: Schüler/innen, die unter günstigen Umständen vielleicht ein Niveau A1 hätten ausweisen können, waren unter den konkreten Bedingungen möglicherweise blockiert und konnten nicht zeigen, dass sie eigentlich zu mehr fähig gewesen wären als andere. Die Verteilung der ursprünglich vergebenen Punktzahlen (anstelle der oben gezeigten, auf Skalierung basierenden, die Aufgabenschwierigkeit berücksichtigenden Measures) lassen in Form einer extremen Linkslastigkeit der Verteilung bei den Sechstklässlern (*skew* positiv; viele Nullen und Einsen auf der Skala 0-5) einen solchen Effekt annehmen. Die folgende Grafik zeigt, wie häufig die verschiedenen Bewertungen von den Sechst- bzw. Neuntklässlern bei einer eher einfachen Aufgabe erreicht wurden.

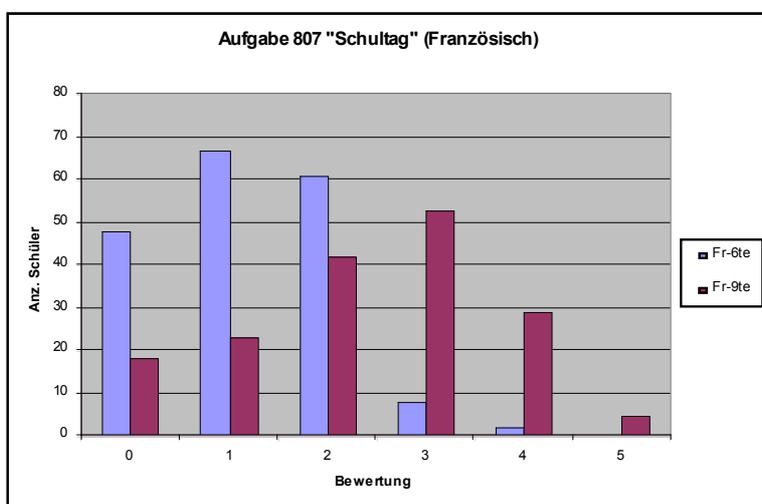


Abb. 17: Verteilung der Punktebewertungen bei Sechst- und Neuntklässlern

Weiter ist nicht auszuschliessen, dass die Richtigkeit der Zuordnung der Leistungen zu den Referenzniveaus wegen der erschwerten Entstehungsbedingungen (ungewohnter Zeitdruck, v. a. für die Jüngeren) mit einem Fragezeichen versehen werden muss: Die Zuordnung der empirischen Skalen zur Referenzskala basiert einerseits auf der Gleichsetzung von Schülertexten (mit ihren Merkmalen) mit Beschreibungen von Schreibkompetenzen und Textmerkmalen, die den verschiedenen Referenzniveaus entsprechen, und andererseits auf einem Vergleich von Schülertexten mit Referenztexten. Die Entstehungsbedingungen der Texte bleiben dabei unberücksichtigt, obschon die Erfahrung zeigt, dass erschwerte Bedingungen systematische Effekte haben können. Diese Möglichkeit muss bei der Interpretation der HarmoS-Ergebnisse mitbedacht werden. Eine „technische“ Aufbesserung der empirischen Ergebnisse kann dagegen nicht vorgenommen werden, weil dazu gesicherte Daten zur Grösse des vermuteten Effekts fehlen.

Bei den Expertenvorschlägen für Basisstandards zum Schreiben gilt – ähnlich wie bei denjenigen zum Leseverstehen – dass sie anspruchsvoll sind und sicher nicht von der angestrebten Zahl von Schülerinnen erreicht werden können, wenn mit den neuen Unterrichtsszenarien nicht wesentlich bessere Ergebnisse erreicht werden als heute.

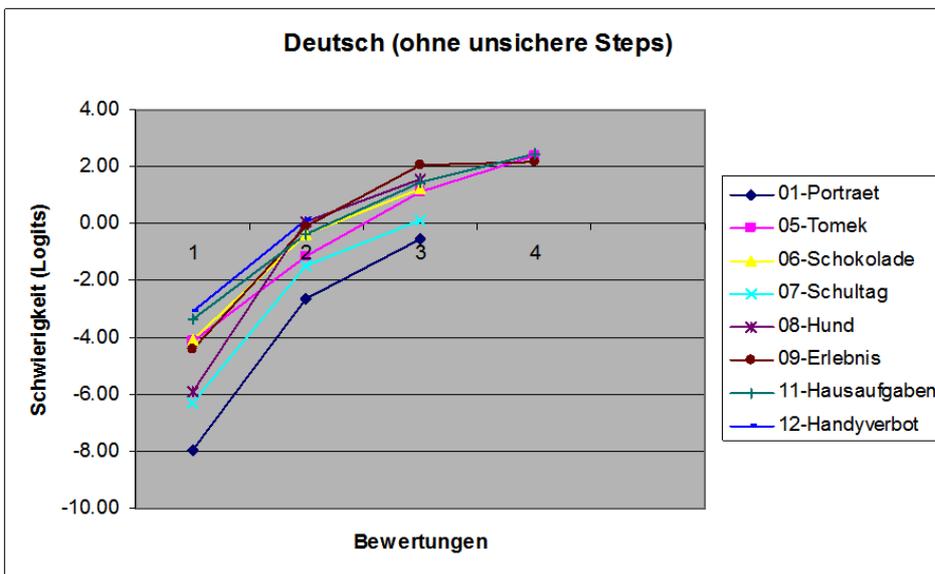
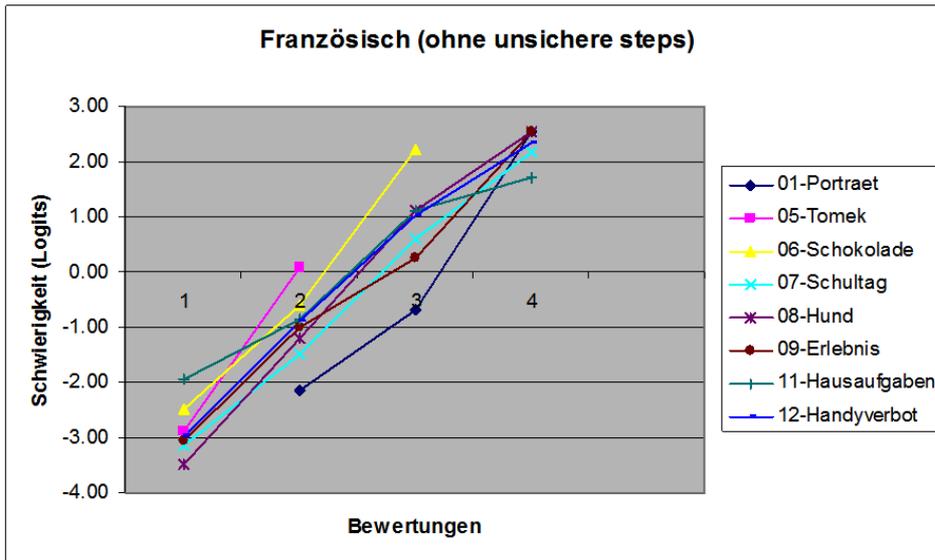
### 3.6.6 Analysen zu den Ergebnissen

Der grosse und vielfältige Datenpool aus HarmoS ruft förmlich nach weit mehr Analysen, als bisher ausgeführt werden konnten. An dieser Stelle soll lediglich auf den Faktor Aufgabenschwierigkeit

(a) sowie auf einige Erkenntnisse aus der Analyse der analytischen Ratings (b) eingegangen werden.

a) Aufgabenschwierigkeit

Wie bereits beschrieben, wurden in den Schreibtests zu den Landessprachen 9 (8 auswertbare), im Englischtest 7 verschiedene Aufgaben eingesetzt. Dabei wurde angenommen, dass die Aufgabensets insgesamt ein breites Kompetenzspektrums bedienen könnten. Die Bewertung der Schülertexte mittels gleich oder analog beschriebener, auf das Produkt zielender holistischer Kategorien ermöglichte eine Schätzung der Schwierigkeit der einzelnen Aufgaben durch *Rasch*-Skalierung – oder etwas genauer: eine Schätzung der Schwierigkeit, bei den einzelnen Aufgaben die Bewertungen 1 bis 5 zu erreichen. Die folgenden Grafiken zeigen die Ergebnisse im Überblick:



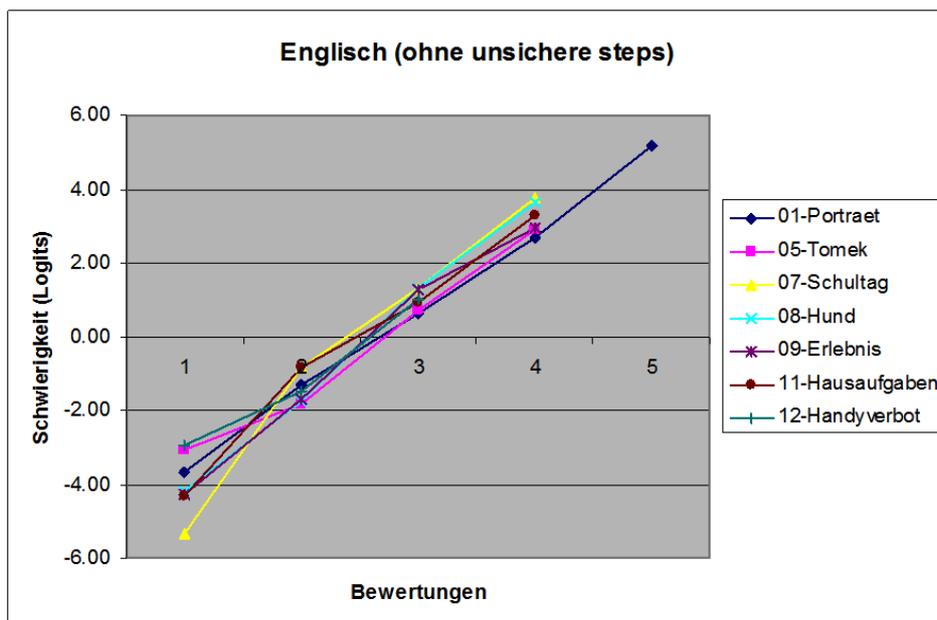


Abb. 18: Vergleich der Schwierigkeiten aller Aufgaben im Französisch-, Deutsch- und Englischtest  
 Dazu vorerst ein Lesebeispiel und eine „technische“ Vorbemerkung:

Lesebeispiel: Auf der Ordinate sind die Schwierigkeits- bzw. Kompetenzeinheiten, *logits* genannt, angegeben, auf der Abszisse die Bewertungen (theoretisch 0 bis 5), genauer die Punkte an denen die jeweilige Bewertung wahrscheinlicher wird als die nächsttiefere. Die Informationen zur Aufgabe „07-Schultag“ (Französisch) können so gelesen werden: Damit ein Schüler bei einem Text zu dieser Aufgabe eher die Bewertung „1“ als „0“ erhält, muss er über eine Kompetenz von -3.13 Logits verfügen, um eher die Bewertung „2“ als „1“ zu erhalten, über eine Kompetenz von -1.49 Logits (auf der Französisch-Schreibskala – zwischen den Sprachen sind in dieser Darstellung nur die relativen Positionierungen vergleichbar). Bei gleicher Kompetenz wäre es nur bei der Aufgabe 01-Portraet leichter gewesen, ebenfalls eine „2“ zu bekommen.

Technische Vorbemerkung: Das Fehlen der Angaben zur Schwierigkeit einzelner (hoher oder tiefer) *steps* (vgl. Knoten auf den Linien) bei verschiedenen Aufgaben ist darauf zurückzuführen, dass die nächsthöhere bzw. nächsttiefere Punktekatgorie zu selten vergeben wurde, als dass verlässliche Skalierungen zustande gekommen wären<sup>61</sup>. Die Anzahl zur Verfügung stehender Ratings war dadurch deutlich eingeschränkt, dass nur diejenigen Schüler/innen in der Skalierung für Schlüsse auf die relative Aufgabenschwierigkeit produktiv genutzt werden konnten, die in der jeweiligen Sprache zwei Texte geschrieben hatten. Auffällige Beispiele für Aufgaben mit nur wenigen statistisch geschätzten *steps* sind bei Französisch die Aufgabe „05-Tomek“<sup>62</sup> und bei Deutsch (und etwas weniger bei Englisch) die Aufgabe „12-Handyverbot“. Die Gründe dürften darin liegen, dass die Aufgabe „05-Tomek“ im Französischtest nur bei Sechstklässler/innen eingesetzt wurden, diese aber ein eingeschränktes Kompetenzspektrum haben; bei der Aufgabe „12-Handyverbot“ dürfte entscheidend sein, dass sie recht anspruchsvoll ist oder zumindest so wirkt.

Bezüglich der Aufgabenschwierigkeit fällt im Überblick auf, dass bei den Landessprachen bzw. dort, wo sowohl Sechst- als auch Neuntklässler/innen getestet wurden, eine Schwierigkeitshierarchie sichtbar wird, die im Grossen und Ganzen dem entspricht, was erwartet wurde (sicher einfacher: „01-Porträt“, „07-Schultag“, „05-Tomek“; sicher schwieriger: „11-Hausaufgaben“, „12-Handyverbot“). Als Ausnahme erscheint im Französischtest die Aufgabe „05-Tomek“, ein Antwortbrief in Form eines E-Mails. Anders als in den beiden anderen Sprachen wurde sie im Französisch nur Sechstklässler/innen vorgelegt. Diese hatten offenbar (eher wider Erwarten) Mühe mit dieser Aufgabe und erreichten kaum je die Bewertung „3“. Zudem gaben über 25% der Sechstklässler/innen ein mehr oder weniger leeres Blatt ab und erhielten die Bewertung „0“. Da diese Aufgabe den

<sup>61</sup> Den gewählten Darstellungen liegt eine Skalierung der *item steps* (50%-Schwelle zwischen zwei Bewertungskategorien, z. B. ‚1‘ und ‚2‘) mit dem *Rasch-partial-credit*-Modell zugrunde.

<sup>62</sup> In diesem Abschnitt werden die Bezeichnungen aus Abb. 18 verwendet. Anders als in der Aufgabenübersicht am Anfang des Kapitels 3.6.2 ist jeweils die erste Ziffer (6, 8 oder 9) weggelassen.

Neuntklässler/innen nicht vorgelegt wurde, fehlt in diesem Fall der solche Effekte oft dämpfende „Blick von oben“ (vgl. dazu das Verhalten von „05-Tomek“ im Deutschttest, wo immerhin rund 20% der Sechstklässler/innen die Bewertung „0“ erhielten, die ebenfalls beteiligten Neuntklässler/innen aber die Schwierigkeitsrelationen offenbar wieder zurechtrückten). Im Deutschttest verhielten sich die beiden Aufgaben, die nur bei Sechstklässler/innen eingesetzt wurden, („01-Portraet“ und „07-Schultag“) den Erwartungen entsprechend und erscheinen leichter als die übrigen Aufgaben. Interessant ist, dass beim Englischttest, an dem nur Neuntklässler/innen beteiligt waren, die Aufgabenschwierigkeiten sehr nah beieinander liegen. Immerhin war bei der Aufgabe „01-Portraet“ die Bewertung „5“ so häufig, dass sie als einzige „5“ in die Skala aufgenommen werden konnte; umgekehrt konnte bei der Aufgabe „12-Handyverbot“ die Bewertung „4“ mangels Daten nicht in die Skala aufgenommen werden. Die geringen Aufgabeneffekte im Englischttest und das überraschende Ergebnis bei der Aufgabe „05-Tomek“ im Französisch könnten Hinweise darauf sein, dass vor allem bei den Sechstklässlern die Aufgaben sehr sorgfältig ausgewählt und formuliert werden müssen, damit es nicht aus Gründen, die nicht ins Zentrum des Schreibkonstrukts gehören, zu einem Floor-Effekt kommt. Umgekehrt zeigen die Ergebnisse auch, dass offenbar auch Aufgaben, die vom Handlungsumfang und -charakter her auf untere Niveaus zielen, auf höheren Niveaus noch differenzieren. Dies dürfte unter anderem darauf zurückzuführen sein, dass in den Texten zu allen Aufgaben für eine Bewertung zum Beispiel mit vier Punkten auch in sprachlicher Hinsicht einiges an positiver und gleichzeitig wenig an negativer Evidenz vorliegen muss. Die Ressourcen, die für eine bestimmte Textqualität und -wirkung mobilisiert werden müssen, scheinen sich unbesehen der Aufgabe weitgehend zu entsprechen<sup>63</sup>. Nicht ganz einfach zu interpretieren ist in diesem Zusammenhang die geringe Zunahme der Schwierigkeitswerte am oberen Ende der Deutschskala. Der parabolische (anstatt lineare) Verlauf der Schwierigkeitskurven und besonders die grossen Sprünge zwischen den Bewertungen „1“ und „2“ deuten am ehesten auf eine strengere Interpretation der Bewertungen „2“ und „3“ durch die Deutschraterinnen hin, sodass schliesslich der inhaltliche Unterschied zwischen den Kategorien „3“ und „4“ nur noch klein war und die beiden fast unterschiedslos vergeben wurden.

b) Die analytischen Ratings sind bisher erst für Englisch näher analysiert worden, und zwar vor allem hinsichtlich der Dimensionalität, die sich innerhalb der analytischen Items zeigt. In diese Analysen wurden teilweise auch die holistischen Ratings einbezogen, dies in der Absicht zu erkennen, wo sich die holistischen Ratings im Verhältnis zu den potenziell innerhalb der analytischen Items sichtbar werdenden Sub-Dimensionen (z. B. „Inhalt“, „Sprache“) situieren<sup>64</sup>. Ziele der explorativen Analysen waren eine Überprüfung der Ratinginstrumente und ein besseres Verständnis der Ratings. Dadurch sollten, im Hinblick auf die Zukunft, auch Hinweise für eine Optimierung der Ratinginstrumente gewonnen werden.

Konkret wurden die Ratings zu drei Englischaufgaben jeweils separat analysiert. Für die Analysen wurde – verzahnt – mit Faktoranalyse und *Rasch*-Skalierung gearbeitet, unter anderem auch mit mehrdimensionalen *Rasch*-Analysen des Programms ConQuest. Die Ergebnisse zu allen drei Englisch-Aufgaben fielen recht ähnlich aus, sodass gewisse Verallgemeinerungen (zumindest für Englisch) nahe liegen. Im Folgenden einige Ergebnisse im Überblick:

- Ein Vergleich zwischen den holistischen und der Gesamtheit der analytischen Ratings bei drei Aufgaben ergibt (nicht-messfehlerbereinigte) Korrelationen von 0.71, 0.74 und 0.83. Leicht höher liegen die Korrelationen, wenn nicht alle analytischen Items in den Vergleich einbezogen werden, sondern lediglich die drei rasch und einfach zu ratenden Items 110 „Kommunikative Wirkung“, 710 „Lexikalisches Spektrum“ und 300 „Textlänge“: 0.72, 0.75, 0.85. (Ein Vergleich zwischen diesen drei Items und den über 20 Items des gesamten analytischen Ratings zeigt seinerseits eine hohe Korrelation von mind. 0.91.) Wenige Prozentpunkte tiefer als das Gesamt der analytischen Items korreliert die Textlänge allein (effektive Wortzahl ausgedrückt in den Intervall-Kategorien 0 bis 5) mit den holistischen Ratings. – Faktoranalysen zu allen drei untersuchten Aufgaben, bei denen die holistischen und die analytischen Daten zusammengelegt

---

<sup>63</sup> Dies scheint zumindest im Fall der vorliegenden Aufgaben und der vorliegenden Rating-Skala zuzutreffen. Damit ist nichts darüber gesagt, inwieweit diese Aussagen auch auf Aufgaben und Texte übertragen werden können, die eine gut ausgebildete CALP (Communicative Academic Language Proficiency) erfordern.

<sup>64</sup> Die Bereiche, die mit dem analytischen Rating erfasst werden sollten, sind weiter oben in Abschnitt 3.6.4. aufgeführt.

wurden, zeigen, dass die Items 110, 300 und 710 tatsächlich sehr ähnlich gerichtet sind wie die holistische Beurteilung, und dass sie zudem sehr stark auf die Rasch-Hauptdimension laden<sup>65</sup>.

- Sowohl die Faktoranalysen als auch darauf aufbauende Berechnungen mit der speziell für die Arbeit mit mehreren psychometrischen Dimensionen geeigneten Software ConQuest weisen darauf hin, dass innerhalb der Items des analytischen Ratinginstruments eine gewisse Dimensionalität besteht, d. h. dass unterschiedliche Kompetenzdimensionen gemessen werden. Grob lassen sich diese Dimensionen charakterisieren als Inhalt+Textaufbau+Textlänge versus Sprache (Wortschatz und Grammatik, inkl. Kohäsion). Die so definierten Itemgruppen zeigen, je nach Aufgabe, Korrelationen von 0.78 bis 0.85. Da ConQuest sog. latente Dimensionen misst, die messfehlerbereinigt sind (und deshalb die obere Grenze des Schätzintervalls angeben), kann tatsächlich von unterschiedlichen Dimensionen gesprochen werden. Sie treten aber nicht so deutlich hervor, dass eine eindimensionale *Rasch*-skalierung unter Einbezug von Items, die diese wahrscheinlichen Dimensionen erfassen, nicht mehr legitim wäre. Umgekehrt dürfte die sich abzeichnende Dimensionalität ein Hinweis darauf sein, dass in einem Ratinginstrument tatsächlich Items beider Gruppen enthalten sein und ggf. gezielt gewichtet werden sollten.
- Im Zuge der Skalierung der analytischen Items wurden auch Analysen zum *differential item functioning* (DIF) gemacht. Dabei ist aufgefallen, dass abhängig von den Kompetenzaspekten, auf die sich die Items beziehen, offenbar nicht-uniformes DIF vorliegt: Für die schwächeren Schüler wären die inhaltsbezogenen Items eigentlich leichter, die sprachlichen Items dagegen schwieriger als sie aufgrund der Skalierung unter Einbezug der ganzen Schülerpopulation erscheinen. Dies steht (vermutlich) im Einklang mit der Feststellung (anlässlich der nach Dimensionen getrennten *Rasch*-Analyse), dass die inhaltsbezogenen Items für weniger Streuung zu sorgen vermögen als die übrigen Items. Diese ersten Erkenntnisse zum nicht-uniformen DIF sollten vertieft werden, weil dieses Phänomen Konsequenzen für die Formulierung von Kompetenzrastern haben könnte, die sowohl inhaltliche als auch sprachliche Aspekte umfassen.
- Für das analytische Rating wurden Items formuliert, die geeignet sein sollten, grammatische Kompetenz konkreter zu erfassen als dies zum Beispiel durch den (einzelsprachübergreifend formulierten) Raster zur Beurteilung der mündlichen Kommunikationsfähigkeit des *Referenzrahmens* getan wird, z. B. solche, die sich auf Verb- oder Satzformen beziehen. Bei der Skalierung (nur für Englisch) haben sich einige dieser Items als problematisch erwiesen insofern, als sie ein relativ schlechtes Fit aufwiesen oder kaum skalierbar waren, weil die Phänomene, auf die sie sich beziehen, nur selten beobachtet werden konnten (z. B. Verneinungen). Andere Items, etwa zu Haupt- und Nebensätzen, haben sich als produktiv erwiesen. Interessant dürfte ein Quervergleich des Funktionierens der Kriterien bei den drei getesteten Fremdsprachen sein.

### 3.6.7 *Stand der Arbeit und Ausblick*

Wie aus dem oben Geschriebenen implizit und explizit hervorgeht, bilden die Daten aus dem analytischen Rating eine gute Grundlage für weitere vertiefende Untersuchungen, vor allem weil jeweils Daten aus drei Tests, die erst noch unterschiedliche Sprachen testen, vorliegen; dadurch sind voreilige Schlüsse aufgrund von Zufälligkeiten viel weniger möglich, als wenn aufgrund einer einzelnen Datenbasis Folgerungen gezogen werden.

Wichtig bzw. lohnenswert wären insbesondere die folgenden Arbeiten:

- Die Finalisierung eines Ratinginstruments, das es erlaubt, über die ganze Skala hinweg effizient die wesentlichen Aspekte so zu erfassen, dass ein faires Gesamturteil zustande kommt. Dazu gehören unter anderem die Identifikation von weiteren wesentlichen Gesichtspunkten zur Beschreibung der grammatischen Aspekte der Schülerleistungen in verschiedenen Sprachen und auf unterschiedlichen Niveaus sowie Entscheidungen bezüglich der Gewichtung verschiedener bzw. unterschiedlicher Kompetenzaspekte.

---

<sup>65</sup> Die Analysen wurden insbesondere mittels Winsteps vorgenommen. Neben den Ergebnissen der Principal-Components-Analysen weisen auch die tiefen Infit-Werte darauf hin, dass in diesen wenigen Items vieles von dem gespiegelt wird, was durch das fusionierte holistische und analytische Rating insgesamt erfasst wird.

- Eine Bereinigung und Erweiterung des Aufgabenpools: Für jüngere und schwächere Lernende sollten mehr Aufgaben vorliegen, die ihnen erlauben zu zeigen, was sie im Bereich des (letztlich) kommunikativen Schreibens können.

Zu überlegen ist auch, ob es tatsächlich sinnvoll ist, mit dem Ziel der Konstruktdeckung, speziell der Berücksichtigung verschiedener Diskurstypen, so viele (7 bzw. 9) Aufgaben einzusetzen. Es dürfte schwierig sein, diesen Punkt aufgrund der empirischen Ergebnisse zu entscheiden, weil mit etwa zwei Aufgaben pro Diskurstyp keine relevante Basis für verallgemeinernde Aussagen vorliegt. Vorhandene Unterschiede zwischen den Ergebnissen zu den einzelnen Aufgaben dürften nicht mit Sicherheit der Zugehörigkeit zu einem bestimmten Diskurstyp zuzuschreiben sein, sondern ebenso gut dem Thema oder der Komplexität des Settings.

- Aussagen bezüglich der Kompetenzprofile von relevanten Untergruppen der getesteten Populationen. Die analytischen Ratings bieten die Datengrundlage zur differenzierenden Beschreibung von Kompetenzprofilen unterschiedlicher Typen von Lernenden über einen relativ grossen Niveaubereich hinweg<sup>66</sup>.

### 3.7 Sprechen

#### 3.7.1 Ausgangspunkt und Grundlagen

Der Fertigungsbereich Sprechen ist bereits heute einer der deklarierten Schwerpunkte des schulischen Fremdsprachenunterrichts – auf der Primarstufe mehr noch als auf der Sekundarstufe. Damit ist allerdings nichts darüber gesagt, wie das Sprechen tatsächlich gewichtet, unterrichtet und evaluiert wird. Insbesondere ist unklar, in welchem Umfang im Unterricht tatsächlich in der Fremdsprache gehandelt wird bzw. inwieweit im mündlichen Unterricht mehr gemacht wird als Übungen und Rollenspiele. Der Sprechunterricht wird auch in den neuen Unterrichtsszenarien seine Bedeutung nicht verlieren. Sowohl im Rahmen einer Austauschpädagogik als auch in einem vermehrt auf Sachinhalte ausgerichteten Unterricht – beides mögliche Entwicklungsrichtungen – spielt das Sprechen eine zentrale Rolle. Im Zusammenhang mit der Beschäftigung mit Sachinhalten dürfte die mündliche Produktion (*Zusammenhängendes Sprechen*) sogar an Bedeutung gewinnen.

Im Hinblick auf diese Szenarien der Sprachverwendung im Zusammenhang mit dem Unterricht wurde auch beim Sprechen, ähnlich wie beim Schreiben, die sprachlich-kommunikative Handlungsfähigkeit mit Hilfe von (grösstenteils) funktional klar eingegrenzten Aufgaben getestet. Alle eingesetzten Aufgaben verlangten relativ spontanes Sprechen mit inhaltlichem Fokus<sup>67</sup>.

Grosse Bedeutung haben in einem handlungsorientierten, sozio-kognitiven Modell pragmatisches und strategisches Wissen und Können (Einsatz von Sprache in unterschiedlichen Situationen, Anpassung des Sprachgebrauchs an die jeweilige Situation; Kompensation, Nachfragen, Kooperation, Selbstkorrektur) (McKay 2006; Europarat 2000, Bachmann/Palmer 1996). Gleichzeitig umfasst die mündliche Sprachfertigkeit als theoretisches Konstrukt aber auch Grammatik- und Textwissen (Wortschatz-, Syntax- und Phonologiekenntnisse; kohärente und strukturierte Sprechweise) wie es beispielsweise mit herkömmlichen auswendig gelernten Rollenspielen einseitig in den Vordergrund gerückt wurde.

Die HarMoS-Tests zur Mündlichen Interaktion umfassten unterschiedliche personelle Konstellationen: einerseits die Interaktion mit Gleichaltrigen Schulkolleg/innen, andererseits ein Gespräch mit einem kompetenten erwachsenen Sprecher der Zielsprache. Die Erwartung dabei war, dass sich durch die unterschiedlichen Konstellationen andere Rolleninterpretationen und damit unterschiedliche Anforderungen vor allem hinsichtlich der pragmatischen und strategischen Kompetenzen ergeben würden. Gemäss Untersuchungen sollte die kognitive und soziale Fertigkeit, Interaktionen

<sup>66</sup> Ein Mitglied des Konsortiums befasst sich auf der Basis der HarMoS-Ergebnisse im Fach Englisch in ihrer Dissertation zum einen mit unterschiedlichen (Typen von) Kompetenzprofilen und zum andern mit der Beschreibung der Leistungen im Rahmen eines Novizen-Experten-Ansatzes (vgl. Bereiter/Scardamalia 1987).

<sup>67</sup> Nicht oder kaum vorbereitete Sprechprodukte entstehen in Ideeneinheiten und sind normalerweise gekennzeichnet durch unvollständige Sätze, Füllwörter (*non-specific words*), die keine oder wenig Information enthalten, und Sprechpausen. Die Ideeneinheiten orientieren sich an der Memorisierungs- und Produktionsfähigkeit des Sprechers / der Sprecherin sowie, in der Interaktion, auch des Zuhörers / der Zuhörerin (Luoma 2004).

kontrolliert zu führen und auf das Gegenüber zu reagieren bis zum zehnten Lebensjahr grundsätzlich entwickelt sein (McKay 2006, Cameron 2001).

Anders als beim Schreiben wurden Interaktion und Produktion getrennt erfasst, weil diese Trennung in der Praxis doch recht stark verankert ist und vor allem weil davon auszugehen war, dass sich die Ergebnisse nicht bei allen Schüler/innen entsprechen würden und dass eventuell in Abhängigkeit vom Kompetenzniveau (oder anderen Variablen wie kognitiver Entwicklung bzw. Alter) Unterschiede auftreten würden. Im Vergleich zur Interaktion, bei welcher der Gesprächspartner den Gesprächsfortgang in der Regel unterstützt, erfordert Mündliche Produktion genügend entwickelte kognitive und linguistische Fertigkeiten, um über längere Zeit ein Gespräch selbstständig aufrecht zu erhalten. Diese Fertigkeiten sind bei jungen Lernenden offenbar stark abhängig von den *literacy skills* in der Erstsprache (McKay 2006).

Wie in Kap. 3.2 beschrieben, wurden zum Sprechen insgesamt 174 Schülerinnen und Schüler getestet, um Anhaltspunkte zum heute vorhandenen Leistungsspektrum zu bekommen. Alle Schülerinnen und Schüler wurden dabei in den drei bereits kurz erwähnten Personenkonstellationen auf ihre Kompetenzen in den beiden Bereichen Mündliche Produktion und Mündliche Interaktion getestet:

- Bei der mündlichen Produktion: selbstständiges monologisches Sprechen;
- Bei der mündlichen Interaktion:
  - Gespräch mit einem gleichaltrigen Gesprächspartner / einer gleichaltrigen Gesprächspartnerin in der Regel aus einer Parallelklasse (*non-native-speaker peer interaction*);
  - Interviewgespräch mit erwachsenem Sprecher / erwachsener Sprecherin der Zielsprache als L1 (*native-speaker / non-native-speaker interaction*).

Dazu kam ein einleitendes Kennenlerngespräch zu dritt.

Im Teil Mündliche Produktion sollten die Schülerinnen und Schüler zeigen, wie gut sie einen längeren, auf einer konkreten schriftlichen Aufgabenstellung basierenden, nur kurz vorbereiteten mündlichen Beitrag abliefern können.

Im Teil *Peer interaction* mussten sie zu zweit eine konkret vorgegebene schriftliche Interaktionsaufgabe lösen, bei der, je nach Aufgabe, eher kooperative oder eher konfrontative Elemente im Vordergrund standen.

Das abschliessende Gespräch zwischen je einem Schüler / einer Schülerin und dem *native speaker* (im Folgenden Interlokutor) hatte weitgehend Interviewcharakter. Anders als in den beiden übrigen Teilen konnten die Lernenden hier auf die Gesprächsführung und Stützung durch die erwachsene, sprachlich kompetentere Person zählen. Diese wiederum hatte den Auftrag, in diesem Gespräch durch ein adaptives Verhalten die sprachlichen Grenzen der Schülerinnen und Schüler auszuloten.

### 3.7.2 Instrumente (Aufgaben)

Die Entwicklung der Aufgaben zum Sprechen folgte im Wesentlichen denselben Schritten und bediente sich derselben Quellen (Dokumente und Instrumente aus dem Umfeld des *Referenzrahmens; lingualevel*) wie die Aufgaben zum Schreiben. Dazu kam Fachliteratur zur Sprechkompetenz und zum Testen und Beurteilen des Sprechens<sup>68</sup>.

Wiederum entstanden **construct maps** in denen für verschiedene Kompetenzniveaus Merkmale von Aufgaben und Leistungen auf verschiedenen Niveaus festgehalten wurden. Sie dienten als Grundlage für die Entwicklung von Aufgaben und Beurteilungsskalen. Auch beim interaktiven und produktiven Sprechen wurde darauf geachtet, dass die verschiedenen Diskurstypen durch geeignete Aufgabenteile abgedeckt wurden. Dabei wurde allerdings eine gewisse Schwerpunktsetzung zwischen den Prüfungsteilen vorgenommen: Der Diskurstyp Erzählen/Berichten wurde in der mündlichen Interaktion zwischen Gleichaltrigen nicht in Aufgaben umgesetzt, weil längeres fortgesetztes Sprechen der Schwerpunkt des Teils zur mündlichen Produktion war. Die Diskurstypen Unterhalten von Beziehung sowie Auffordern/Veranlassen wurden primär bei der Interaktion zwi-

---

<sup>68</sup> Insbesondere Fulcher (2003), Kurtz (2001), Luoma (2004), McKay (2006), Weir (2005)

schen Gleichaltrigen berücksichtigt, indem diese selbständig ein Gespräch aufrecht erhalten und im sprachlichen Austausch etwas aushandeln mussten (z. B. sich auf einen idealen Ort für das Klassenlager einigen oder eine Party organisieren) bzw. den andern zum Handeln auffordern mussten (z. B. einen Gegenstand reichen).

Die folgende Tabelle gibt einen Überblick über Aufgaben, angezielte Diskurstypen und die Verteilung der Aufgaben auf die Klassenstufen.

### Mündliche Produktion

Diskurstypen	P1 Kleines Porträt	P2 Klassenporträt	P3 Von einem Tag erzählen	P4 „Deine“ Stadt	P5 TV – Computer – Telefon	P6 Das Leben frü- her und heute
Informieren/ Beschreiben	X	X	X	X	X	X
Auffordern/ Veranlassen						
Erzählen/ Berichten			X	X	X	
Meinung äussern/ Argumentieren				X	X	X
Unterhalten von Bezie- hung						
<b>Einsätze:</b>						
6. Kl. (D/F)						
9. Kl. (D/F)						
Englisch, 9. Kl., D-CH und W-CH						

### Mündliche Interaktion mit Gleichaltrigen (*peer interaction, non-native*)

Diskurstypen	IP1 Bilder legen	IP2 Klassenlager	IP3 Ein Abschluss- fest planen	IP4 Gemein- same Aktivitä- ten planen	IP5 Schuluniformen?	IP6 Ein Klassenfest organisieren
Informieren/ Beschreiben	X	X	X	X	X	
Auffordern/ Veranlassen	X		X			X
Erzählen/ Berichten						
Meinung äussern/ Argumentieren	X	X	X	X	X	X
Unterhalten von Bezie- hung						
<b>Einsätze:</b>						
6. Kl. (D/F)						
9. Kl. (D/F)						
Englisch, 9. Kl., D-CH und W-CH						

### Mündliche Interaktion mit Erwachsenen (*native speaker*)

Diskurstypen	Themenbereich A Familie/Haustiere Wetter/Jahreszeiten Freizeitaktivitäten Ferien/Reisen	Themenbereich B Fernsehen oder Lesen? Fastfood oder Bio? Gesundheit Technik	Themenbereich C Träume Zukunft Im Lotto gewonnen! Überraschung
Informieren/ Beschreiben	x	x	x
Auffordern/ Veranlassen			
Erzählen/ Berichten	x	x	x
Meinung äussern/ Argumentieren		x	x
Unterhalten von Bezie- hung			
<b>Einsätze:</b>			
6. Kl. (D/F)	1)	2)	2)
9. Kl. (D/F)	1)	2)	2)
Englisch, 9. Kl., D-CH und W-CH	1)	2)	2)

1) Bezeichnet den Startthemenbereich für alle Schulstufen, 2) anspruchsvollere Themenbereiche. Da es sich beim Interviewgespräch um eine Form des adaptiven Testens handelte, wurden je nach Sprechleistung, unabhängig von der Schulstufe, nach Themen aus dem Bereich A solche aus den anspruchsvolleren Themenbereichen B oder – als weitere Steigerung – C gewählt.

Der Aufbau der Aufgaben ist vergleichbar mit dem Aufbau der Schreibaufgaben. Folgende Gliederung wurde dabei berücksichtigt:

- 1) Situierung: Der Schüler / die Schülerin soll sich gedanklich leicht in den Sprechkontext versetzen können:

Situation
<b>In eurer Schule wird diskutiert, ob eine Schuluniform eingeführt werden soll. Eure Meinung interessiert uns!</b>

Abb. 21: Situierung der Aufgabe IP5

- 2) Die Aufgabe mit dem beabsichtigten Sprechziel, Leitpunkten zu den erwarteten Inhalten sowie Hinweisen zu den verlangten Diskurstypen:

Aufgabe	
1) <b>Beschreibe</b> , welche Kleider du normalerweise zur Schule trägst.	
2) <b>Diskutiert</b> : Schuluniformen ja – nein? <b>Vor- und Nachteile?</b>	
3) <b>Wähle die Uniform aus</b> , die dir besser gefällt (1 oder 2). <b>Überzeuge</b> dein Gegenüber von dieser Uniform.	
<p><b>Schuluniform 1</b></p> 	<p><b>Schuluniform 2</b></p> 
4) <b>Einigt euch</b> : <b>Welche Uniform</b> wäre für eure Schule möglich?	

Abb. 22: Strukturierende Aufgabenstellung zur Aufgabe IP5

Beim Kennenlerngespräch zu dritt und bei der Mündlichen Interaktion mit einem erwachsenen Sprecher / einer erwachsenen Sprecherin sollte eine möglichst natürliche spontane Sprechsitua-

tion geschaffen werden, weshalb den Schülerinnen und Schülern weder ein Auftragsblatt vorgelegt, noch Vorbereitungszeit eingeräumt wurde. Beim Kennenlernen waren jeweils zwei Schülerinnen/Schüler gleichzeitig anwesend. Ihnen wurden konkrete, dem erwachsenen Interlokutor / der erwachsenen Interlokutorin vorliegende Fragen gestellt, wobei deren Reihenfolge für die Schülerinnen/die Schüler nicht vorhersehbar war, jedoch einem genauen Prinzip entsprach (gleiche Anzahl Erst- bzw. Folgefragen).

Die Mündliche Interaktion mit einem Interlokutor wurde mit jedem Schüler / jeder Schülerin einzeln geführt. Die Interlokutoren waren angehalten, sich auf das Gespräch mit Interesse einzulassen und sich in der Wahl der sprachlichen Mittel und im Sprechtempo dem Niveau des Schülers / der Schülerin anzupassen. Der Schüler / die Schülerin durfte zu Beginn von vier Themenkarten (Themenbereich A) eine auswählen. Um die Grenzen der sprachlichen Leistungsfähigkeit auszuloten, wurde im Verlauf des Gesprächs die Schwierigkeit auf zwei Ebenen gesteigert: zuerst auf der Ebene der Prompts bzw. Fragen (von offeneren zu enger gefassten Fragen) und dann auf der Ebene der Themen (Einstieg mit Themenkreis A: Informieren und Beschreiben aus dem persönlichen Bereich; Themenkreis B: Meinungsäußerung zu (noch) vertrauten Themen; Themenkreis C: Beschreiben von Visionen und Möglichkeiten im Bereich abstrakter Themen). Diese adaptive Testanlage sollte dem Schüler / der Schülerin die Möglichkeit bieten, ohne Einschränkung durch eine stark vorstrukturierende Aufgabenstellung seine/ihre Sprachkompetenz zu zeigen.

Die **Testaufgaben** wurden in einem Kernteam (Sandra Hutterli und Peter Lenz) aufgrund der *construct map* entwickelt und in mehreren Schritten einem erweiterten Kreis von Expertinnen und Experten aus dem Konsortium und aus dem Beirat zur Diskussion vorgelegt und darauf gründend angepasst bzw. verworfen. Mit einer Stichprobe von Schülerinnen und Schülern aus je zwei ausgewählten Schulklassen (6. Primar- und 9. Sekundarklasse) wurden die Aufgaben in einer *Pretest*-Phase in der Praxis erprobt und videografiert. Die Schülerinnen und Schüler wurden anschliessend an die Erprobung zur Testdurchführung und zu ihrer persönlichen Einschätzung der Tests befragt. Die Auswertung dieser Befragungen und die Analyse der Videoaufnahmen haben folgende Hinweise zu den Aufgaben ergeben:

- 1) Es ist sinnvoll, dass die Anweisungen zu den Aufgaben in der Schulsprache gehalten sind, da die Vorbereitungszeit sehr knapp ist und somit die Aufgabe schnell verstanden werden muss.
- 2) Die Situierung der Sprechhandlung hilft den Schülerinnen und Schülern, sich in die Situation zu versetzen. Sie muss jedoch kurz gefasst (in einem bis maximal drei Sätzen) und direkt auf das Sprechziel ausgerichtet sein.
- 3) Die Inhaltspunkte werden am einfachsten verstanden, wenn Schlüsselbegriffe farblich hervorgehoben und zentrale Aspekte mit Bildern illustriert werden. Dies unterstützt ein möglichst freies, vom Aufgabenblatt losgelöstes Sprechen.
- 4) Die Schülerinnen und Schüler reagierten positiv auf diese Art von Tests und bezeichneten sie als Möglichkeit, ihr Können zu zeigen. Positiv hervorgehoben wurde insbesondere, dass die Themen einen gewissen sprachlichen und inhaltlichen Gestaltungsraum zulassen und nicht explizites Wissen abgefragt wird.

Als **Interlokutoren** wurden Erwachsene *native speaker* gewählt, welche einen Zugang zu Kindern und Jugendlichen sowie Erfahrungen im Umgang mit Fremdsprachigen haben, aber idealerweise nicht als Lehrpersonen tätig sind. Gefordert war ein sensibles und ermutigendes Verhalten, nicht erwünscht war eine schulisch-korrigierende oder allzu stark unterstützende Haltung vor allem im letzten Testteil. Weiter erschien es wünschenswert, dass sie altersmässig nicht zu nah an den Neuntklässlern wären, damit die Rollen intuitiv klar wären. Letzteres Kriterium wurde von den Interlokutorinnen deutscher Muttersprache nicht ganz erfüllt; sie waren mit 25 Jahren noch relativ jung und passten sich in den Gesprächen zuweilen zu stark den Jugendlichen an. Als zusätzlich problematisch erwies sich, dass sie mangels Erfahrung mit der Zielgruppe gelegentlich in ein (zu) stark schweizerisch geprägtes Deutsch verfielen oder die verwendeten sprachlichen Mittel nicht flexibel genug den Möglichkeiten des Gegenübers anpassen konnten.

Die Interlokutoren wurden vorgängig während zwei Tagen geschult. Sie führten mit Schülerinnen und Schülern einer 6. und einer 9. Klasse frühere Fassungen der Tests durch und wurden dabei

gefilmt. Mit Hilfe der Videoanalysen und offenen Feedbacks wurden sie in ihre Aufgabe eingeführt. Zu den Schwerpunkten der Schulung gehörten neben einer generellen Einführung in die Testanlage die Sensibilisierung der Interlokutoren für eine vergleichbare Gesprächsführung sowie eine möglichst natürliche, aber klare Gesprächslenkung.

### 3.7.3 Datenerhebung

Die Mündlichkeitstests fanden in den jeweiligen Schulen statt, was für die Prüferteams von jeweils drei Personen einen beachtlichen Aufwand bedeutete, der bei statistisch repräsentativen Stichproben kaum zu leisten wäre. Alle Tests wurden gefilmt.

Jeder Schüler / jede Schülerin musste alle Testteile durchlaufen:

- Kennenlerngespräch zu dritt;
- Mündliche Produktion;
- Mündliche Interaktion mit gleichaltrigem Gesprächspartner / gleichaltriger Gesprächspartnerin;
- Mündliche Interaktion mit erwachsenem Interlokutor.

Da relativ viele verschiedene Aufgaben vorlagen, wurde die einzelne Aufgabe in jeder der drei Sprachen nur relativ wenige Male bearbeitet. Deutsch- oder Französischaufgaben, die nur entweder bei Sechst- oder Neuntklässlern zum Einsatz kamen, wurden von 8 Schüler/innen gelöst, schulstufenübergreifende Aufgaben bzw. die Englischaufgaben von 14 Schüler/innen. Insgesamt ergab dies 540 Sprechprodukte.

Zeitliche Vorgaben galten lediglich für die Vorbereitung. Vor den Teilen Mündliche Produktion und Mündliche Interaktion mit Gleichaltrigen hatten die Schüler/innen nach einer ersten Lektüre der schriftlichen Aufgabe eine Minute Zeit, um sich gedanklich auf das Thema einzustimmen, das entsprechende Wortfeld zu aktivieren und Stichworte (nicht mehr!) zu notieren. Die Vorbereitungszeit sollte den Schüler/innen die Möglichkeit geben, sich in die Handlungssituation zu versetzen und ihre kommunikativen Ressourcen zu aktivieren. Diese Vorgaben wurden mit dem Ziel einer Annäherung der Bedingungen an vergleichbare Gesprächssituationen ausserhalb der Testsituation gemacht. Vor den beiden anderen Teilen war keine Vorbereitung nötig, da die Gesprächsplanung im Prinzip durch den Interlokutor geleistet wurde und möglichst spontan kommuniziert werden sollte. Bezüglich der Sprechzeit bei den einzelnen Aufgaben bestanden keine festen Zeitvorgaben. Die Schülerinnen und Schüler wurden in der Regel so lange sprechen gelassen, wie sie sich zur Aufgabe äusserten. Der Wechsel zwischen den Testteilen wurde möglichst natürlich gestaltet. Die Testzeit betrug pro Schüler/in insgesamt rund 20 Minuten.

### 3.7.4 Rating

Jeder Test wurde in drei unabhängigen Schritten geratet: 1) Aufgabe zur Mündlichen Produktion; 2) Aufgabe zur *peer interaction* (bei jedem Rating-Durchgang Fokus auf nur einem Schüler); 3) Interviewgespräch (inkl. Kennenlernen ganz zu Beginn).

Die Testteile 1) und 2) zur mündlichen Produktion und Interaktion wurden sehr detailliert geratet. Dabei kamen Instrumente zur Anwendung, die im Konsortium eigens für diesen Zweck entwickelt wurden. Mittels der gewählten Kategorien wird versucht, die für die handlungsorientierte Sicht des *Referenzrahmens* relevanten Kategorien ziemlich umfassend umzusetzen. Viele der Beschreibungen übernehmen Kategorien und Beschreibungselemente aus dem *Referenzrahmen*. Intendiert wird aber keine direkte Zuordnung zum *Referenzrahmen*, sondern eine Beschreibung der Performanz, die absieht zum Beispiel von den unterschiedlichen Anforderungen der verschiedenen Aufgaben – so wie dies auch bei den Aufgaben zum Schreiben gemacht wurde. Die eingesetzten Kriterien sind je nachdem aufgabenspezifisch (Inhaltlich-Thematisches), diskurstypspezifisch (z. B.: Qualität der Informations- oder Argumentationshandlung), fertigkeitsspezifisch (Interaktionsfähigkeit, Produktionsfähigkeit) oder allgemein auf alle mündlichen Leistungen bezogen (linguistische und bestimmte pragmatische Aspekte). Sie sind aber nie nur auf eine der drei getesteten Sprachen bezogen. Die Skalen mit den analytischen Kriterien umfassen jeweils Abstufungen von 0 bis 4 bzw. 5 und sind geeignet, ein grosses Kompetenzspektrum zu erfassen, sodass *Floor-* oder *Ceiling-Effekte* weitgehend ausgeschlossen werden können, d. h. auch bei Leistungen an den Skalenrändern noch differenziert wird. Verschiedene Aufgaben umfassen zwei Teile, die sich je-

weils um einen anderen Diskurstyp gruppieren, z. B. zuerst Informieren/Beschreiben, dann Meinungsäußerung/Argumentieren. In diesen Fällen wurden die beiden Teile getrennt beurteilt, weil grundsätzlich damit zu rechnen ist, dass die unterschiedlichen Anforderungen durch den Diskurstyp Auswirkungen haben darauf, inwieweit die sprachlichen Ressourcen mobilisiert werden können.

Die Verflechtung der Kriterien über Aufgaben- und Fertigungsgrenzen hinweg soll anhand von je zwei Aufgaben zur Mündlichen Interaktion und zur Mündlichen Produktion kurz aufgezeigt werden:

IP2 „Klassenlager“	IP5 „Schuluniform“ (Teil II)	P2 „Klassenporträt“	P5 „TV – Computer – Telefon“
Produktionsfähigkeit/Diskurskompetenz (2 Aspekte)		Interaktionsfähigkeit (8 Aspekte)	
Globale Beurteilung der Aufgabenbewältigung (Typ Informieren/ Beschreiben)	Globale Beurteilung der Aufgabenbewältigung (Typ Meinungsäußerung/ Argumentieren)	Globale Beurteilung der Aufgabenbewältigung (Typ Informieren/ Beschreiben)	Globale Beurteilung der Aufgabenbewältigung (Typ Informieren/ Beschreiben)
Erfüllung der Leitpunkte zu Aufgabe IP2	Erfüllung der Leitpunkte zu Aufgabe IP5	Erfüllung der Leitpunkte zu Aufgabe P2	Erfüllung der Leitpunkte zu Aufgabe P5
Linguistische und pragmatische Aspekte: Aussprache/Intonation Flüssigkeit Kohärenz und Kohäsion Grammatik (Verbalbereich: 3 Aspekte; Satzbereich: 3 Aspekte) Wortschatz (2 Aspekte)			

Auf alle Leistungen zur mündlichen Produktion werden also dieselben Kriterien zur „Produktionsfähigkeit“ angewendet; Gleiches gilt für die mündliche Interaktion. Dagegen wird die globale Aufgabenbewältigung bei verschiedenen Aufgaben nur dann mit identischen Kriterien beurteilt, wenn diese denselben Diskurstyp erfordern. Die Erfüllung der Leitpunkte wiederum wird für jede einzelne Aufgabe spezifisch beurteilt.

Beim dritten Testteil (Interviewgespräch inkl. Kennenlernen) waren die Absicht und das Vorgehen beim Rating grundsätzlich anders: Die Leistungen der Schüler/innen sollten durch Kriterien aus dem *Referenzrahmen* sowie weitere, dem *Referenzrahmen* zugeordnete Kriterien direkt den Referenzniveaus zugeordnet werden. Dies sollte von der Anlage dieses Testteils her auch möglich und sinnvoll sein, lautete doch der Auftrag an die Interlokutoren, im Gespräch die Grenzen der sprachlichen Leistungsfähigkeit auszuloten.

Beim Rating wurde in einem ersten Schritt holistisch eine Globaleinschätzung der Sprachhandlung vorgenommen. Das dabei eingesetzte Instrument entspricht weitgehend der „Globalskala zur mündlichen Kommunikation“ (Schneider/North 2000, 341) aus dem Schweizer Forschungsprojekt, aus dem die meisten Deskriptoren des *Referenzrahmens* hervorgegangen sind, plus einigen Ergänzungen aus dem Projekt *IEF/lingualevel* (im unteren Niveaubereich). Diese erste Einschätzung entspricht einer ersten Aussage dazu, wie gut ein Schüler / eine Schülerin sprachlich handeln kann. Beim darauffolgenden analytischen Rating wurden verschiedene qualitative Aspekte des sprachlichen Handelns separat beurteilt. Das eingesetzte Instrument ist eng an die Tabelle 3 des *Referenzrahmens* (Europarat 2000, 37f.) angelehnt (und wie die Globalskala aus *IEF/lingualevel* ergänzt); sie umfasst die qualitativen Aspekte Interaktion(sfähigkeit), (Wortschatz-)Spektrum, (grammatische) Korrektheit, Flüssigkeit, Aussprache/Intonation und Kohärenz (primär im Sinne von Kohäsion).

Als Ratende wurden pro Sprache je zwei fortgeschrittene Linguistik-Studierende (eine Doktorandin, fünf Lizentiand/innen) eingesetzt. Davon waren zwei deutscher, eine französischer Mutter-

sprache, die andern drei verfügen über eine hohe Sprachkompetenz und Erfahrung im jeweiligen Sprachgebiet.

Die Familiarisierung der Ratenden mit den Instrumenten und eine grundsätzliche Eichung auf die Niveaus (genauer: die Feinniveaus A1.1, A1.2 usw.) des *Referenzrahmens* erfolgten sowohl für die Interaktion als auch die Produktion in sechs Schritten:

1. **Informationen** zu HarmoS, insbesondere zum Kompetenzverständnis; Bezug zum *Gemeinsamen europäischen Referenzrahmen*;
2. **Erläuterung der Aufgaben**, mit Fokus auf den Ablauf, die Inhalte und Ziele;
3. **Selbsteinschätzung der eigenen Sprechkompetenz** in einer individuell gewählten Fremdsprache mit der Globalskala zum Sprechen sowie den Deskriptoren aus dem *Europäischen Sprachenportfolio für Jugendliche und Erwachsene* (Schneider / North / Koch 2000).
4. **Konkrete Familiarisierung** mit der Globalskala für die Kodierung des Testteils 3, Interaktion mit Muttersprachler/innen: Die Ratenden ordnen die einzelnen Beschreibungen nach Niveau und arbeiten anschliessend im detaillierten analytischen Beurteilungsraster zur Interaktion die Charakteristika der einzelnen Niveaubeschreibungen heraus, markieren wichtige Punkte und vertiefen ihr Niveauverständnis in der Diskussion.
5. **Gemeinsames Rating** von sechs Interviews (je zwei pro Sprache) in der Gesamtgruppe. Nach dem Anhören ca. der ersten Hälfte jedes Gesprächs wird eine Globaleinschätzung vorgenommen, im Anschluss daran folgt eine detaillierte Einschätzung mittels der analytischen Kriterien.
6. **Individuelles Rating** von drei weiteren Beispielen (eines pro Sprache), mit anschliessendem Vergleich und einer Diskussion der Ergebnisse.

Um die Vergleichbarkeit der Ratings über die Sprachen hinweg aufrecht zu erhalten, wurden zu Beginn jedes Ratingtages jeweils drei Schülerleistungen (eine pro Sprache) gemeinsam geratet. Dabei wurde zuerst individuell geratet und anschliessend diskutiert, bis ein Konsens erreicht war. Als weitere Massnahme zur Gewährleistung der Interrater-Reliabilität beurteilte eine Expertin aus dem Konsortium zusätzlich eine Stichprobe von 36 einzelnen mündlichen Leistungen. Bei Divergenzen wurde das Gespräch gesucht. Die Interrater-Reliabilität konnte zudem laufend anhand der rund 10% der Schülerleistungen überprüft werden, die jeweils von beiden für eine Zielsprache zuständigen Rater/innen beurteilt wurden. Zur Gewährleistung einer besseren Intrarater-Reliabilität wurde jeweils zu Beginn jedes Morgens vom selben Rater die am Vortag zuletzt geratete Schülerleistung ein weiteres Mal beurteilt; anschliessend wurden die Ratings verglichen und, wenn nötig, Korrekturen vorgenommen. Jeweils nach der Mittagspause beurteilten die beiden Rater/innen jeweils dieselbe Leistung, verglichen die Ratings und diskutierten Differenzen aus.

### 3.7.5 *Analysen und Ergebnisse*

In nennenswerter Weise statistisch ausgewertet werden konnten nur die Ratingdaten zum Testteil 3, dem Interviewgespräch mit erwachsenem muttersprachlichem Gesprächspartner. Die Erstellung der Aufgaben und Ratinginstrumente, die Interlokutorenschulung, die Testdurchführung und das Rating ergaben aber zahlreiche wichtige Erkenntnisse, auf denen in Zukunft aufgebaut werden kann. Im Weiteren entstanden rund 50 Stunden an Filmmaterial sowie eine Reihe von wissenschaftlichen Transkriptionen von Schülerleistungen, die für weiterführende Analysen genutzt werden können.

Die Daten zum Teil 3 wurden in erster Linie mit Hilfe der Statistiksoftware *Facets* ausgewertet, indem eine Mehr-Facetten-*Rasch*-Analyse unternommen wurde. Dadurch konnten besonders gut die unterschiedliche Strenge verschiedener Beurteiler ausgeglichen und die psychometrischen Eigenschaften der Kriterien überprüft werden. Die Ergebnisse zu den Kriterien bestätigen zum Teil, was schon aus früheren Untersuchungen bekannt war, werfen aber auch neue Fragen auf: Wie sich bereits beim Rating abzeichnete, trägt das Kriterium Aussprache/Intonation wenig zu einer Differenzierung der Schüler/innen bei. Die Schüler/innen sind in Bezug auf dieses Merkmal fast alle ziemlich gut; zudem sind nicht unbedingt die sonst Schwachen auch in dieser Hinsicht weniger gut als die anderen. Deshalb ist es keine Überraschung, dass diese Variable aufgrund der psychometrischen Kennzahlen ausgeschlossen werden musste. Wie immer in solchen Fällen ist zu entscheiden, ob und, wenn ja, in welcher Form dieser vor allem auch in den Augen von Laien wichtige Aspekt in zukünftige Ratings einbezogen werden soll. Angesichts der Tatsache, dass es

offenbar fast selbstverständlich ist, dass bei den heutigen Schüler/innen der Akzent die Verständlichkeit kaum je beeinträchtigt, könnte dieses Kriterium wohl ganz weggelassen werden. Eher überraschend zeigte sich bei der Analyse der Beurteilungskriterien, dass auch das Kriterium „(grammatische) Korrektheit“ zumindest bei Deutsch und Englisch keine guten Kennwerte erreichte. Offenbar ist die grammatische Elaboriertheit bzw. Korrektheit für erfolgreiches Kommunizieren nicht zentral wichtig; möglicherweise zeigen die Ergebnisse auch, dass ein Fokussieren auf grammatische Aspekte die generelle Kommunikationsleistung (z. B. in den Bereichen Flüssigkeit und Interaktionsfähigkeit) eher beeinträchtigen. Auch im Fall dieses Kriteriums stellt sich die Frage nach den Konsequenzen für künftige Beurteilungsinstrumente. Wenn der Bereich der Grammatik als eigene Dimension von den übrigen Gesichtspunkten getrennt wird, sollte er besser und vielfältiger gefasst werden. Die auch informell geäußerten Schwierigkeiten der Rater/innen mit diesem Kriterienbereich deuten darauf hin, dass hier ohnehin Handlungsbedarf besteht. Eine bessere und präzisere, d. h. möglicherweise einzelsprachspezifische Fassung der Kriterien auf den verschiedenen Niveaus ist eine Arbeit, die bald an die Hand genommen werden sollte, beispielsweise in Seminar- oder Masterarbeiten.

Zu den primären Zielen der Untersuchung zum Sprechen gehörte ganz einfach auch das Anlegen einer Sammlung von mündlichen Schülerleistungen, die in der ganzen Diskussion um Standards (und Lehrpläne) zur Illustration und als Diskussionsgrundlage genutzt werden kann. Die folgende Grafik gibt einen Überblick darüber, in welcher Sprache welches Material vorhanden ist:

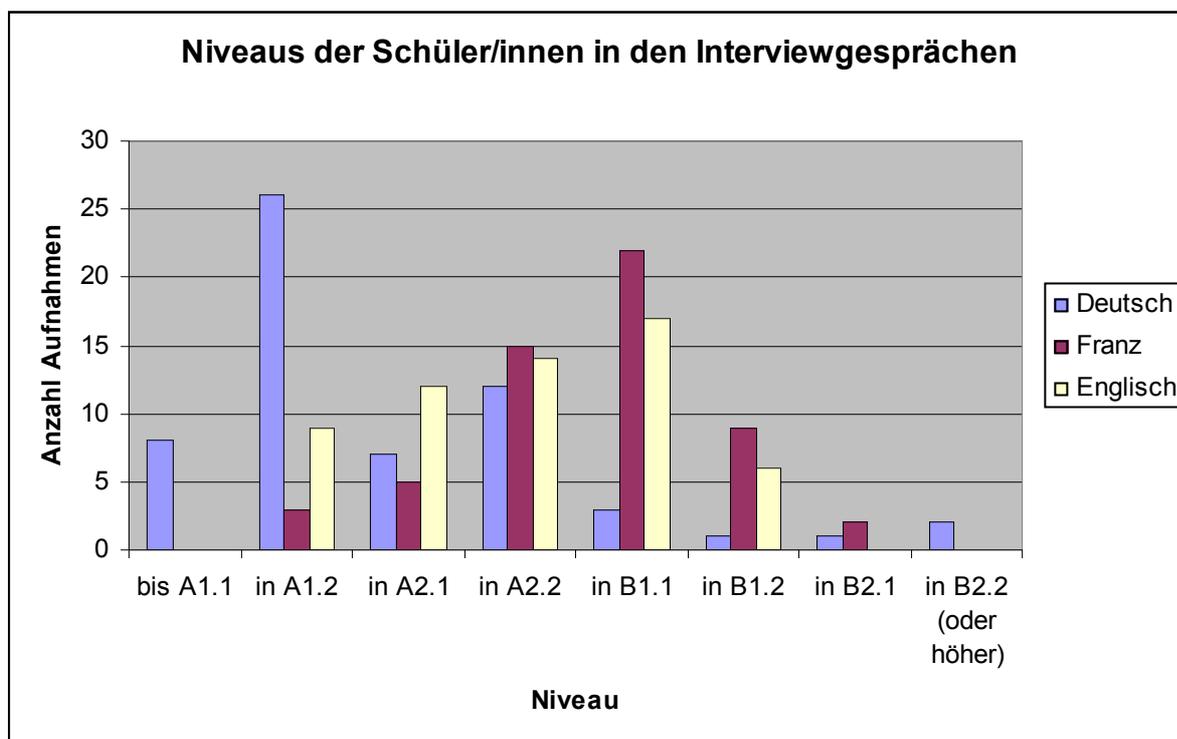


Abb. 19: Übersicht über die Niveaus der in der HarmoS-Untersuchung zum Sprechen getesteten Schüler/innen (basiert auf den Roh-Ratings, nicht auf der Rasch-Skalierung).

Es fällt auf, dass das gewünschte Niveauspektrum nicht für jede Sprache bereitgestellt werden konnte. Für Deutsch als Fremdsprache wurden, in den wenigen, recht zufällig ausgewählten Klassen viele sehr bescheidene Leistungen gefilmt, dies vor allem in den sechsten Klassen. Für Französisch liegen umgekehrt unter anderem ausgezeichnete, aber untypische Leistungen von Sechstklässler/innen aus Deutschfreiburg vor, sodass da eher im unteren Niveaubereich Lücken bestehen. Insgesamt ist aber ein schönes Korpus zusammengekommen, das vielfältig genutzt werden kann. Wie realistisch die als Basisstandards vorgeschlagenen Niveaus sind, lässt sich aufgrund der vorliegenden Ergebnisse zum Sprechen nicht abschliessend sagen, Tatsache ist aber,

dass sich unter den getesteten Schüler/innen recht viele finden, die die Basisstandards ihrer Stufe<sup>69</sup> nicht erfüllen würden.

Der dritte Testteil wurde, wie erwähnt, mit dem Anspruch durchgeführt, durch ein adaptives Verhalten der Interlokutoren das „wahre“ Kompetenzniveau der Schüler/innen zu Tage zu fördern. Nach Ansicht der Beteiligten ist dies in vielen Fällen wohl auch gelungen, obschon sich die Interlokutoren, gerade weil sie viel Gestaltungsfreiheit hatten und flexibel sein mussten, natürlich nicht immer ideal verhielten. Es ist sicher auch so, dass der Faktor „Interlokutor“ auch in diesem Setting bis zu einem gewissen Grad verzerrend wirkte, weil verdeckt wirkende Verhaltensweisen wie Wohlwollen oder Kritik nie völlig kontrolliert werden können. Es kommt dazu, dass einige Schüler/innen – beispielsweise die Sechstklässler/innen aus Freiburg – kaum Mühe hatten, mit Sprechern der Zielsprache authentisch zu kommunizieren, weil sie über eine grosse Erfahrung verfügten, während andere offensichtlich auf keine Alltagserfahrungen (auch nicht aus der Schule) zurückgreifen konnten und deshalb emotional viel mehr herausgefordert waren.

Mehr als im dritten Testteil bekommt man bei den Leistungen im zweiten Testteil „*peer interaction*“ den Eindruck, dass immer wieder Schüler/innen nicht zeigen konnten oder zeigen wollten, was sie vermutlich könnten, wenn ihnen Gelegenheit dazu gegeben würde oder wenn sie explizit zum Handeln aufgefordert würden. Partnertests zur Interaktion oder auch nur inhaltliches Kommunizieren (anstelle von Rollenspielen oder Strukturübungen) mit Gleichaltrigen in der Fremdsprache war vielen Schüler/innen offensichtlich nicht vertraut. Unter anderem hatten relativ viele Schüler/innen die Tendenz sich eher zu verstecken als zu zeigen, sodass anschliessend (oft bei einem von beiden Schülern) nur sehr wenig Sprachmaterial vorhanden war, um Aussagen über die Kommunikationsfähigkeit machen zu können. Einen gewissen Anteil daran, dass in der *peer interaction* zum Teil wenig gesprochen wurde, hatten auch Aufgaben, bei denen Material manipuliert werden musste (v. a. Aufgabe IP1). Vieles konnte dort nonverbal oder mit wenigen, sich oft wiederholenden Worten einigermaßen erfolgreich erledigt werden. – Die *peer interaction* hat sich in den vergangenen Jahren in den bekannten internationalen Diplomprüfungen zu einem Standardelement der Prüfungen zum Sprechen entwickelt und dort auch grundsätzlich bewährt. Dies hat den positiven Washback-Effekt, dass interaktives Sprechen in der Zweierkonstellation zur Prüfungsvorbereitung tatsächlich auch geübt wird. Gerade im Unterricht mit Klassen von 20 oder mehr Schüler/innen ist diese Sozialform eine der wenigen Möglichkeiten, wie Interaktion regelmässig praktiziert und auch evaluiert werden kann. Deshalb sollten die sicher zum Teil skeptisch stimmenden Erfahrungen in HarmoS nicht zum Schluss führen, dass diese Testform ungeeignet ist, sondern dass diese Form der Zusammenarbeit im mündlichen Fremdsprachenunterricht tatsächlich gefördert werden sollte.

Die Daten zur mündlichen Produktion konnten wie diejenigen zur *peer interaction* mangels Datenvolumen nicht skaliert werden. Im Zusammenhang mit dem Versuch, die Vorschläge für Basisstandards zu illustrieren, wurden aber insgesamt über 50 Schülerleistungen zur Mündlichen Produktion von Expert/innen aus dem Konsortium gemeinsam den Referenzniveaus zugeordnet<sup>70</sup>. In einem einfachen Vergleich dieser Zuordnungen mit den Niveauzuordnungen aufgrund von Testteil 3 zeichnet sich tendenziell ein Schwelleneffekt ab: Schwächere Schüler/innen im Niveaubereich A1.2/A2.1 konnten in der Interaktion mit dem muttersprachlichen Interlokutor eine bessere Performanz zeigen (bzw. eine höhere kommunikative Sprachkompetenz glaubhaft machen) als in der Mündlichen Produktion, wo sie oft stecken blieben und grosse Mühe hatten, ohne die aktive Unterstützung einer sprachlich kompetenteren Person voranzukommen.

Wie erwähnt, konnten die ausführlichen Beurteilungsinstrumente, die zur Kodierung der Leistungen in den Testteilen 1 und 2 eingesetzt wurden, mangels Daten nicht auf ihre statistischen Eigenschaften hin geprüft und skaliert werden. Im Verlauf der sehr guten und intensiven Zusammenarbeit mit den Ratern konnten aber immerhin zahlreiche Formulierungen optimiert sowie Rückmel-

---

<sup>69</sup> Klasse 8 (heute 6. Klasse), 1. Fremdsprache: A2.1; 6; Klasse 8, 2. Fremdsprache: A1.2; Klasse 11, 1. oder 2. Fremdsprache: A2.2.

<sup>70</sup> Insgesamt erwies es sich als schwierig, Schüler/innen besonders der untersten Niveaus aufgrund ihres Erfolgs (bzw. Scheiterns) bei einer einzigen Aufgabe zur mündlichen Produktion mit einiger Sicherheit einem der Referenzniveaus zuzuordnen. Ein ausgesprochen auf Fehlervermeidung ausgerichteter „Kommunikationsstil“ bei einer Reihe von schwächeren Schüler/innen (v. a. Sechstklässler/innen aus der Westschweiz) machte die Aufgabe besonders schwierig bzw. eine faire Beurteilung wohl unmöglich.

dungen aller Art zu den Kriterien gesammelt werden. Bei der Beschreibung der Kompetenzniveaus sind Elemente dieser Formulierungen eingeflossen, sofern damit Phänomene charakterisiert werden, die wiederholt in Leistungen der entsprechenden Niveaus beobachtet werden konnten.

### **3.7.6 Ausblick: Überprüfung der Sprechkompetenz im Bildungsmonitoring**

Das Sprechen gehört in der Volksschule zweifellos zu den zentralen kommunikativen Sprachkompetenzen. Gleichzeitig sind herkömmliche Tests zum Sprechen – wie die eben besprochenen – in der Durchführung unvergleichlich viel aufwändiger als es vor allem Tests zu den rezeptiven Sprachfertigkeiten sind. Trotzdem wäre es sicher ein problematisches Signal, wenn das Sprechen aus Kostengründen in einem künftigen Bildungsmonitoring keinen Platz hätte.

Das IQB in Berlin arbeitet im Auftrag der deutschen KMK an computerbasierten Tests zum Sprechen. Angestrebt wird bei computerbasierten Tests ein Verfahren, bei dem Durchführung und Rating vollkommen automatisch ablaufen und deshalb sehr billig sind. Der Preis dafür sind in der Regel Validitätsprobleme in dem Sinn, dass Gespräche bis zu einem gewissen Grad auf Abfolgen von Computerprompts und Schülerreaktionen reduziert werden. Die Anforderungen liegen eher im Bereich der sprachlichen Mittel als der Interaktionsfähigkeit. Das Potenzial von computerbasierten Tests zum Sprechen hängt stark vom Stand der Technik im Bereich der Verarbeitung von natürlicher Sprache (*natural language processing*) ab, und zwar sowohl was die Authentizität und Natürlichkeit der Interaktion als auch was die Verlässlichkeit der Beurteilung betrifft. Die Fortschritte in diesem Bereich sind beträchtlich, die Berichte über computerbasierte Tests zum Sprechen (und auch zum Schreiben) sind nicht mehr nur negativ, sodass heute ein Potenzial vorhanden ist, das zumindest gründlich geprüft werden sollte, zumal die Alternative, ein Bildungsmonitoring ohne Tests zum Sprechen, keine gute ist.

Über das reine Bildungsmonitoring hinaus ist zu überlegen, inwieweit andere Mittel der Qualitätssicherung das Bildungsmonitoring etwas entlasten und zu einer Förderung der mündlichen Sprachkompetenz beitragen könnten. Ein eher informeller Ansatz könnte im Zurverfügungstellen und in der aktiven Einführung von Beurteilungsinstrumenten wie *lingualevel* bestehen, die es im Klassenrahmen ermöglichen, die Leistungen der Schüler/innen mit externen, vergleichbaren Massstäben zu messen. Eine weitere Möglichkeit wäre, in den öffentlichen Schulen Sprachzertifikate anzubieten. Im Rahmen von Zertifikatsprüfungen sind gut ausgebaute mündliche Prüfungsteile längst üblich und offenbar auch finanzierbar<sup>71</sup>. Ein gezielteres informelles Testen, ergänzt durch ein Angebot an Zertifikatsprüfungen mit ausführlichen mündlichen Teilen, würde es bis zu einem gewissen Grad erlauben, im Rahmen des Bildungsmonitorings auf Maximallösungen zu verzichten.

## **4 Kompetenzbereiche, die nicht empirisch untersucht wurden**

### **4.1 Sprachmittlung**

Der *Referenzrahmen* unterscheidet neben der schriftlichen und mündlichen Rezeption, Produktion und Interaktion als einen vierten Bereich die mündliche und schriftliche Sprachmittlung. Die mündlichen und schriftlichen Aktivitäten der Sprachmittlung ermöglichen die Kommunikation zwischen Menschen, die nicht direkt in derselben Sprache kommunizieren. Zu den sprachmittelnden Aktivitäten gehören Dolmetschen und Übersetzen, wobei in unserem Zusammenhang nicht an ein wörtliches Dolmetschen oder Übersetzen zu denken ist, sowie das zusammenfassende und paraphrasierende Berichten von Texten in derselben Sprache oder in einer anderen Sprache. In Sprachmittlungsaktivitäten werden Mitteilungen, die man in einer Sprache gehört oder gelesen hat, für jemanden, der die verwendete Sprache nicht oder nicht genügend versteht, in einer anderen für ihn verständlichen Sprache mündlich oder schriftlich sinngemäss weitergegeben.

---

<sup>71</sup> Im Bereich der öffentlichen Schule ist es allerdings sehr problematisch, wenn wesentliche Massnahmen der Qualitätssicherung in den Bereich der Freiwilligkeit bzw. der Privatwirtschaft verschoben und durch die Eltern finanziert werden müssen.

Der *Referenzrahmen* betont: „Sprachmittelnde Aktivitäten, also die Umformung eines schon vorhandenen Textes, nehmen eine wichtige Stellung im alltäglichen sprachlichen Funktionieren unserer Gesellschaften ein.“ (GER, 26). Konsequenterweise wird postuliert, dass sprachmittelnde Aktivitäten auch in die Beurteilung und Zertifizierung aufgenommen werden sollten. Die Beurteilung würde sich nicht auf die Kompetenz in *einer* Sprache beziehen, sondern auf die Fähigkeit, mit verschiedenen Sprachen, mit einem mehrsprachigen Repertoire umzugehen (GER, 169).

In Übereinstimmung mit dem *Referenzrahmen* hat das Fremdsprachenkonsortium die Sprachmittlung als einen der Kompetenzbereiche in das Kompetenzmodell aufgenommen. Dies geht eigentlich über den erhaltenen Auftrag hinaus. Denn die Entwicklung von Basisstandards im Projekt HarmoS ist auf Einzelfächer hin angelegt. Der Auftrag bestand – leider, aber im gegebenen Zeitrahmen realistischerweise – darin, ein Kompetenzmodell und Standards für die Fächer Französisch, Deutsch, Englisch, d. h. für Einzelsprachen zu erarbeiten, nicht aber ein Modell, das auf Mehrsprachigkeitskompetenz hin ausgelegt ist.

Sprachmittlung ist bisher nur in wenigen Lehrplänen erwähnt und sprachmittelnde Aktivitäten werden gegenwärtig im Fremdsprachenunterricht kaum eine Rolle spielen. Man muss aber davon ausgehen, dass die Sprachmittlung einen grösseren Stellenwert erhält, wenn im Rahmen des Szenarios 2010/12 das gleichzeitige Erlernen von zwei Fremdsprachen und – hoffentlich – eine entsprechende integrierte Sprachendidaktik umgesetzt werden.

Im Hinblick darauf wollten wir zumindest ansatzweise und exemplarisch zeigen, in welche Richtung die Formulierung von Standards gehen könnte.

Empirische Untersuchungen zur Sprachmittlung konnten in diesem Projekt nicht durchgeführt werden. Tests zu den Sprachmittlungs Kompetenzen wären im gegenwärtigen Zeitpunkt auch wenig aussagekräftig gewesen, da Sprachmittlung zurzeit im Fremdsprachenunterricht kaum geübt oder praktiziert wird.

Die Niveaubeschreibungen, die in das Papier zu den Basisstandards aufgenommen wurden, beruhen also – wie bisher alle Niveaubeschreibungen zur Sprachmittlung – nicht auf empirischen Befunden, die aus spezifische Überprüfungsaufgaben zur Sprachmittlung gewonnen wären, sondern sind aus anderen Quellen abgeleitet, die im Folgenden kurz genannt werden.

Der *Referenzrahmen* enthält keine Kompetenzbeschreibungen für Sprachmittlungsaktivitäten. In den von der deutschen Kultusministerkonferenz (KMK) herausgegebenen Bildungsstandards und auch in Bildungsstandards und Kerncurricula verschiedener deutscher Bundesländer sind zwar einige wenige Kannbeschreibungen für die Sprachmittlung enthalten, doch sind diese vielfach sehr allgemein. Eine recht detaillierte systematische Sammlung von Kannbeschreibungen zur Sprachmittlung liegt in *Profile deutsch* vor (Glaboniat et al. 2005). Zusätzlich zu den Kannbeschreibungen finden sich dort jeweils Beispiele für entsprechende Sprachmittlungsaktivitäten in verschiedenen Lebensbereichen (von Erwachsenen), insbesondere im privaten und öffentlichen Bereich sowie in den Bereichen Ausbildung und Beruf. Eine Umsetzung der Kannbeschreibungen und/oder Beispiele in konkrete Aufgaben gibt es in *Profile deutsch* nicht. Zudem wurden die Kannbeschreibungen selbst bisher noch nicht empirisch validiert.

Die Kannbeschreibungen dieser Quellen konnten als Anregung genutzt, nicht aber übernommen werden, diejenigen der KMK-Bildungsstandards nicht, weil sie zu vage und zu wenig niveauspezifisch sind, diejenigen der umfassenden Sammlung von *Profile deutsch* nicht, weil sie eher auf Erwachsene zugeschnitten und die dort gewählte Systematik zu recht komplizierten Formulierungen geführt hat.

Wir sind bei der Auswahl und Formulierung der Sprachmittlungsaktivitäten folgendermassen vorgegangen.

In einem ersten Schritt haben wir das Feld eingeschränkt und den Bereich der Sprachmittlung innerhalb ein und derselben Sprache wie das vereinfachende Zusammenfassen eines komplizierten Textes, z. B. das Paraphrasieren von Fachtexten für Laien, ausgeklammert. Wir beschränken uns also auf Sprachmittlung erstens zwischen einer Fremdsprache und der lokalen Schulsprache und zweitens zwischen einer ersten und einer zweiten Fremdsprache.

In einem zweiten Schritt wurden innerhalb des so eingeschränkten Feldes Typen von Sprachmittlungs-Aktivitäten identifiziert. Sie ergeben sich aus der Kombination der Faktoren: 1. beteiligte

Sprachen; 2. Mündlichkeit – Schriftlichkeit; 3. Art der Kommunikationssituation: direkte (*face to face*) vs. zeitverschobene Kommunikation.

Ohne Anspruch auf Vollständigkeit wurden die folgenden Konstellationen ausgewählt, die als relevant für die Zielgruppe eingeschätzt werden:

- a) Sprachmittlung mündlich ↔ mündlich: wechselseitig aus der Fremdsprache in die lokale Schulsprache und aus der lokalen Schulsprache in die Fremdsprache (informelles Dolmetschen);
- b) Sprachmittlung mündlich ↔ mündlich: wechselseitig aus einer Fremdsprache in eine zweite Fremdsprache (informelles Dolmetschen);
- c) Sprachmittlung mündlich → mündlich: aus der lokalen Schulsprache in die Fremdsprache;
- d) Sprachmittlung mündlich → schriftlich: aus der lokalen Schulsprache in die Fremdsprache;
- e) Sprachmittlung mündlich → mündlich/schriftlich: aus der Fremdsprache in die lokale Schulsprache;
- f) Sprachmittlung schriftlich → mündlich/schriftlich: aus der Fremdsprache in die lokale Schulsprache;
- g) Sprachmittlung schriftlich → schriftlich: aus einer Fremdsprache in die andere Fremdsprache.

In einem dritten Schritt wurden Kannbeschreibungen formuliert, die sich hinsichtlich des Handlungspotenzials, das jemand bei Tätigkeiten der Sprachmittlung hat, an den Kompetenzen orientieren, über die jemand in den jeweils beteiligten rezeptiven, produktiven und interaktiven Sprachfertigkeiten verfügt. Je nachdem ob etwas in mündlicher oder schriftlicher Form weitervermittelt wird, fallen die Anforderungen anders aus, da die Basisstandards für den Bereich des Sprechens höher angesetzt wurden als für das Schreiben. In Fällen, wo etwas in einer Fremdsprache gehört oder gelesen und in der lokalen Schulsprache weitergegeben wird, gelten die Basisstandards für Hörverstehen bzw. Leseverstehen in der betroffenen Fremdsprache.

Die Formulierungen der Kannbeschreibungen zur Sprachmittlung übernehmen daher wörtlich Elemente aus den Kannbeschreibungen der anderen Kompetenzbereiche, um niveautypische Präzisierungen und Einschränkungen anzugeben (z. B. „vertraute Themen“, „mit einfachen Sätzen und Wendungen“ usw.).

Die (fettgedruckten) Kannbeschreibungen werden jeweils durch einzelne Beispiele für Sprachmittlungsaktivitäten in Alltags- und Schulsituationen etwas konkretisiert.

Was die Qualität der mündlichen bzw. schriftlichen Äusserungen in der Fremdsprache angeht, könnte man ergänzend die Beschreibungen zum verfügbaren Spektrum sprachlicher Mittel, zu Flüssigkeit und Korrektheit heranziehen, die für das Sprechen und Schreiben für den jeweiligen Basisstandard vermerkt sind. Allerdings hat die Anlehnung an Kompetenzbeschreibungen der rezeptiven und produktiven mündlichen und schriftlichen Einzelfertigkeiten hypothetischen Charakter. Es müsste in Nachfolgeuntersuchungen empirisch überprüft werden, ob und inwieweit z. B. die begrenzte Verarbeitungskapazität beim gleichzeitigen Umgang mit zwei Sprachen oder die Anforderung, die Bedürfnisse der Kommunikationspartner abzuwägen, in den verschiedenen Konstellationen der Sprachmittlung noch gleichartige Leistungen ermöglicht, wie sie für die jeweiligen Einzelfertigkeiten beschrieben sind.

Die exemplarisch für die Klasse 11 formulierten Kannbeschreibungen und Beispiele für Sprachmittlungsaktivitäten sollen einerseits Anregungen geben für die stärkere Berücksichtigung dieses Kompetenzbereichs in Lehrplänen und Lehrwerken. Sie können andererseits als Ausgangspunkt für die Entwicklung von Aufgaben genutzt werden. Denn wenn einmal das Szenario 2010/2012 für die Reform des Fremdsprachenunterrichts realisiert ist und eine integrierte Sprachendidaktik umgesetzt wird, müssten in einer Folgeuntersuchung zum HarMoS-Projekt Testaufgaben zu relevanten Konstellationen der Sprachmittlung entwickelt und aufgabenbasiert auf empirischer Basis Standards für diesen Kompetenzbereich formuliert werden.

## 4.2 Überfachliche Kompetenzen

### 4.2.1 Allgemeines

Im Kompetenzmodell Fremdsprachen sind - ebenso wie im zugrundeliegenden Kompetenzmodell des *Referenzrahmens* – interkulturelle und methodische Kompetenzen als Komponenten der „Allgemeinen Kompetenzen“ enthalten; die Bedeutung dieser Kompetenzbereiche wird ausdrücklich hervorhoben, aber es fehlen ausformulierte Kompetenzbeschreibungen und Skalen für diese wichtigen Bereiche. Der Referenzrahmen ist deshalb immer wieder kritisiert worden und die Autoren sowohl des *Referenzrahmens* als auch von Folgeprojekten im Umkreis des *Referenzrahmens* sind sich dieser Lücken bewusst. So betont z. B. Daniel Coste: „Tout ce qui touche à la compétence (inter)culturelle reste peu développé dans le *Cadre* et demande là aussi, dans bien des contextes, des descripteurs, voire des types de descripteurs, autres que ceux dont on dispose déjà.“ (Coste 2007, 51). Es ist verständlich, dass manche Praktiker, vor allem Lehrplanentwickler und Lehrkräfte, von *HarmoS Fremdsprachen* auch Vorschläge für Standards zu diesen wichtigen Bildungsbereichen erwarten. Doch kann diese Erwartung aus mehrfachen Gründen im Rahmen des Projekts *HarmoS Fremdsprachen* nicht erfüllt werden:

Von Beginn an hat das Konsortium Fremdsprachen darauf hingewiesen, dass noch viel grundlegende Arbeit geleistet werden muss, um zu theoretisch und empirisch abgestützten Kompetenzbestimmungen und -stufen zu gelangen und dass dazu zusätzliche Forschungsprojekte nötig wären (z. B. im Rahmen des Nationalen Forschungsprogramms NFP 56 zur „Sprachenvielfalt und Sprachkompetenz in der Schweiz“). Solche Projekte sind jedoch nicht zustande gekommen.

HarmoS ist, der Klieme-Expertise folgend, auf die Kompetenz in Einzelfächern, d. h. in unserem Fall auf die Kompetenz in einzelnen Fremdsprachen, hin angelegt, nicht aber auf fächerübergreifende Kompetenzen. Interkulturelle Kompetenzen, Methodenkompetenz und Sprachbewusstheit stehen in einer engen Beziehung zu Fremdsprachenfächern, aber sie werden nicht nur in einem einzelnen Fremdsprachenfach und nicht nur in den Fremdsprachenfächern ausgebildet.

Im engen Zeitplan des HarmoS-Projekts war es auch nicht möglich, Aufgaben oder auch Fragebogen etwa zu Haltungen, Problemlösungsverhalten, kulturellem Wissen zu erproben. Unserer Auffassung nach wäre es problematisch, ohne empirische Grundlagen gestufte Kompetenzbeschreibungen formulieren und Basisstandards vorschlagen zu wollen. Man würde riskieren, dass der unterschiedliche Status von Beschreibungen für die kommunikative Sprachkompetenz auf der einen und der interkulturellen und methodischen Kompetenzen auf der anderen Seite in der Praxis nicht genügend beachtet würde.

Wir sehen diese Sicht bestätigt, wenn es beispielsweise im Handbuch zu den österreichischen Bildungsstandards, die eine Liste von Fähigkeiten dieser Kompetenzbereiche enthalten, heisst: „Deskriptoren zu den dynamischen Fähigkeitsbereichen *kommunikative, interkulturelle, soziale Kompetenz* und *Lernstrategien* lassen sich nur sehr schwer verschiedenen Niveaus zuordnen und wurden deshalb zumindest vorläufig stufenübergreifend eingetragen<sup>72</sup>“. Und weiter: „Die Deskriptoren der *dynamischen Fähigkeiten* laufen Gefahr – und dessen sind sich die Autoren/innen sehr wohl bewusst – in überzogene Ansprüche an interkulturelle, soziale und lernstrategische "Fitness" abzugleiten. Dennoch haben sie ihre Berechtigung: Nicht als unbedingt zu erreichende Unterrichtsziele, sondern als Denk- und Handlungsansätze, die zwar nicht wirklich kategorisierbar und nach Schwierigkeitsgraden messbar sind, aber besonderer Aufmerksamkeit und Übung bedürfen, um eine reale Chance zu erhalten, zur intellektuellen bzw. emotiven Reifung und Persönlichkeitsbildung beizutragen“ (Buchberger et al. 2002, 19). Offensichtlich handelt es sich hier also nicht mehr um die Definition von Standards weder im Sinne von Basis noch von Regelstandards, sondern um die Formulierung von Lernzielen.

Wie die österreichischen enthalten auch die deutschen Bildungsstandards (beide sind Regelstandards) Beschreibungen der interkulturellen und methodischen Kompetenzen, die jedoch aus mehreren Gründen nicht unproblematisch sind: Sie sind nicht theoretisch mit Bezug auf ein Kompetenzmodell begründet; sie sind nicht oder kaum stufenbezogen formuliert; sie werden nicht durch Aufgabenbeispiele illustriert; sie mischen ganz generelle, für alle Fächer geltende Bildungsziele z.

---

<sup>72</sup> Unter dem missverständlichen Titel „kommunikative Kompetenz“ sind in der Liste der Deskriptoren zu den dynamischen Fähigkeiten der österreichischen Bildungsstandards kommunikationssichernde Strategien wie Nachfragen oder Achten auf nonverbale Signale aufgeführt.

B. „selbstständig und kooperativ arbeiten“ (KMK) oder „Kann in einer Gruppe arbeiten und dort eigene Ideen einbringen“ (Gassner 2005) mit Zielen, die enger mit den Fremdsprachen verbunden sind, z. B. „können sich in Bezug auf die Befindlichkeiten und Denkweisen in den fremdkulturellen Partner hineinversetzen“ (KMK) oder „Weiß, dass auch alltägliche Umgangsformen und Vorschriften wie Verkehrsregeln, Tischsitten oder Körpersprache kulturabhängig und damit grundsätzlich gleichwertig sind“ (Gassner 2005). Die Auswahl erscheint (für Standards) manchmal etwas beliebig und würde wahrscheinlich von Fachdidaktikern nicht unbedingt übereinstimmend hoch gewichtet werden, z. B. „Macht Nachsprechübungen und imitiert dabei Intonation und Aussprache des Sprachmodells“ oder „Versucht manchmal zum Spaß, Selbstgespräche auch in der Fremdsprache zu führen“ (Gassner 2005).

An einem Expertentreffen (Beiratsmitglieder und zusätzlich eingeladene Fachdidaktiker/innen) haben wir rund 50 Deskriptoren zu den interkulturellen und methodischen Kompetenzen, die aus verschiedenen Quellen (darunter auch die deutschen und österreichischen Bildungsstandards) stammten, gewichten, beurteilen und kommentieren lassen. Dabei zeigte sich erstens, dass bei über der Hälfte Deskriptoren recht unterschiedliche Prioritäten gesetzt wurden, zweitens, dass nur selten eine deutliche Zuordnung zu den Klassenstufen 8 oder 11 sowie zur ersten und zweiten Fremdsprache möglich schien, und drittens, dass die Expertinnen und Experten in ihren Kommentaren vielfach skeptisch auf den idealistischen Charakter und auf die Nicht-Überprüfbarkeit von solchen Kompetenzen hinwiesen.

Aus all diesen Gründen haben wir in die Papiere zum Kompetenzmodell und zu den Basisstandards keine Beschreibungen interkultureller und methodischer und anderer im *Referenzrahmen* als „allgemeine Kompetenzen“ verorteter Dispositionen und Fertigkeiten aufgenommen, sondern durch eine Anmerkung signalisiert, dass es sich bei den interkulturellen und methodischen Kompetenzen um zentrale Bildungsziele der Schule handelt, die in den regionalen Lehrplänen den entsprechenden Stellenwert erhalten müssen. Zudem plädieren wir dafür, angesichts des gegenwärtigen Kenntnisstands auf entsprechende Tests im Rahmen des Bildungsmonitorings zu verzichten und ganz auf die (formative) Evaluation durch das Sprachenportfolio zu setzen. Denn das Sprachenportfolio ermöglicht eine individuelle, reflexive Auswertung sowohl von kulturellen und interkulturellen Erfahrungen als auch von Sprachlernerfahrungen und Erfahrungen mit sprachlicher Vielfalt. Diese Position wird auch durch einschlägige Literatur bestätigt, in der die zentrale Rolle der Selbstbeurteilung und des Portfolioansatzes verschiedentlich hervorgehoben wird. So schreibt Vollmer (2005, 267): „Für die Bewertung von *interkulturellen Kompetenzen* scheint mir ohnehin die Selbstbeurteilung die primäre oder gar die einzig taugliche Form der Evaluation zu sein;“ (vgl. z. B. Beacco 2004, 251; Byram/ Gribkova/ Starkey 2002, 33; Lázár et al. 2007, 29; Scarino 2007, 5).

Wir haben uns für unsere Entscheidung auf den gegenwärtigen Forschungsstand berufen, der im Folgenden zusammenfassend skizziert wird, um Hinweise zu geben zum einen im Hinblick auf die Entwicklung von Lehrplänen und Lehrwerken, zum andern aber auch im Hinblick auf Folgeprojekte zu HarmoS. Die Forschungs- und Entwicklungsarbeit in diesen Bereichen ist in Bewegung. Es ist zu hoffen, dass sowohl in Bezug auf die Konzeptualisierung als auch in Bezug auf Evaluierungsmöglichkeiten Fortschritte erzielt werden. In Folgeprojekten zu HarmoS sollten die hier besprochenen Kompetenzbereiche besondere Beachtung finden und es sollten für die notwendigen Untersuchungen auch die entsprechenden Ressourcen bereitgestellt werden. Anders als in den gegenwärtigen Projekten müsste gezielt eine enge Zusammenarbeit der Konsortien verschiedener Fachbereiche eingeplant werden, insbesondere eine enge Koordination zwischen Projekten für die Schulsprachen (L1) und die Fremdsprachen (L2/L3). Dies bedeutet, dass ermöglicht werden sollte, von einem Konzept plurilingualer und plurikultureller Kompetenz auszugehen und die Standards unter diesem Gesichtspunkt und im Hinblick auf eine integrierte Sprachendidaktik zu überarbeiten.

#### **4.2.2 Kulturelle und interkulturelle Kompetenzen**

Der Bereich der kulturellen und interkulturellen Kompetenzen ist sowohl, was die Komponenten oder Dimensionen als auch was die Entwicklung und Abstufung anbelangt, unterschiedlich konzeptualisiert worden.

##### **Komponenten/Dimensionen der (inter)kulturellen Kompetenz**

Im *Referenzrahmen* werden – weitgehend den Vorarbeiten von Byram und anderen folgend – die interkulturellen Kompetenzen als Teil der allgemeinen Kompetenzen gesehen und vier Wissens- bzw. Könnenskategorien zugeordnet:

1. *Deklaratives Wissen (savoir)*: Dazu gehören neben dem allgemeinen Wissen a) das soziokulturelle Wissen über die jeweilige Zielkultur, z. B. charakteristische, kulturspezifische Merkmale des täglichen Lebens, der zwischenmenschlichen Beziehungen, Wertsysteme, soziale Konventionen, Körpersprache u. Ä. und b) das interkulturelle Bewusstsein, z. B. das Wissen um Ähnlichkeiten und Unterschiede oder auch um stereotype Sichtweisen.

2. *Fertigkeiten und prozedurales Wissen (savoir-faire)*: Dazu zählen neben praktischen Fertigkeiten die interkulturellen Fertigkeiten, nämlich:

- die Fähigkeit, die Ausgangskultur und die fremde Kultur miteinander in Beziehung zu setzen;
- kulturelle Sensibilität und die Fähigkeit, eine Reihe verschiedener Strategien für den Kontakt mit Angehörigen anderer Kulturen zu identifizieren und zu verwenden;
- die Fähigkeit, als kultureller Mittler zwischen der eigenen und der fremden Kultur zu agieren und wirksam mit interkulturellen Missverständnissen und Konfliktsituationen umzugehen;
- die Fähigkeit, stereotype Beziehungen zu überwinden (Europarat 2001, 106).

3. *Persönlichkeitsbezogene Kompetenz (savoir-être)*: Neben Motivationen, kognitiven Stilen und anderen Persönlichkeitsfaktoren, sind hier besonders die Einstellungen in Bezug auf Fremdes und Neues bedeutsam, nämlich:

- Offenheit für und Interesse an neuen Erfahrungen, anderen Menschen, Ideen, Völkern, Gesellschaften und Kulturen;
- Bereitschaft, die eigene kulturelle Sichtweise und das eigene kulturelle Wertesystem zu relativieren;
- Bereitschaft und Fähigkeit, sich von konventionellen Einstellungen gegenüber kulturellen Unterschieden zu distanzieren (Europarat 2001, 107).

4. *Lernfähigkeit (savoir-apprendre)*: Hier werden erstens allgemeine Lern- und Arbeitstechniken aufgeführt wie „in Partner- oder Gruppenarbeit effektiv zusammenarbeiten“, zweitens Lerntechniken und heuristische Fertigkeiten, die sich auf das Sprachenlernen im engeren Sinn beziehen, und drittens solche mit Bezug auf neue (inter)kulturelle Erfahrungen:

- die Fähigkeit, aus direkter Beobachtung und Teilnahme an Kommunikationsereignissen effektiv zu lernen (sowohl linguistisch als auch soziokulturell), indem man perzeptuelle, analytische und heuristische Fertigkeiten ausbildet; (...)
- die Fähigkeit der Lernenden, mit neuen Erfahrungen umzugehen (mit einer neuen Sprache, mit neuen Menschen, neuen Verhaltensweisen usw.) und in einer Lernsituation andere Kompetenzen einzusetzen (z. B. durch Beobachten, Erfassen der Bedeutung des Beobachteten, Analysieren, Schlüsse ziehen, Memorieren usw.) (Europarat 2001, 109).

Wie erwähnt, enthält der *Referenzrahmen* keine Aussagen zur Entwicklung oder Stufung dieser Kompetenzbereiche. Die interkulturellen Kompetenzen erscheinen im *Referenzrahmen* grundsätzlich auch nicht an sprachliche Kompetenzen eines bestimmten Niveaus gebunden.

Byram (1997) differenziert die *Savoir*-Typen etwas anders und nennt ergänzend zum *savoir* als dem eigentlichen Wissen ein *savoir-comprendre*: das Verstehen unter Nutzung des kulturellen Wissens. Er fügt ein *savoir s'engager* hinzu, d. h. die kritische Urteilsfähigkeit in Bezug auf Eigen- und Fremdkulturelles.<sup>73</sup>

Die Unterscheidung der Dimensionen des Wissens, der Einstellungen, der Fertigkeiten und der Bewusstheit findet sich auch in vielen anderen Modellierungen der interkulturellen Kompetenz. Vielfach werden mehrere der folgenden Merkmale aufgeführt: Respekt, Empathie, Flexibilität, Geduld, Interesse, Neugier, Offenheit, Bereitschaft, Sinn für Humor, Ambiguitätstoleranz, Zurückhaltung im Beurteilen u. a. (Eine gute Übersicht verschiedener Modellierungen findet sich in Humphrey 2007 sowie in Sinicrope/ Norris/ Watanabe 2007). Bolten (2007) weist darauf hin, „dass immer wieder als signifikant beschriebene Merkmale interkultureller Kompetenz wie ‚Empathie‘,

---

<sup>73</sup> Das Modell von Byram (1997) wurde im EU-Projekt LACE zusammen mit dem Modell von Chen/Starosta (1998) und dem (im deutschen Projektbericht seltsamerweise immer als „Bezugsrahmen“, „GEBR“, zitierten) Referenzrahmen für die Analyse von Lehrplänen und Unterrichtspraxis in Bezug auf interkulturelle Ziele verwendet (Franklin et al. 2007). Trotz methodischer Mängel der Studie dürften die Grundergebnisse der Länderuntersuchung wohl – auch für die nicht beteiligte Schweiz – zutreffen: Der Stellenwert interkultureller Ziele ist in Lehrplänen und Unterrichtspraxis relativ gering. Die Aspekte des Wissens und der Einstellung finden am ehesten Beachtung; interkulturelle Handlungskompetenz dagegen recht selten.

‚Rollendistanz‘, ‚Ambiguitätstoleranz‘, ‚(Meta)kommunikationsfähigkeit‘ usw. auch über den Erfolg oder Nicht-Erfolg von Handlungen entscheiden, die nicht durch kulturelle Überschneidungssituationen gekennzeichnet sind“. Im Hinblick auf die Förderung interkultureller Kompetenz als Querschnittsaufgabe der Schule gibt er zu bedenken, dass beispielsweise Einfühlungsvermögen, Rollendistanz oder Flexibilität auch intrakulturell unabdingbare Voraussetzungen“ sind, „um soziales Handeln erfolgreich bewältigen zu können. Vor diesem Hintergrund erscheint es sinnvoll, interkulturelle Kompetenz nicht als einen eigenständigen Kompetenzbereich zu verstehen, sondern vielmehr als Fähigkeit, individuelle, soziale, fachliche und strategische Teilkompetenzen in ihrer bestmöglichen Verknüpfung auf interkulturelle Handlungskontexte beziehen zu können“ (Bolten 2001).

Im europäischen Leonardo-Projekt INCA (= Intercultural Competence Assessment), das auf dem *Referenzrahmen* aufbaut und an dem auch Michael Byram federführend beteiligt war, werden sechs Kompetenzelemente unterschieden, nämlich: Empathie, Offenheit gegenüber anderen Kulturen, Wissenserwerb, Kommunikationsbewusstsein, Ambiguitätstoleranz, Verhaltensflexibilität. Diese sechs Elemente werden zur besseren Verständlichkeit für die Benutzer zu den drei „Säulen“ interkultureller Kompetenz *Offenheit*, *Wissen* und *Anpassungsfähigkeit* zusammengefasst:

*Offenheit* umfasst *Respekt vor anderen Kulturen* und *Ambiguitätstoleranz*.

*Wissen* umfasst *Wissenserwerb* und *Empathie*.

*Anpassungsfähigkeit* umfasst *Verhaltensflexibilität* und *Kommunikationsbewusstsein*. (INCA 2004)

Im Sokrates-Projekt LOLIPOP wurde ein Online-Sprachenportfolio erarbeitet, das den Anspruch erhebt, einem besonderen Aspekt auf das interkulturelle Lernen zu legen. Dort werden die drei Dimensionen Wissen (savoir), Fertigkeiten (savoir-faire) und Einstellungen (savoir-être) unterschieden und durch – allerdings nicht empirisch abgesicherte – Deskriptoren zu diesen Dimensionen beschrieben.

Die drei Savoir-Typen bilden auch die Dimension des Konzepts interkultureller Kompetenz, das dem Projekt des Europäischen Fremdsprachenzentrums in Graz zur Entwicklung und Evaluierung interkultureller kommunikativer Kompetenz zu Grunde liegt (Lázár/ Huber-Kriegler/ Lussier/ Matei/ Peck 2007).

Jean-Claude Beacco, Berater der Abteilung Sprachpolitik des Europarats, nimmt in seinem Vorschlag für ein „Référentiel pour les compétences culturelles dans l’enseignement des langues“ eine Einteilung in die Dimensionen Wissen, Einstellungen, (Interaktions-)Situationen und Diskursformen vor. Er konzentriert sich auf die (sozio-)kulturellen Kompetenzen und versucht, sie möglichst in Nähe zu kommunikativen sprachlichen Kompetenzen zu beschreiben. Er unterscheidet - nicht immer leicht nachvollziehbar - folgende Komponenten:

les **dimensions actionnelles** (dites aussi sociopragmatiques) : elles permettent de décrire un « savoir agir » minimum dans une communauté peu connue ou inconnue, de manière à pouvoir y gérer sa vie matérielle et relationnelle (...)

les **dimensions ethnolinguistiques** permettent de caractériser le « vivre ensemble verbal », comme ensemble de normes concernant les comportements communicatifs (relevant de l’ethnologie de la communication) et qui ont des effets directs sur la réussite de la communication. (...)

les **dimensions relationnelles** concernent les attitudes et les compétences verbales nécessaires à une gestion appropriée d’interactions portant explicitement, même partiellement, sur la matière culturelle et sociétale elle-même. Ces interactions ne relèvent pas uniquement d’un savoir-faire d’ordre interactionnel : elles constituent l’une des formes de la « communication verbale interculturelle » et impliquent la constitution ou le renforcement d’attitudes positives comme la curiosité, la bienveillance et la tolérance culturelle ;

les **dimensions interprétatives** concernent la manière dont les utilisateurs/apprenants donnent sens à des sociétés qui ne leur sont pas familières, dont ils ne partagent pas les normes et bien des référents et en rendent compte (pour eux-mêmes et à d’autres). (...)

les **dimensions éducatives ou interculturelles** concernent la matière culturelle en tant qu’elle est objet de jugements évaluatifs, susceptibles, en particulier, d’activer des réactions d’ethnocentrisme et d’intolérance ou des processus d’acculturation et de dépendance culturelle, d’enfermement identitaire ou d’aliénation. (Beacco 2004, 253-254).

Im Detail enthält der Vorschlag von Beacco eine ganze Reihe interessanter Anregungen, er bleibt jedoch auf einem hohen Abstraktionsniveau, was eine Umsetzung in die Praxis der Curriculum- oder Lehrplanentwicklung erschwert.

Das Problem theoretisch begründeter und zugleich plausibler Abgrenzungen und damit das Problem von Überschneidungen und Redundanzen, das sich im Vorschlag Beaccos zeigt, ist auch im Projekt CARAP evident und wird im Projektbericht mehrfach reflektiert. In diesem Projekt des Europäischen Fremdsprachenzentrum Graz wurde der Versuch unternommen, einen Referenzrahmen für plurale Zugänge zu Sprachen und Kulturen zu entwickeln (CARAP = Cadre de Référence pour les Approches Plurielles des Langues et des Cultures). Wie im *Referenzrahmen*, wo formelhaft immer wieder von der „mehrsprachigen und plurikulturellen Kompetenz“ gesprochen wird, versucht CARAP Mehrsprachigkeitskompetenz und interkulturelle Kompetenz zusammen zu sehen.

Es werden drei „Zonen“ unterschieden:

### 1. *La zone de la gestion de la communication linguistique et culturelle en contexte d'altérité*

Zu diesem Kompetenzbereich der Kommunikation im Kontext kultureller Heterogenität gehören:

- die Kompetenz zur Konfliktlösung, Hindernisbeseitigung, zur Klärung von Missverständnissen im interkulturellen Bereich;
- Verhandlungskompetenz in der interkulturellen Interaktion;
- die Kompetenz der Sprachmittlung;
- Adaptionskompetenz.

### 2. *La zone de construction et d'élargissement d'un répertoire linguistique et culturel pluriel*

Zu diesem Kompetenzbereich werden zwei sogenannte Mikro-Kompetenzen angeführt:

- die Kompetenz, die eigenen interkulturellen und plurilingualen Erfahrungen zu nutzen;
- die Kompetenz, bezogen auf interkulturelle Kontexte sowohl im schulischen als auch im ausserschulischen Rahmen systematische und kontrollierte Lernprozesse zu initiieren.

### 3. *Une zone intermédiaire*

Für diesen Mischbereich werden Mikro-Kompetenzen unterschieden, die als relevant für die beiden genannten Zonen angesehen werden, nämlich:

- Kompetenz zum Perspektivenwechsel;
- Kompetenz, dem sprachlich und/oder kulturell Nichtvertrauten einen Sinn zu geben
- Kompetenz der Distanzierung;
- Kompetenz, die eigene kommunikative Situation oder eine Lernsituation und die anhängigen Aktivitäten kritisch zu analysieren;
- Kompetenz, die fremde Alterität in ihren Unterschieden und Ähnlichkeiten zu erkennen (Candelier et al. 2007).

Die grosse Menge von – vorwiegend aus der Literatur abgeleiteten - Kompetenzbeschreibungen ist beeindruckend und CARAP kann von Praktikern gewinnbringend als Steinbruch verwendet werden. Ein Referenzrahmen für mehrsprachige und mehrkulturelle Kompetenz aber ist damit noch nicht geschaffen. Es bleibt weitgehend bei einer Materialsammlung. Daher wurde auch eine Weiterentwicklung im Rahmen des nächsten Entwicklungsprogramms des Europäischen Fremdsprachenzentrums in Graz beantragt.

Lehrpläne brauchen eine einfachere und überschaubare, für Lehrwerkautoren und Lehrpersonen leicht nachvollziehbare Aufteilung der Dimensionen interkultureller Kompetenz, wie sie in den Bildungsstandards der KMK zu finden ist. Dort werden analog zu den drei *Savoir*-Typen (*savoirs*, *savoir-être*, *savoir-faire*) die folgenden Bereiche unterschieden:

- Soziokulturelles Orientierungswissen
- Verständnisvoller Umgang mit kultureller Differenz
- Praktische Bewältigung interkultureller Begegnungssituationen. (KMK 2005, 8)

Zu recht wird dort auch betont, dass es sich bei der Entwicklung der Kompetenzen in diesen Bereichen um eine übergreifende Aufgabe der Schule handelt<sup>74</sup>.

Die Entwicklung interkultureller Kompetenzen ist eine übergreifende Aufgabe von Schule, zu der fremdsprachlicher Unterricht einen besonderen Beitrag leistet. Angesichts der zunehmenden persönlichen und medialen Erfahrung kultureller Vielfalt ist es Aufgabe auch des Fremdsprachenunterrichts, Schülerinnen und Schüler zu kommunikationsfähigen und damit offenen, toleranten und mündigen Bürgern in einem zusammenwachsenden Europa zu erziehen. Mit der Fähigkeit, eigene Sichtweisen, Wertvorstellungen und gesellschaftliche Zusammenhänge in ausgewählten Kontexten mit denen anderer Kulturen tolerant und kritisch zu vergleichen, und mit der Bereitschaft, Interesse und Verständnis für Denk- und Lebensweisen, Werte und Normen und die Lebensbedingungen der Menschen eines anderen Kulturkreises aufzubringen, erleben die Schülerinnen und Schüler einen Zuwachs an Erfahrung und Stärkung der eigenen Identität. Der Fremdsprachenunterricht trägt zu dieser Mehrperspektivität der Persönlichkeitsentwicklung vor allem bei durch Orientierungswissen zu exemplarischen Themen und Inhalten sowie durch den Aufbau von Fähigkeiten zur interkulturellen Kommunikation. (KMK 2005, 6)

Am ehesten spezifisch für das einzelne Fremdsprachenfach sind Ziele und Inhalte in Bezug auf das soziokulturelle und kulturelle Wissen. Dabei bleibt zu berücksichtigen:

- im Hinblick auf Ziele im Fach zweite Landessprache: dass es über die Sprachregionen hinweg viele Gemeinsamkeiten und gemeinsames soziokulturelles Wissen gibt (in Bezug auf Institutionen, Politik, Geschichte u. a.);
- dass sich mehr und mehr eine internationale Jugendkultur entwickelt (in der das Englische - und amerikanische Kultur - eine wichtige Rolle spielen);
- dass die Kultur, in der die Lernenden leben (Schule und Umwelt), sowie die Kultur der Zielsprache(n) nicht homogen und in sich abgegrenzt, sondern durchmischt und „transkulturell“ geprägt sind; diese „Hybridisierung“ von Kultur wird sich durch die zunehmende Globalisierung weiter verstärken.

Interkulturelle Handlungskompetenz sollte möglichst in Verbindung mit realen und virtuellen Austausch- und Begegnungsprojekten gefördert werden. Im Rahmen einer Austauschpädagogik macht auch die Orientierung an möglichen „Critical incidents“, an „Hotspots“ der interkulturellen Kommunikation und an „Hotwords“ Sinn (Heringer 2004). Auch hier ist aber nicht zu vergessen, dass interkulturelle Begegnungssituationen nicht unbedingt ausserhalb der eigenen Sprachregion gesucht werden müssen, sondern sich auch im Umfeld der Lernenden und im multikulturellen Klassenzimmer finden.

### **Niveaus der (inter-)kulturellen Kompetenz**

Der *Referenzrahmen* enthält, wie mehrfach vermerkt, keine Skalen zu soziokulturellen und interkulturellen Kompetenzen. Im Rahmen des Nationalen Forschungsprogramms NFP 33 war es nicht gelungen, Deskriptoren zu diesen Kompetenzbereichen zu skalieren (Schneider/North 2000, 89-90). Es liegen verschiedene Versuche vor, - unabhängig vom *Referenzrahmen* oder mit Bezug auf den *Referenzrahmen* - Stufen (inter-)kultureller Kompetenz zu unterscheiden.

Das einflussreiche Entwicklungsmodell von Milton Bennett, auf das oft mit dem Kürzel DMIS für „Developmental Model of Intercultural Sensivity“ Bezug genommen wird, unterscheidet unter Berücksichtigung von kognitiven, affektiven und Verhaltensaspekten sechs Niveaus<sup>75</sup>. Davon sind jeweils drei Niveaus durch ethnozentrische und drei durch ethnorelative Orientierungen bestimmt (Bennett 1993). Die Differenzierung in sechs Orientierungen und deren Abfolge liess sich empirisch nicht in allen Untersuchungen replizieren. Deshalb wurden im DESI-Projekt die sechs interkulturellen Orientierungen aus Bennetts Modell „als Klassifikationen interkultureller Kompetenz

---

<sup>74</sup> Vgl. „Beschluss der Kultusministerkonferenz vom 25.10.1996. In: Bundeszentrale für politische Bildung (Hg.), Interkulturelles Lernen. Bonn 1998: „Der interkulturelle Aspekt ist dabei nicht in einzelnen Themen, Fächern oder Projekten zu isolieren, sondern eine Querschnittsaufgabe in der Schule.“ Zur Problematik der Umsetzung des Postulats vgl. Bolten 2001.

<sup>75</sup> Wie der Name des Modells besagt, bezieht es sich auf interkulturelle Sensibilität. Diese wird definiert als „ability to discriminate and experience relevant cultural differences“ im Unterschied zur interkulturellen Kompetenz als „ability to think and act in interculturally appropriate way“ (Hammer, Bennet & Wiseman 2003).

verstanden ohne Annahmen über deren Reihenfolge zu machen“ (Hesse/Göbel 2007, 266). Ausserdem wurden wegen Überlappungen und „aus Gründen der Darstellungsökonomie“ die ursprünglich sechs Klassen zu drei Klassen interkultureller Kompetenz zusammengefasst.

Das vorsichtige Vorgehen in der DESI-Untersuchung zeigt sehr deutlich die Schwierigkeiten, Niveaustufen interkultureller Kompetenz empirisch abzustützen.

Weniger vorsichtig werden oft auf intuitiver Basis Niveaus interkultureller Kompetenz angesetzt. So wurden in manchen frühen Versionen des europäischen Sprachenportfolios und neu auch im Online-Portfolio LOLIPOP Beschreibungen interkultureller Kompetenzen der Typen *savoir*, *savoir-comprendre*, *savoir-faire*, *savoir-être* und *savoir-apprendre* intuitiv und teilweise in problematischer Weise den Sprachkompetenzniveaus A1-C2 zugeordnet (zur Diskussion dieser Problematik siehe Schneider/Lenz 2001; Little/Simpson 2003; Lenz/Schneider 2004)<sup>76</sup>.

Beacco (2004) ordnet ebenfalls seine Beschreibungen kultureller und interkultureller Kompetenzen den Niveaus A1-C2 zu. Sein Vorschlag, den er ausdrücklich zur Diskussion stellt, will vor allem eine Planungshilfe für den Fremdsprachenunterricht sein. Er vermerkt selbst mit aller Deutlichkeit:

« Ces spécifications des compétences devraient être en mesure de constituer la base d'un référentiel culturel, sans que soient pour autant résolues la question de la diversité des situations éducatives et celle de leur organisation en niveaux successifs de compétences, comme pour les compétences langagières et communicationnelles. C'est peu dire que la détermination de niveaux de maîtrise distincts pour ces compétences culturelles demeure extrêmement problématique » (Beacco 2004, 253).

Im Projekt INCA werden drei Kompetenzstufen unterschieden: Stufe 1 = *Niedrige Kompetenz*, Stufe 2 = *Mittlere Kompetenz* und Stufe 3 = *Hohe Kompetenz*. Die drei Stufen werden in einer Version für die Benutzer (Assessee-Version) und einer Version für Beurteilende (Assessoren-Version) beschrieben. Beide Versionen enthalten eine (etwas unterschiedlich formulierte) Globalskala sowie Skalen zu den sechs Kompetenzelementen Empathie, Offenheit gegenüber anderen Kulturen, Wissenserwerb, Kommunikationsbewusstsein, Ambiguitätstoleranz, Verhaltensflexibilität. Bemerkenswert ist, dass in der Assessee-Version im Unterschied zur Assessoren-Version die Niveaubeschreibungen weitgehend positiv formuliert sind. Die Beschreibungen der drei Niveaus sind nicht unabhängig voneinander – als Stand-alone-Statements – formuliert, sondern nehmen auf die frühere Kompetenzstufe Bezug. So heisst es in der „Allgemeinen Definition“ der Assessee-Version für die mittlere Kompetenz: „Aufgrund meiner Erfahrung bzw. meines Trainings beginne ich, Zusammenhänge in einigen Aspekten interkultureller Begegnungen, die ich bisher als einmalige Ereignisse betrachtet habe, zu erkennen.“ Die Beschreibung für Stufe 3 (Hohe Kompetenz) beginnt: „Viele der Kompetenzen, die ich bewusst in der zweiten Kompetenzstufe entwickelt habe, wende ich nun intuitiv an.“ Die Assessee-Version der dreistufigen Skalen ist auch im „Pass für interkulturelle Kompetenz“ des INCA-Portfolios enthalten und soll der „fortlaufenden Leistungsbeurteilung am Arbeitsplatz und im Zuge von Trainingsmaßnahmen“ dienen.

Auch im Projekt des Europäischen Fremdsprachenzentrums in Graz zur Entwicklung und Beurteilung interkultureller kommunikativer Kompetenzen werden – aufbauend auf den Arbeiten von Denise Lussier – drei Stufen unterschieden. Lázár et al. beziehen sich für die summative Evaluation zum einen auf die dreistufige Skala des INCA-Projekts und zum andern unverständlicherweise auf die sechsstufige Skala von Valdes (1986), die allerdings mit ihrer Ausrichtung am Ideal des Native-Speaker kaum mit der ansonsten postulierten Ausrichtung am *Referenzrahmen* und dem INCA-Projekt in Einklang zu bringen ist. Ausserdem ist auf der CD und der Website des Projekts noch zu den drei Bereichen Interkulturelles Wissen, Interkulturelle Fertigkeiten und Interkulturelle Einstellungen jeweils eine dreistufige Skala enthalten. Die Skalen sollen als Hilfe für Assessoren und für den Gebrauch im Klassenzimmer dienen können. Die Brauchbarkeit ist jedoch fraglich. Denn die Deskriptoren – im Sinne von Globalskalen zu längeren Abschnitten zusammengefasst – sind vielfach sehr allgemein, vage und uneinheitlich und teilweise kompliziert formuliert. Die Abgrenzung der drei Bereiche scheint nicht immer eindeutig und plausibel. es gibt keine klaren Informationen über die Herkunft der Deskriptoren und die Validierung der Stufeneinteilung. Für eine Selbstbeurteilung sind die Deskriptoren sicherlich nicht geeignet und auch nicht bestimmt.

---

<sup>76</sup> Nicht empirisch validiert sind auch z. B. die Stufungsvorschläge zur interkulturellen Bildung von Gogolin (2003) und zur Kompetenz im Umgang mit literarischen Texten von Burwitz-Melzer (2005 und 2008).

## **Evaluierung interkultureller Kompetenzen**

Scarino (2007) postuliert, dass eine Evaluierung des interkulturellen Sprachenlernens u. a. eine Reihe von Anforderungen genügen muss:

Given the features of intercultural language learning, its assessment needs to:

- position the student as both performer and analyser in interaction, though in any individual procedure one or the other role may be foregrounded for different purposes; (...)
- draw upon a range of assessment-types including interviews, conferences, journals, observations, story-telling, and capture students' cumulative learning so that development and progress can be taken into account, for example, through portfolios;
- include self-assessment that recognises learning as a personal process;
- include dimensions that require reflection on the part of students on their developing knowledge and understanding.

Auf die notwendige Vielfalt von Beurteilungsverfahren und -instrumenten weisen auch Lázár et al. (2007) hin und nennen als Verfahren die nützliche Informationen liefern können « les instruments d'enquêtes en pré- et post-formation ou enseignement, l'auto-évaluation des apprenants, l'observation de la démarche d'apprentissage et des progrès de l'apprenant à l'aide de grilles d'évaluation remplies par l'enseignant, les évaluations de l'enseignant et le portfolio de l'apprenant ». An gleicher Stelle werden dann allerdings etwas stark simplifizierend den drei Savoir-Bereichen bestimmte Verfahren zugewiesen:

Les méthodes d'évaluation vont donc varier en conséquence afin d'évaluer l'apprenant aussi efficacement que possible : procédures d'évaluation indirectes pour les « savoirs », réalisation de tâches pour le « savoir-faire », auto-évaluation, questionnement sur les attitudes, grilles de l'enseignant et portfolio de l'apprenant pour le « savoir-être ». (Lázár et al. 2007)

Am konkretesten ist die Evaluierung interkultureller Kompetenz wohl im INCA-Projekt ausgearbeitet. Das INCA-Assessment umfasst drei als „Testarten“ bezeichnete Instrumente bzw. Verfahren

1. Zwei Fragebögen, die über den beruflichen Hintergrund und die interkulturelle Erfahrung der Assessee Auskunft geben;
2. Szenarios. Zu den text- oder videobasierten Szenarios mit interkulturellen Situationen müssen Fragen beantwortet werden, wobei jede Frage einen Aspekt interkultureller Kompetenz abdecken soll;
3. Rollenspiele, in denen die Assessee in einer vorgegebenen Rolle interagieren und kommunizieren müssen. Die Rollenspiele sollen alle sechs Kompetenzaspekte abdecken.

Für die drei „Testarten“ sind unterschiedliche Assessoren vorgesehen. Die Ergebnisse werden in einer Assessorenkonferenz diskutiert, die sich auf ein Gesamtergebnis für die drei Säulen interkultureller Kompetenz einigt. Das aufwändige Verfahren nach Art eines Assessment Center mit Elementen eines elektronischen Assessment (E-Assessment), setzt eine gründliche Schulung der Assessoren voraus (vgl. Bolten 2007). Das auf Erwachsene und ihre Berufswelt ausgerichtete Evaluierungskonzept von INCA ist schon wegen des notwendigen Aufwands nicht für Tests im Rahmen des Bildungsmonitorings für die Sekundarstufe I geeignet.

Bezeichnenderweise wird das Assessmentverfahren von INCA ergänzt durch ein „Interkulturelles Portfolio“, das mit seinen drei Teilen und in vielen Aspekten dem „Europäischen Sprachenportfolio“ nachempfunden ist. Es enthält u. a. auch einen Selbstbeurteilungsraster.

Wie einleitend vermerkt ist es unserer Meinung nach angemessen, in der Sekundarstufe I ganz auf die formative Evaluation in der Arbeit mit dem Sprachenportfolio zu setzen. Die Hinweise und Instrumente im ESP II formulieren das Programm in einer für die Lernenden verständlichen Form:

Die Sprachbiografie hilft dir

- zu beschreiben, welche Erfahrungen du im Kontakt mit anderen Sprachen sammeln konntest und was du dabei über andere Menschen, Länder und Kulturen gelernt hast;

- die Sprachen und Kulturen der Menschen in deiner Umgebung besser zu kennen und zu respektieren.

Die Schülerinnen und Schüler werden angeleitet, ihre Erfahrungen mit anderen Kulturen und Sprachen zu reflektieren und zu diskutieren, zu beschreiben und zu dokumentieren. Ausdrücklich eingeschlossen sind auch Erfahrungen im Kontakt mit Schulkameradinnen und -kameraden, aus anderen Kulturen, Brieffreundschaften und ausserschulische Begegnungen und Erfahrungen (ESP II, Anleitung S.10).

In der einleitend zu diesem Abschnitt erwähnten Expertentagung, an der u. a. auch originale und adaptierte aus verschiedenen Quellen stammende Deskriptoren zur interkulturellen Kompetenz gewichtet und beurteilt wurden, wurde das folgende Ziel, ein Deskriptor, in dem das Portfolio erwähnt ist, als relevant für die Zielgruppe von HarmoS hoch gewichtet:

- die Bereitschaft und Fähigkeit, über Erfahrungen im Kontakt mit anderen Kulturen z. B. im *Sprachenportfolio* zu berichten und eigene Gedanken dazu festzuhalten.

Die beiden anderen ähnlich hoch gewichteten Deskriptoren mit Zielen für interkulturelles Lernen in der Schule waren:

- die Bereitschaft und Fähigkeit, den Alltag anderer Sprachregionen (auch mit Hilfe der lokalen Schulsprache) zu erkunden, mit der eigenen Lebenswelt zu vergleichen und sich auf eine reale Begegnung mit Sprechern anderer Sprachregionen vorzubereiten;
- Wissen und Verständnis dafür, dass es zwei- bzw. mehrsprachige und zwei- bzw. mehrkulturelle Identitäten gibt.

#### 4.2.3 Methodische Kompetenzen

In den Ergebnissen der kommunikativen HarmoS-Testaufgaben sind Einflüsse der Anwendung verschiedener Strategien enthalten, u. a.:

- die Auswirkung von innerhalb der schulischen Lerngeschichte angewendeten Lernstrategien und von lernfördernden Kommunikationsstrategien, die in der Vergangenheit in schulischen und ausserschulischen Kommunikationssituationen verwendet wurden;
- die Wirkung von Sprachgebrauchstrategien, die in der Testsituation eingesetzt werden (z. B. Strategien zur Texterschliessung, Planungsstrategien beim Schreiben oder Kommunikationsstrategien in der Interaktion);
- die Wirkung von Prüfungsstrategien, die vor oder in der Testsituation eingesetzt werden.

Der Einfluss solcher Strategien wird in den Testaufgaben nicht direkt und kontrolliert erfasst und es wurden für Lernerstrategien auch keine Basisstandards definiert. Das darf auf keinen Fall so interpretiert, als wären diese Kompetenzbereiche weniger wichtig. Sie müssen im Gegenteil in den regionalen Lehrplänen, im Unterricht und in der Lehrerbildung einen zentralen Stellenwert haben.

Seit den achtziger Jahren des 20. Jahrhunderts haben Lernerstrategien in der Fremdsprachenforschung und Fremdsprachendidaktik grosse Beachtung gefunden. Es gibt unterschiedlich weite Definitionen und viele unterschiedliche Einteilungen. Es ist charakteristisch für die Entwicklung ist, dass der Bereich der Lernerstrategien zunehmend weiter und komplexer gefasst wurde. So hat Rebecca Oxford (1990) in ihr bekanntes Strategien-Inventar (*Strategy Inventory for Language Learning, SILL*) neben den Kommunikationsstrategien sowie den kognitiven und metakognitiven Strategien (O'Malley/Chamot 1990) auch Gruppen von affektiven und sozialen Strategien sowie Gedächtnisstrategien aufgenommen.

In neuerer Zeit wurde das Konzept der Lernerstrategien stark in Verbindung gesetzt zu den Konzepten der Sprachbewusstheit/Sprachreflexion (*Language Awareness*) einerseits (z. B. Eckerth/Riemer 2000; McIntosh/Noels 2004; Morkötter 2005; Svalberg 2007; Candelier et al. 2007) und der Lernerautonomie andererseits (z. B. Wolff 1989; Bimmel/Rampillon 2000). So heisst es beispielsweise im EDK-Bericht zu „JALING“ (Saudan et al. 2005, 9):

„Begegnung mit Sprachen soll:

- das Bewusstsein und das Interesse für die eigene(n) und andere Sprachen wecken,

- die Motivation für das Erlernen von Fremd- und Schulsprachen unterstützen,
- Methodenkompetenz im Sinne von Verständigungs-, Lern- und «Erforschungs»- Strategien und selbstgesteuertes Arbeiten beim sprachlichen Lernen schulen,
- und mittels vergleichender Reflexionen über Sprache(n) und Sprachenlernen metasprachliche und metakognitive Fähigkeiten aufbauen helfen.“

Es liegen manche anregende Materialien für die Entwicklung solcher Kompetenzbereiche vor (z. B. Balsiger et al. 2005), doch sind Aktivitäten im Sinne von ELBE / EOLE (*Language Awareness, Begegnung mit Sprachen / Education et ouverture aux langues à l' école*) noch nicht fester Bestandteil der Lehrpläne. In den neuen regionalen Lehrplänen müssen sie einen angemessenen Platz finden. Das gilt ebenfalls für die im Sinne einer integrierten Sprachendidaktik bzw. Mehrsprachigkeitsdidaktik wichtigen Fertigkeiten des Vergleichens sowie für Interkomprehensions- und Transfer-Strategien (vgl. z. B. Berthele 2007; Candelier et. al. 2007; Doyé 2006; Edmondson 2001; Meißner 2002, 2004, 2006).

Lange Zeit wurden Lernerstrategien nur in Verbindung mit jeweils einzelnen Fremdsprachen gesehen. In Lehrwerken und in der Unterrichtspraxis sind vor allem Strategien und Tipps zum Leseverstehen und zum Wortschatzlernen präsent. Seltener werden z. B. Strategien des Planens und Überarbeitens beim Schreiben in der Fremdsprache und noch viel seltener Interkomprehensionsstrategien explizit gefördert.

Im Kapitel 2 der KMK Bildungsstandards für die erste Fremdsprache für den Hauptschulabschluss (9. Klasse) „Kompetenzbereiche für die erste Fremdsprache“ heisst es unter dem Titel „Methodische Kompetenzen“:

Im Unterricht in der Fremdsprache werden systematisch methodische Kompetenzen entwickelt. In diesem Kompetenzbereich können Schülerinnen und Schüler Kompetenzen entwickelt. In diesem Kompetenzbereich können Schülerinnen und Schüler

- ihrem persönlichen Lernstil entsprechende Entscheidungen über geeignete Lernwege treffen,
- verschiedene Verfahren zum Umgang mit und zur Auswertung von gesprochenen und geschriebenen Texten nutzen,
- Verfahren zum Gestalten von mündlichen und schriftlichen Texten anwenden,
- kooperative Formen des Arbeitens anwenden,
- Lerntechniken und -strategien für den Ausbau ihrer Kenntnisse in der ersten Fremdsprache sowie für den Erwerb weiterer Sprachen einsetzen.

Damit verfügen Schülerinnen und Schüler über Grundlagen für lebenslanges Sprachenlernen<sup>77</sup>. (KMK 2005, 10; ähnlich KMK 2004, 10)

Im Kapitel „Standards für die Kompetenzbereiche“ werden dann „fachliche und fachübergreifende Arbeitstechniken und Methoden“, über die die Schülerinnen und Schüler verfügen können sollen, für die folgenden Bereiche aufgeführt (KMK 2005, 15-16; KMK2004, 17-18):

- Textrezeption (Leseverstehen und Hörverstehen),  
z. B. : „verschiedene Hör- und Lesetechniken (u. a. globales, suchendes, detailliertes Hören und Lesen) aufgabenbezogen/funktionsbezogen einsetzen“;
- Interaktion;  
z. B. „grundlegende Regeln des Gesprächsablaufs beachten und Verständigungsprobleme durch Rückfragen und nonverbale Mittel überwinden“;
- Textproduktion (Sprechen und Schreiben);  
z. B. „Techniken zur Vorbereitung eigener Texte oder Präsentationen anwenden, z. B. Stichworte notieren, Gliederungen erstellen, Handlungsgeländer anfertigen, Bilder verwenden“;
- Lernstrategien;  
z. B. „Verfahren zum Memorieren und Abrufen von Wörtern und Redemitteln anwenden“;
- Präsentation und Mediennutzung;  
„Neue Medien zur Informationsbeschaffung, zur kommunikativen Interaktion und zum Lernen einsetzen“

---

<sup>77</sup> In den Bildungsstandards für den Mittleren Schulabschluss (10. Klasse) lautet die zusammenfassende Beschreibung: „Damit verfügen Schülerinnen und Schüler über die Fähigkeit, selbstgesteuertes und kooperatives Sprachlernverhalten als Grundlage für den Erwerb von weiteren Sprachen sowie für das lebenslange, selbstständige Sprachenlernen einzusetzen“ (KMK 2004, 10).

und „Präsentationstechniken zur Darstellung von Arbeitsergebnissen einsetzen (Medien auswählen, Gliederungs- und Visualisierungstechniken anwenden)“;

– Lernbewusstheit und Lernorganisation

z. B. „selbstständig und kooperativ arbeiten“; „für sie förderliche Lernbedingungen erkennen und nutzen, ihre Lernarbeit organisieren und die Zeit einteilen“ (KMK 2005, 15-16; ähnlich KMK 2004, 17-18).

Wenn für die methodischen Kompetenzen Standards, im Falle der KMK Regelstandards, festgelegt werden sollen, dann stellt sich die Frage nach Möglichkeiten und Verfahren einer begründeten Auswahl und Stufung von Methodenkompetenzen.

Nun haben die Forschungen und Diskussionen über den „guten Sprachlerner“ seit Wenden/Rubin (1987) deutlich gemacht, dass der Lernerfolg nicht einem bestimmten Set von Lernerstrategien zugeschrieben werden kann. Zwar gibt es Anzeichen dafür, dass erfolgreiche Lerner oft über ein grösseres Inventar von Strategien verfügen, das sie flexibel und vielfach bewusst nutzen (Skehan 1989; Cohen 1998; Dornyei 2005 u. a.). Aber die Präferenz für bestimmte Strategien und deren Nutzung ist stark durch den Lernertyp und den persönlichen Lernstil bestimmt. Lernstile sind aber nur schwer veränderbar (Grotjahn 2003). Entscheidend ist, dass die Lernenden das für sie jeweils individuell am besten geeignete Strategierepertoire entwickeln können, und das bedeutet, dass im Unterricht ein entsprechend breites Angebot gemacht und viel Gelegenheit gegeben werden muss, unterschiedliche Strategien auszuprobieren und über deren Erfolg und individuelle Passung zu reflektieren (vgl. z. B. Tönshoff 1998, 182).

Es ist allerdings äusserst schwierig, zu überprüfen, ob denn die Lernenden tatsächlich, über das für sie geeignete Repertoire verfügen und damit einen Standard erfüllen.

Mit Verweis darauf, dass der *Referenzrahmen* für die methodischen Kompetenzen „nur allgemeine Ausführungen ohne Niveaustufen“ enthält, wurden in den deutschen Bildungsstandards, wie es heisst, „erfahrungsgestützte Festlegungen vorgenommen“ (KMK 2005, 11). Die Beschreibungen der erwarteten methodischen Kompetenzen für die 9. Jahrgangsstufe (Hauptschule) und die 10. Jahrgangsstufe (Mittlerer Schulabschluss) sind teilweise unterschiedlich. Doch scheint es sich dabei eher um Unterschiede in der Formulierung zu handeln, darauf zurückzuführen, dass unterschiedliche Autorengruppen formuliert haben. Wirklich inhaltliche Unterschiede und ein klares Konzept für unterschiedliche Anspruchsniveaus sind kaum zu erkennen. Teilweise sind beispielsweise die Formulierungen für die 9. Klasse allgemeiner und scheinen daher umfassendere Kompetenzen zu beschreiben als die entsprechenden detaillierteren Beschreibungen für die 10. Klasse.

Rampillon (2003) versucht eine Stufung der Lernkompetenz vorzunehmen und, so der Titel ihres Beitrags, ein „Lernstrategisches Minimalprofil an der Schwelle von L2 und L3“ zu bestimmen. Sie schreibt relativierend dazu:

Lernstrategien sind in ihrer Ausprägung wie auch in der Anzahl abhängig von der Kreativität der Lehrenden wie der Lernenden. Daher kann es keine fest definierte Liste von Lernstrategien geben, sondern lediglich eine Zusammenstellung wichtiger und häufig angewandter Verfahren. Auf dieser Basis ist auch die folgende Liste der lernstrategischen Grundkenntnisse erstellt worden. Sie gibt ein Minimalprofil für die erreichte Lernkompetenz in L2 wieder und stellt damit gleichzeitig eine Grundlage für die Weiterführung der Lernkompetenz in L3 dar. Ergänzungen sind jederzeit denkbar (Rampillon 2003, 90).

Sie ist sich bewusst, dass es sich bei ihrer Auflistung um Setzungen handelt und stellt ihren Versuch ausdrücklich zu Diskussion:

Eine Schwierigkeit bei der Festlegung des Basiswissens für L2 bzw. L3 besteht darin, dass es bisher kaum Grundlagen gibt, an denen man sich orientieren könnte, um ein Minimalprofil, ein Aufbauprofil und vielleicht sogar ein Perfektionsprofil zu beschreiben. Darüber hinaus ist die Ausgangsbasis für das lernstrategische Wissen der Lernenden hypothetisch, da auch nicht bekannt ist, welche Lernstrategien sie im Einzelnen bereits aus dem vorschulischen Lernen und aus dem Unterricht der Primarstufe mitbringen und inwieweit sie diese auch tatsächlich beherrschen und anwenden. Da dieses zum Teil in den Bereich der Spekulation hineinreicht, mussten Setzungen vorgenommen werden (Rampillon 2003, 91).

Zu dem interessanten Ansatz lernstrategische Ziele für die erste und die zweite Fremdsprache zu unterscheiden, merkt sie an, dass auch die Verteilung der Lernstrategien auf L2 oder auf L3 „in

vielen Fällen willkürlich gesetzt“ wurde „da eine Progression nach Schwierigkeit nicht immer gegeben ist. Es handelt sich also oftmals eher um eine quantitative Verteilung, bei der die genannten Lernstrategien letztlich auch austauschbar sind“ (Rampillon 2003, 91). Anzumerken ist, dass die nicht immer leicht nachzuvollziehende Verteilung auf die erste und zweite Fremdsprache angepasst werden müsste, je nachdem in welchem Alter mit dem Unterricht der zweiten Fremdsprache begonnen wird. Rampillon selbst betont, dass die „Zahl der Imponderabilien derzeit noch sehr hoch ist und dass es weiterer Forschungen bedarf, um zu gesicherteren Ergebnissen zu kommen“ (Rampillon 2003, 103).

Bezogen auf die Kompetenzbereiche „Sprachreflexionskompetenz (Sprachbewusstheit, Sprachlernkompetenz (Sprachlernbewusstheit) und Strategiekompetenz (Strategiebewusstheit)“ plädiert auch Vollmer 2004) dafür „trotz aller Schwierigkeit und Vorläufigkeit bestimmte Niveaustufen“ zu definieren. Erschlägt vor, „dass konkrete Maßnahmen sowohl in ein Kompetenz-Standard-Modell als auch in ein didaktisches Entwicklungs-Modell eingebettet werden. Dies ist nicht leicht, aber leider auf lange Sicht unabdingbar. Erst die Entwicklung eines solchen Rahmens, die Operationalisierung der involvierten Kompetenzen und Kompetenzniveaus sowie deren empirische Überprüfung geben einer Mehrsprachigkeitsdidaktik und -methodik die angemessene Perspektive“ (Vollmer 2004, 244).

Bis dies erreicht wird, ist zweifellos noch sehr viel Forschungsarbeit zu leisten. es wäre wünschenswert, wenn auch Folgeprojekte zu HarmoS zu dieser Forschungsarbeit beitragen könnten. Die Lehrplanarbeit aber kann nicht warten, bis die neuen Forschungsergebnisse vorliegen, sondern muss die methodischen Ziele auf der Basis der vorhandenen Kenntnisse und Erfahrungen festlegen. In den vorliegenden Arbeiten und auch in den diversen Listen methodischer Kompetenzen (z. B. Bimmel/Rampillon 2000; Rampillon 2003; KMK 2004, 2005) und nicht zuletzt in den vielen Sprachenportfolios lassen sich viele Anregungen finden.

Sowohl für Forschungsprojekte als auch für die Lehrplanarbeit müssten unbedingt Vertreter der Schulsprache, der Fremdsprachenfächer und der Sachfächer zusammenarbeiten. Denn Sprachlernstrategien sind, wie es im Titel von Grenfell/Harris (2004) heisst „a case for cross-curricular collaboration“ (vgl. Vollmer 2006a, 2006b). Noch gibt es auch in der Schweiz noch manche Widerstände und Skepsis gegenüber der Umsetzung einer integrierten Sprachendidaktik. Meißner (2006, 163) betont zu Recht: „Dabei wäre es gerade auf dem Feld der Lernkompetenzentwicklung angesagt, ‚mehrsprachigkeitsdidaktisch‘ vorzugehen“ (vgl. u. a. Behr 2007). Das aber verlangt unbedingt Kooperation (Behr 2007, 91; Meißner 2004, 156; Vollmer 2004, 243; Schneider 2007, 150).

Kooperation über die Einzelfächer hinweg verlangt auch die Arbeit mit dem „Europäischen Sprachenportfolio“, dem als einem nicht an ein Einzelfach gebundenem Instrument eine entscheidende Bedeutung zukommt. Das Sprachenportfolio in der Version für Kinder und Jugendliche von 11 bis 15 Jahren (ESP II) regt an, die eigene Sprachbiografie und Sprachlernbiografie zu reflektieren und gibt in den Rubriken „Wie ich am erfolgreichsten lerne“ viele Anstöße, unterschiedliche Strategien auszuprobieren, bewusst zu machen, zu üben und zu evaluieren. Damit das Sprachenportfolio aber seine Wirkung voll entfalten kann, darf es nicht nur punktuell eingesetzt und als Pflichtübung abgearbeitet werden. Noch betrachten viele Lehrpersonen die explizite Förderung von Strategien und die Arbeit mit dem Sprachenportfolio als eine zusätzliche zeitliche Belastung. Deshalb muss noch viel in die Lehrerbildung investiert werden: Die Lehrerinnen und Lehrer sollten in ihrer Aus- und Weiterbildung die Gelegenheit haben, den Ansatz einer integrierten Didaktik selbst zu erfahren: in Sprachkursen, in sprachenübergreifenden Veranstaltungen und Projekten, im Praktikum und durch Austausch. Sie sollten in ihrer Aus- und Weiterbildung das Sprachenportfolio nicht nur kennen lernen und lernen, wie die entsprechenden Portfolioversionen mit unterschiedlichen Lerngruppen eingesetzt werden können, sondern sollten auch ihr eigenes Portfolio füllen mit Dokumenten, Berichten und Reflexionen über ihre eigene Sprachbiographie, ihre Lernwege, Lerngewohnheiten und Strategien (Schneider 2007, 150).

#### 4.2.4 Literatur zu den überfachlichen Kompetenzen

##### Literatur zu den interkulturellen Kompetenzen

- Bättig, Brigitte / Ruf, Barbara (2005): HARMOS: Lehrplanuntersuchung erste und zweite Fremdsprachen. EDK: Informationsstelle IDES.
- Beacco, Jean-Claude (2004): Une proposition de référentiel pour les compétences culturelles dans les enseignements de langues. In : Beacco, Jean-Claude/ Bouquet, Simon/ Porquier, Rémy (2004): Niveau B2 pour le Français. Textes et Références. Paris: Didier, 251-286.
- Bennett, Milton .J. (1993). Towards ethnorelativism: A developmental model of intercultural sensitivity. In: M. Paige (Ed.), Education for the intercultural experience. Yarmouth, ME: Intercultural Press, 21-71.
- Bersinger, Simone / Jordi, Urs / Tschang, Martine (2005): Europäisches Sprachenportfolio - European Language Portfolio - Portfolio Européen des langues - Version für Kinder und Jugendliche von 11 bis 15 Jahren ESP II. Bern: Schulverlag blmv.
- Boeckmann, Klaus-Börge (2006): Dimensionen von Interkulturalität im Kontext des Fremd- und Zweitsprachenunterrichts. In: Zeitschrift für Interkulturellen Fremdsprachenunterricht [Online] 11/3, 14 S. Online: <http://www.ualberta.ca/~german/ejournal/Boeckmann1.htm> (23.01.08).
- Bolten, Jürgen (2007): Was heißt „Interkulturelle Kompetenz?“ Perspektiven für die internationale Personalentwicklung. In: V.Künzer/ J. Berninghausen (Hrsg.): Wirtschaft als interkulturelle Herausforderung. Frankfurt/M, 21-42. Online: [http://www2.uni-jena.de/philosophie/iwk/publikationen/interkulturelle\\_kompetenz\\_personalentwicklung\\_bolten.pdf](http://www2.uni-jena.de/philosophie/iwk/publikationen/interkulturelle_kompetenz_personalentwicklung_bolten.pdf) (05.05.08).
- Bolten, Jürgen (2001):Thesen zum interkulturellen Lernen in der Schule. In: Bolten, Jürgen / Schröter, Daniela (Hrsg.): Interkulturelle Kommunikation. Sternenfels/ Berlin: Verlag Wissen und Praxis.
- Buchberger, Gerlinde / Horak, Angela / Matzer, Edith / Mewald, Claudia / Seidlhofer Barbara / Spenger, Jörg / Stefan, Ferdinand (2002). Standards für die 8. Schulstufe Fachbereich Fremdsprachen (am Beispiel des Englischen). Arbeitsentwürfe für die Hand der Lehrer/innen. Erprobungsversion, Teil II. Online: [http://arge.stvg.at/arge.nsf/073757dce34766ffc12569960053dced/00b8deacb440a538c1256c5b002ae95d/\\$FILE/Standards\\_Erprobungsversion\\_Teil\\_2\\_Juni\\_02.pdf](http://arge.stvg.at/arge.nsf/073757dce34766ffc12569960053dced/00b8deacb440a538c1256c5b002ae95d/$FILE/Standards_Erprobungsversion_Teil_2_Juni_02.pdf). (HTML-Version 24.01.08).
- Burwitz-Melzer, Eva (2008): Interkulturelles und sprachliches Lernen mit fremdsprachlichen literarischen Texten: Zwei zentrale Elemente eines neuen Lesekompetenzmodells. In: FLUL 35, 104-120.
- Burwitz-Melzer, Eva (2005): Kompetenzen für den Literaturunterricht heute. Ein Beitrag zur standardorientierten Didaktik des Fremdsprachenunterrichts . In: FLUL 34, 94-120.
- Byram, Michael (2003): Intercultural competence. Strasbourg: Council of Europe.
- Byram, Michael / Gribkova, Bella / Starkey, Hugh (2002): Développer la dimension interculturelle de l'enseignement des langues. Une introduction pratique a l'usage des enseignants. Strasbourg: Conseil de l'Europe.
- Byram, Michael / Neuner, Gerhard / Zarate, Geneviève (1997): Sociocultural competence in language learning and teaching. Studies towards a Common European Framework of reference for language teaching and learning. Strasbourg: Council of Europe Publishing.
- Candelier, Michel / Camilleri-Grima, Antoinette / Castellotti, Véronique / de Pietro, Jean-François / Lörincz, Ildikó / Meissner, Franz-Joseph / Schröder-Sura, Anna / Noguero, Artur (2007): CARAP Cadre de Référence pour les Approches Plurielles des Langues et des Cultures. Version 1. Graz: Centre Européen pour les Langues Vivantes.
- Candelier, Michel / Camilleri-Grima, Antoinette / Castellotti, Véronique / de Pietro, Jean-François / Lörincz, Ildikó / Meissner, Franz-Joseph / Schröder-Sura, Anna / Noguero, Artur (2007): CARAP - Framework of reference for pluralistic approaches to languages and cultures . Version 2. Graz: Centre Européen pour les Langues Vivantes.
- Chen, Guo-Ming/ Starosta, William G. (1998): Foundations of Intercultural Communication. Needham Height, Mass: Allyn and Bacon.
- Condat, Sophie (2008): L'éducation interculturelle. Bibliographie. Paris: Centre de ressources documentaires du CIEP. Online: [http://www.ciep.fr/bibliographie/Education\\_interculturelle.pdf](http://www.ciep.fr/bibliographie/Education_interculturelle.pdf) (24.04.08).
- Coste, Daniel (2007): Contextualiser les utilisations du Cadre européen commun de référence pour les langues. In : Conseil de l'Europe : « Le Cadre européen commun de référence pour les langues (CECR) et l'élaboration de politiques linguistiques : défis et responsabilités ». Forum intergouvernemental sur les politiques linguistiques. Rapport, 44-53. Online: [http://www.coe.int/t/dg4/linguistic/Source/SourceForum07/ForumFeb07\\_%20Report\\_0807\\_FR.doc](http://www.coe.int/t/dg4/linguistic/Source/SourceForum07/ForumFeb07_%20Report_0807_FR.doc) (16.07.08).
- Doyé, Peter (2006): Interkomprehension. In: ZFF - Zeitschrift für Fremdsprachenforschung 17 / 2, 245-256.
- EU, European Commission Directorate General for Education and Culture, Multilingualism Policies: Languages and cultures in Europe. LACE. Website: <http://www.lace2007.eu/> (17.07.08).
- Franklin, Peter / Langé, Gisella / Marsh, David / Spencer-Oatey, Helen (2007): Studie über die Vermittlung interkultureller Kompetenzen im Fremdsprachenunterricht im Pflichtschulbereich in der Europäischen Union. Sprachen und Kulturen in Europa (LACE). Online: <http://www.lace2007.eu/docs/de/final.pdf> (06.05.08).

- Gassner, Otmar/ Horak, Angela/ Mewald, Claudia/ Moser, Wolfgang/ Schober, Michael/ Stefan, Ferdinand/ Valsky, Claudia (2005): Bildungsstandards in Österreich. Fremdsprachen. Englisch 8. Schulstufe. Version September'05. Wien: Bundesministerium für Bildung, Wissenschaft und Kultur.
- Glaboniat, Manuela / Müller, Martin / Rusch, Paul / Schmitz, Helen / Wertenschlag, Lukas (2005): Profile deutsch. Gemeinsamer europäischer Referenzrahmen. Lernzielbestimmungen - Kannbeschreibungen - Kommunikative Mittel - Niveau A1 - A2 - B1 - B2 - C1 - C2 mit CD-ROM. München: Langenscheidt.
- Gogolin, Ingrid (2003): Fähigkeitsstufen der Interkulturellen Bildung. Hamburg: Institut für International und Interkulturell Vergleichende Erziehungswissenschaft. Mimeo. Online: [http://www.erzwiss.uni-hamburg.de/Personal/Gogolin/cosmea/core/corebase/mediabase/foermig/website\\_gogolin/dokumente/Gogolin\\_2002\\_F\\_higkeitsstufen\\_Mimeo.pdf](http://www.erzwiss.uni-hamburg.de/Personal/Gogolin/cosmea/core/corebase/mediabase/foermig/website_gogolin/dokumente/Gogolin_2002_F_higkeitsstufen_Mimeo.pdf) (17.07.08).
- Hammer, M. R., Bennett, M. J., & Wiseman, R. (2003): Measuring intercultural sensitivity: The intercultural development inventory. In: *International Journal of Intercultural Relations*, 27, 421–443.
- Heringer, Hans Jürgen (2004): *Interkulturelle Kommunikation*. Tübingen: Francke (=UTB 2550).
- Hesse, Hermann-Günter / Göbel, Kerstin (2007): *Interkulturelle Kompetenz*. In: Beck, Bärbel/ Klieme, Eckhard (Hrsg.): *Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)*. Weinheim und Basel: Beltz, 256-272.
- Humphrey, Donna (2007): *Intercultural Communication Competence: The State of Knowledge*. Report prepared for CILT The National Centre for Languages: Development of NOS in intercultural Skills Project. Online: <http://www.cilt.org.uk/standards/donnareport.pdf>
- INCA (2004.): *Intercultural Competence Assessment*. Online: <http://www.incaproject.org/> und [http://www.incaproject.org/de\\_downloads/21\\_INCA\\_Assessor\\_Manual\\_germ.pdf](http://www.incaproject.org/de_downloads/21_INCA_Assessor_Manual_germ.pdf) (21.01.08).
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2004) (Hrsg.): *Bildungsstandards für die erste Fremdsprache (Englisch/ Französisch) für den Mittleren Schulabschluss*. Neuwied: Luchterhand.
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005) (Hrsg.): *Bildungsstandards für die erste Fremdsprache (Englisch/ Französisch) für den Hauptschulabschluss*. Neuwied: Luchterhand.
- Lázár, Ildikó (2003): *Incorporating intercultural communicative competence in language teacher education*. Strasbourg: Conseil de l' Europe.
- Lázár, Ildikó (2007): *Developing and assessing intercultural communicative competence*. In: *Babylonia* 3, 9-13.
- Lázár, Ildikó / Huber-Kriegler, Martina / Lussier, Denise / Matei, Gabriela S. / Peck, Christiane (2008): *Développer et évaluer la compétence en communication interculturelle - Un guide à l'usage des enseignants de langues et des formateurs d'enseignants*. Graz : Centre européen pour les langues vivantes.
- Lázár, Ildikó / Huber-Kriegler, Martina / Lussier, Denise / Matei, Gabriela S. / Peck, Christiane (Hrsg.) (2007): *Developing and assessing intercultural communicative competence – A guide for language teachers and teacher educators*. Graz: European Centre for Modern Languages.
- Lazar, Ildiko / Huber-Kriegler, Martina / Lussier, Denise / Matei, Gabriela S. / Peck, Christiane (2008): *Développer et évaluer la compétence en communication interculturelle - Un guide à l'usage des enseignants de langues et des formateurs d'enseignants avec cdr*. Strasbourg: Council of Europe.
- Lenz, Peter / Studer, Thomas (2005): *Neue Instrumente für die Beurteilung der Französisch- und Englischkompetenzen von Deutschschweizer Schülerinnen und Schülern*. In: Aline Gohard-Radenkovic (Hrsg.), *Mehrsprachigkeit, Interkulturalität und Fremdsprachendidaktik in einem zweisprachigen Kontext*, Transversales, Bern: Peter Lang, 37-56.
- Little, David (2006): *State-of-the-Art Article: The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact*. In: *Language Teaching* 2006, 3.
- Little, David / Simpson, Barbara (2003): *Portfolio européen des langues. La composante interculturelle et Apprendre à apprendre*. Strasbourg: Conseil de l'Europe. Online: <http://www.coe.int/T/DG4/Portfolio/documents/Maquettes.pdf> (24.01.08).
- Little, David / Hodel, Hans-Peter / Kohonen, Viljo / Meijer, Dick / Perclová, Radka (2008): *Préparer les enseignants du Portfolio européen des langues - arguments, matériels et ressources*. Livre et CDROM. Graz: European Centre for Modern Languages.
- LOLIPOP: *Language On-line Portfolio Project*. <http://www.isoc.siu.no/isocii.nsf/projectlist/116998> (17.07.08).
- Newby, David (2007): *The European Portfolio for Student Teachers of Languages*. In: *Babylonia*. 3, 23-26.
- Newby, David / Allan, Rebecca / Fenner, Anne-Brit / Jones, Barry / Komorowska, Hanna / Soghikyan, Kristine (2008): *Portfolio européen pour les enseignants en langues en formation initiale - Un outil de réflexion pour la formation des enseignants en langues*. Strasbourg: Conseil de l'Europe.
- Scarino, Angela (2007): *Assessing intercultural language learning*. Australian Government Quality Teacher Program: *Intercultural Language Teaching and Learning in Practice*. Discussion Paper 6. Online: [http://www.iltlp.unisa.edu.au/doclibpapers/iltlp\\_paper6.doc](http://www.iltlp.unisa.edu.au/doclibpapers/iltlp_paper6.doc) (23.01.08).

- Schneider, Günther (2003): Das Europäische Sprachenportfolio - Grundidee, Modelle, Erfahrungen. In: Schneider, Günther / Clalüna, Monika (Hrsg.): Mehr Sprache - mehrsprachig - mit Deutsch. Didaktische und politische Perspektiven, 256-275.
- Schneider, Günther (2005): Der ""Gemeinsame europäische Referenzrahmen für Sprachen"" als Grundlage von Bildungsstandards für die Fremdsprachen - Methodologische Probleme der Entwicklung und Adaptierung von Kompetenzbeschreibungen. In: Schweizerische Zeitschrift für Bildungswissenschaften 1, 13-37.
- Sercu, Lies (2002): Autonomes Lernen im interkulturellen Fremdsprachenunterricht. Kriterien für die Auswahl von Lerninhalten und Lernaufgaben. Zeitschrift für Interkulturellen Fremdsprachenunterricht [Online], 7/2, 16 S. Online: <http://zif.spz.tu-darmstadt.de/jg-07-2/beitrag/sercu1.htm> (23.01.08).
- Sercu, Lies (2004): Assessing intercultural competence: a framework for systematic test development in foreign language education and beyond. In: Intercultural Education 15/1, 73-90.
- Sinicrope, Castle., Norris, John .M., and Watanabe, Yukiko. (2007): Understanding and assessing intercultural competence: A summary of theory, research, and practice. Technical report for the Foreign Language Program Evaluation Project. Honolulu: University of Hawai'i, National Foreign Language Resource Center. Online: <http://nflrc.hawaii.edu/evaluation/files/ICC%20Final%20Report%20for%20publication.doc> (25.01.08).
- Vollmer, Hermut J. (2005): Das Gespenst der Standardisierung geht um ODER: Lehren und Lernen fremder Sprachen auf der Basis von Bildungsstandards. In: Bausch, Karl-Richard/ Burwitz-Melzer, Eva/ Königs, Frank G. (Hrsg.): Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand. Arbeitspapiere der 25. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, 258-271.
- Zarate, Geneviève / Gohard-Radenkovic, Aline / Lussier, Denise / Penz, Hermine (2003). Médiation culturelle et didactique des langues. Strasbourg: Centre européen pour les langues vivantes. Graz/Strasbourg: Editions du Conseil de l'Europe.

### **Literatur zu den methodischen Kompetenzen**

- Babylonia. Themenheft: Begegnung mit Sprachen / Ouverture aux langues. 1999/2.
- Babylonia. Themenheft: Strategien für den Sprachunterricht und das Sprachenlernen. 2002/2.
- Babylonia. Themenschwerpunkt: Integrierte Sprachendidaktik. 2005/4.
- Balsiger, Claudine / Berger, Claudia / Dufour, Janine / Gremion, Lise / de Pietro, Danièle / Zurbriggen, Elisabeth (2003): Education et ouverture aux langues à l' école eole. Volume 2 3e année primaire - 6e année. Documents reproductibles. Neuchâtel: SG / CIIP.
- Bausch, Karl-Richard / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2004): Mehrsprachigkeit im Fokus. Arbeitspapiere der 24. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr (= Giessener Beiträge zur Fremdsprachendidaktik).
- Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. (2005): Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand. Arbeitspapiere der 25. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr (= Giessener Beiträge zur Fremdsprachendidaktik).
- Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2006): Aufgabenorientierung als Aufgabe. Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr (= Giessener Beiträge zur Fremdsprachendidaktik).
- Behr, Ursula (2007): Sprachenübergreifendes Lernen und Lehren in der Sekundarstufe I. Ergebnisse eines Kooperationsprojektes der drei Phasen der Lehrerbildung. Tübingen: Narr (= Giessener Beiträge zur Fremdsprachendidaktik).
- Bernaus ,Mercè Andrade, Ana Isabel / Kervran, Martine / Murkowska, Anna / Trujillo Sáez, Fernando (2007) : La dimension plurilingue et pluriculturelle dans la formation des enseignants de langues : kit de formation. Strasbourg / Graz: Conseil de l'Europe / Centre européen pour les langues vivantes. Online : [http://www.ecml.at/mtp2/Lea/html/LEA\\_E\\_Results.htm](http://www.ecml.at/mtp2/Lea/html/LEA_E_Results.htm) (05.05.08).
- Bersinger, Simone / Jordi, Urs / Tschang, Martine (2005): Europäisches Sprachenportfolio - European Language Portfolio - Portfolio Européen des langues - Version für Kinder und Jugendliche von 11 bis 15 Jahren ESP II. Bern: Schulverlag blmv.
- Berthele, Raphael (2007): Zum Prozess des Verstehens und Erschliessens In: Hufeisen, Britta / Marx, Nicole (Hrsg.) (2007): EuroComGerm - Die sieben Siebe: Germanische Sprachen lesen lernen. Aachen: Shaker, 15-26.
- Bimmel, Peter / Rampillon, Ute (2000): Lernerautonomie und Lernstrategien. München: Langenscheidt.
- Candelier, Michel (2003): L' éveil aux langues à l' école primaire. Evlang: bilan d'une innovation européenne. Bruxelles: De Boeck (= Pratiques pédagogiques).
- Candelier, Michel et al. (2003): Janua Linguarum. La porte des langues L'introduction de l'éveil aux langues dans le curriculum. Strasbourg/Graz : Centre européen pour les langues vivantes / Editions du Conseil de l'Europe.
- Candelier, Michel / Camilleri-Grima, Antoinette / Castellotti, Véronique / de Pietro, Jean-François / Lörincz, Ildikó / Meissner, Franz-Joseph / Schröder-Sura, Anna / Noguero, Artur (2007): CARAP - Framework of reference for pluralistic approaches to languages and cultures . Version 2. Graz: Centre Européen pour les Langues Vivantes.

- Chamot, Anna Uhl (2004): Issues in Language Learning Strategy Research and Teaching. In: Electronic Journal of Foreign Language Teaching 1/1, 14-26. Online: <http://e-flt.nus.edu.sg/v1n12004/chamot.htm> (23.01.08).
- Cohen, Andrew (1998): Strategies in learning and using a second language. London: Longman.
- Department for children, schools and families (o.J.): Key Stage 2 Framework for languages. Language learning strategies (LLS): Planning: Analysing and evaluating ways of learning. Online: <http://www.standards.dfes.gov.uk/primary/publications/languages/framework/lfs/1205069/#1205935> (25.01.08).
- Dornyei, Zoltan (2001): Motivation strategies in the language classroom. Cambridge: University Press (= Cambridge Language Teaching Library).
- Dornyei, Zoltán (2005): The Psychology of the Language Learner. Individual Differences in Second Language Acquisition. London: Erlbaum.
- Doyé, Peter (2006): Interkomprehension. In: ZFF - Zeitschrift für Fremdsprachenforschung 17 / 2, 245-256.
- Eckerth, Johannes / Riemer, Claudia (2000): Awareness und Motivation: Noticing als Bindeglied zwischen kognitiven und affektiven Faktoren des Fremdsprachenlernens. In: Riemer, Claudia (Hrsg.): Kognitive Aspekte des Lehrens und Lernens von Fremdsprachen. Festschrift für Willis J. Edmondson zum 60. Geburtstag, 228-246.
- Edmondson, Willis J. (2001): Transfer beim Erlernen einer weiteren Fremdsprache: die L1-Transfer-Vermeidungsstrategie. In: Aguado, Karin / Riemer, Claudia: Wege und Ziele. Zur Theorie, Empirie und Praxis des Deutschen als Fremdsprache (und anderer Fremdsprachen). Festschrift für Gerd Henrici zum 60. Geburtstag. Hohengehren: Schneider, 137-154.
- Edmondson, Willis / House, Juliane (2006, 3. Aufl.): Einführung in die Sprachlehrforschung. Tübingen: Francke (= UTB 1697).
- Faerch, Claus / Kasper, Gabriele (1983): Strategies in Interlanguage Communication. London u. a.: Longman.
- Glaboniat, Manuela / Müller, Martin / Rusch, Paul / Schmitz, Helen / Wertenschlag, Lukas (2005): Profile deutsch. Gemeinsamer europäischer Referenzrahmen. Lernzielbestimmungen - Kannbeschreibungen - Kommunikative Mittel - Niveau A1 - A2 - B1 - B2 - C1 - C2 mit CD-ROM. München: Langenscheidt.
- Grenfell, Michael / Harris, Vee (2004): Language-learning strategies: a case for cross-curricular collaboration. In: Language Awareness, 13/2, 116-130. Online: <http://www.multilingual-matters.net/la/013/0116/la0130116.pdf> (17.07.08).
- Grotjahn, Rüdiger (2003): Lernstile/Lerntypen. In: Bausch, Karl-Richard / Christ, Herbert / Krumm, Hans-Jürgen (Hrsg.): Handbuch Fremdsprachenunterricht. Tübingen: Francke, 326-331.
- Harris, Vee (2001): Helping learners learn: exploring strategy instruction in language classrooms across Europe. Strasbourg: European Centre for Modern Languages / Conseil de l'Europe.
- Holec, Henri / Little, David / Richterich, René (1996): Strategies in language learning and use. Studies towards a Common European Framework of reference for language learning and teaching. Strasbourg: Council of Europe Publishing.
- Hufeisen, Britta / Marx, Nicole (Hrsg.) (2007): EuroComGerm - Die sieben Siebe: Germanische Sprachen lesen lernen. Aachen: Shaker.
- Jessner, Ulrike (2006): Linguistic Awareness in Multilinguals. English as a Third Language. Edinburgh: University Press.
- Klein, Horst G. / Rutke, Dorothea (Hrsg.) (2004): Neuere Forschungen zur Europäischen Interkomprehension. Aachen: Shaker (= Editiones EuroCom 21).
- Kleppin, Karin (2006): Selbstreflexion und Selbstevaluation : ein vernachlässigtes Potential bei Aufgaben. In: Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.): Aufgabenorientierung als Aufgabe. Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, S. 102-108.
- Kohonen, Viljo (2000): Student reflection in portfolio assessment: making language learning more visible. In: Babylonia. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen, 1, 15-18.
- Little, David / Hodel, Hans-Peter / Kohonen, Viljo / Meijer, Dick / Perclová, Radka (2008): Préparer les enseignants du Portfolio européen des langues - arguments, matériels et ressources. Livre et CDROM. Strasbourg: Conseil de l'Europe.
- Macaro, Ernesto (2001): Learning strategies in foreign and second language classrooms. London: Continuum.
- 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), Erhältlich unter [http://zif.spz.tu-darmstadt.de/projekt\\_ejournal/jg-09-2/beitrag/Mcintosh2.htm](http://zif.spz.tu-darmstadt.de/projekt_ejournal/jg-09-2/beitrag/Mcintosh2.htm) (30.05.08).
- Meißner, Franz-Joseph (2002): Transfer aus der Sicht der Mehrsprachigkeitsdidaktik. In: Materialien Deutsch als Fremdsprache 65, 128-142.
- Meißner, Franz-Joseph (2004): Sprachenlernen zusammendenken. Aufhören, Rücken an Rücken zu forschen, zu planen, zu lehren und zu lernen. In: Bausch, Karl-Richard / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2004): Mehrsprachigkeit im Fokus. Arbeitspapiere der 24. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, 148-162.

- Meißner, Franz-Joseph (2006): Lernerautonomisierung durch Aufgaben? In: Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.): Aufgabenorientierung als Aufgabe. Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, S. 159-170.
- Meißner, Franz-Joseph / Senger, Ulrike (2001): Vom induktiven zum konstruktiven Lehr- und Lernparadigma. Methodische Folgerungen aus der mehrsprachigkeitsdidaktischen Forschung. In: Meißner, Franz-Joseph / Reinfried, Marcus (Hrsg.): Bausteine für einen neokommunikativen Französischunterricht: Lernerzentrierung, Ganzheitlichkeit, Handlungsorientierung, Interkulturalität, Mehrsprachigkeitsdidaktik. Tübingen: Narr, S. 22-51.
- Morkötter, Steffi (2005): Language Awareness und Mehrsprachigkeit. Frankfurt am Main u. a.: Lang (= Kolloquium Fremdsprachenunterricht 21).
- Newby, David (2007): The European Portfolio for Student Teachers of Languages. In: Babylonia. Themenheft: Projekte des Europäischen Fremdsprachenzentrums in Graz. 3, 23-26.
- Newby, David / Allan, Rebecca / Fenner, Anne-Brit / Jones, Barry / Komorowska, Hanna / Soghikyan, Kristine (2008): Portfolio européen pour les enseignants en langues en formation initiale - Un outil de réflexion pour la formation des enseignants en langues. Strasbourg: Conseil de l'Europe.
- O'Malley, Michael J. / Chamot, Anna Uhl (1990): Learning Strategies in Second Language Acquisition. Cambridge: Cambridge University Press (= Cambridge Applied Linguistics).
- Oxford, Rebecca (1990): Language Learning Strategies: What Every Teacher Should Know. New York: Newbury House.
- Rampillon, Ute (2003): Lernstrategisches Minimalprofil an der Schwelle von L2 und L3. In: Hufeisen, Britta / Neuner, Gerhard (Hrsg.): Mehrsprachigkeitskonzept - Tertiärsprachenlernen - Deutsch nach Englisch, 87-106.
- Rampillon, Ute / Zimmermann, Günther (Hrsg.) (1997): Strategien und Techniken beim Erwerb fremder Sprachen. Ismaning: Hueber (= Forum Sprache).
- Robinson, Peter (ed.) (2002): Individual Differences and Instructed Language Learning. Amsterdam: Benjamins (= Language Learning & Language Teaching 2).
- Saudan, Victor (2007): Ursache Zukunft: Die Neukonzeption des Fremdsprachenunterrichts und ihre Auswirkungen auf die Lehrerinnen- und Lehrerbildung. In: Beiträge zur Lehrerbildung 25/2, 214-222.
- Saudan, Victor / Perregaux, Christiane / Mettler, Monika / Deschoux, Carole-Anne / Sauer, Esther / Ladner, Esther (2005): Lernen durch die Sprachenvielfalt. Schlussbericht zum Projekt JALING Suisse. Apprendre par et pour la diversité linguistique. Rapport final sur le projet JALING Suisse. Bern: EDK. Online: <http://www.edk.ch/dyn/17206.php> (20.07.08).
- Schneider, Günther (2007): Fremdsprachenforschung und die Ausbildung von Fremdsprachenlehrerinnen und -lehrern. In: Beiträge zur Lehrerbildung 25/2, 143-155.
- Skehan, Peter (1989): Individual Differences in Second-Language Learning. London: Arnold.
- Steinhauer, Britta (2006): Transfer im Fremdspracherwerb. Ein Forschungsüberblick und eine empirische Untersuchung des individuellen Transferverhaltens. Frankfurt am Main: Lang (= Europäische Hochschulschriften: Reihe 21, Linguistik 294).
- Svalberg, Agneta M-L. (2007): Language awareness and language learning. In: Language teaching. Surveys and Studies October 4 / 40, 287-308.
- Tönshoff, Wolfgang (1997): Bewusstmachendes Training von Fremdsprachenstrategien. In: Materialien Deutsch als Fremdsprache 46, 96-111.
- Tönshoff, Wolfgang (1998): Zur Rolle des Begriffs ‚Kognition‘ beim Lernen(lernen) von Fremdsprachen. In: Bausch, Karl-Richard / Christ, Herbert / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) Kognition als Schlüsselbegriff bei der Erforschung des Lehrens und Lernens fremder Sprachen, 179-186.
- Vollmer, Helmut J. (1996): Zum Stand der Erforschung des Fremdspracherwerbs in der Bundesrepublik Deutschland - Fokus: Der Lerner. In: Bausch, Karl-Richard / Christ, Herbert / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.): Erforschung des Lehrens und Lernens fremder Sprachen. Zwischenbilanz und Perspektiven. Arbeitspapiere der 16. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, 144-158.
- Vollmer, Helmut J. (1998): Dem Lerner auf der Spur: Kognitive, emotionale und interaktive Aspekte des Fremdspracherwerbs. In: Bausch, Karl-Richard / Christ, Herbert / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) Kognition als Schlüsselbegriff bei der Erforschung des Lehrens und Lernens fremder Sprachen, 197-206.
- Vollmer, Helmut J. et al. (2001): Lernen und Lehren von Fremdsprachen: Kognition, Affektion, Interaktion. Ein Forschungsüberblick. In: Zeitschrift für Fremdsprachenforschung 2/2001, 1-143.
- Vollmer, Helmut (2004): Auf dem Wege zu Mehrsprachigkeit – Ansätze, Erfahrungen, Aufbruch. In: Bausch, Karl-Richard / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2004): Mehrsprachigkeit im Fokus. Arbeitspapiere der 24. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, 238-248.
- Vollmer, Helmut (2005): Das Gespenst der Standardisierung geht um ODER: Lehren Lernen fremder Sprachen auf der Basis von Bildungsstandards. In: Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. (2005): Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand. Arbeitspapiere der 25. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, 258-271.

- Vollmer, Helmut (2006a): Towards a Common European Instrument for Language(s) of Education. Preliminary Study. Strasbourg: Language Policy Division. Online: [http://www.coe.int/t/dg4/linguistic/Source/Vollmer\\_Beacco\\_final\\_EN.doc](http://www.coe.int/t/dg4/linguistic/Source/Vollmer_Beacco_final_EN.doc) (20.07.08).
- Vollmer, Helmut (2006b): Language Across the Curriculum. Preliminary Study. Strasbourg: Language Policy Division. Online: [http://www.coe.int/t/dg4/linguistic/Source/Vollmer\\_LAC\\_EN.doc](http://www.coe.int/t/dg4/linguistic/Source/Vollmer_LAC_EN.doc) (20.07.08).
- Vollmer, Helmut (2007): Language Across the Curriculum – A Way towards Plurilingualism. In: Martyniuk, Waldemar (Hrsg): Towards a Common European Framework of Reference for Languages of School Education. Krakow: Universitas, 177-192.
- Wokusch, Susanne / Lys, Irène (2007): Überlegungen zu einer integrativen Fremdsprachendidaktik. In: Beiträge zur Lehrerbildung 25/2, 168-179.
- Wolff, Dieter (1998): Lernstrategien: Ein Weg zu mehr Lernerautonomie. Online: <http://paedpsych.jk.uni-linz.ac.at:4711/LEHRTEXTE/Wolff98.html> (01.2002).
- Wenden, Anita / Rubin, Joan (Hrsg.) (1987): Learner Strategies in Language Learning. Englewood Cliffs: Prentice-Hall.

## 5 Die Expertenvorschläge für Basisstandards

Im Kapitel 3.6 dieses Berichts werden im Rahmen der HarmoS-Untersuchungen gemessene Kompetenzen im Lesen und Schreiben mit den Expertenvorschlägen für Basisstandards konfrontiert. Dabei entsteht im Grossen und Ganzen das Bild, dass – mit leichten Unterschieden – nur rund 50% der heutigen Sechst- bzw. Neuntklässler/innen die als Basis-, d. h. Mindeststandards definierten schweizerischen Bildungsstandards erreichen würden. Basisstandards sollten aber, sind sie einmal in Kraft, eigentlich von allen Schüler/innen und Schülern der entsprechenden Stufen erreicht werden, sofern sie keine gravierenden Lernschwierigkeiten haben oder unter ausnehmend widrigen Umständen gelernt haben, die ein Erreichen der Standards verunmöglichten.

Das Fremdsprachenkonsortium stand bei seinem Expertenvorschlag vor der schwierigen Situation, dass es aufgrund von Kompetenzen, die bei den Schüler/innen des Jahres 2007 gemessen wurden, die (Mindest-)Kompetenzen von Schüler/innen voraussagen musste, die zum Zeitpunkt der nächsten Messungen in einigen Jahren unter ganz anderen, erfolgversprechenderen Bedingungen gelernt haben werden.

Um den Vorschlag für Basisstandards breiter abzustützen, hat das Fremdsprachenkonsortium am 23./24. November 2007 ein Expertentreffen (Panel) zum Setzen der Basisstandards durchgeführt, an dem die Mitglieder und der wissenschaftliche Beirat des Konsortiums sowie weitere Experten aus dem Bereich der Fremdsprachendidaktik – insgesamt über 20 Personen – teilnahmen.

Grundsätzlich wurde mit Niveaubeschreibungen und Aufgabenbeispielen bzw. Schülerleistungen gearbeitet. Dazu kamen Informationen über die Leistungen der Schüler/innen anlässlich der HarmoS-Tests – insbesondere zum Leseverstehen. Konkret verlief die Entscheidungsfindung bezüglich aller Fertigkeiten nach dem folgenden Muster:

- 1) Die Panel-Teilnehmer/innen wurden mit den erwähnten Materialien konfrontiert und mussten aufgrund der Kompetenzbeschreibungen (Grundlage: *Referenzrahmen* und *IEF/lingualevel*) vorläufig entscheiden, welches Niveau im Minimum von praktisch allen Schüler/innen und Schülern einer künftigen sechsten bzw. neunten Klassen erreicht werden sollte.
- 2) Anschliessend wurden sie mit Aufgaben bzw. Leistungen konfrontiert, die aus der Sicht des Konsortiums diese Niveaus illustrierten. Die erste Frage dazu lautete, ob die gezeigten Aufgaben/Leistungen geeignet seien, das betreffende Niveau zu illustrieren.
- 3) Sobald geeignete Leistungen gefunden waren, hatten die Teilnehmenden Gelegenheit, auf ihre Niveaumentcheidung zurückzukommen bzw. die Diskussion zum „richtigen“ Niveau zu eröffnen.
- 4) Die Entscheidung über das Niveau jedes Basisstandards wurde schliesslich mittels Abstimmung (Handerheben) gefällt.

Nicht für alle Fertigkeiten (insbesondere nicht für das Hörverstehen) konnte dieses Verfahren in allen Schritten durchlaufen werden, es wurden aber alle Entscheidungen zumindest aufgrund von Kompetenzbeschreibungen diskutiert und anschliessend getroffen.

Der Konsens innerhalb der Gruppe war recht hoch. Es gab lediglich einige wenige abweichende Meinungen, und zwar sowohl gegen unten als auch gegen oben. Im Verlauf der Tagung wurde recht schnell klar, dass es der grossen Mehrheit der Expertinnen und Experten nicht um eine Fortschreibung der heutigen Praxis im Fremdsprachenunterricht geht. Vielmehr sehen sie die Bildungsstandards als Möglichkeit, einen deutlichen Schritt vorwärts zu machen, hin zu einem anderen, anwendungsnäheren und zielführenderen Fremdsprachenunterricht.

Als Schwierigkeit erwies sich immer wieder, ganz konsequent nur die Basisstandards im Blick zu behalten. In den Argumenten der Teilnehmenden schien immer wieder auf, dass sie versucht waren, an den „Durchschnittsschüler“ zu denken und daran, was dieser am Ende der jeweiligen Stufe können sollte.

Zur leichteren Orientierung werden die Niveaumentscheidungen des Expertenpanels – d. h. der vollständige Expertenvorschlag – an dieser Stelle noch einmal im Überblick abgebildet:

**Standard Stufe 8/ 1. Fremdsprache**

Referenzniveaus		A1.1	A1.2	A2.1	A2.2	B1.1	B1.2	B2
Harmo S-Niveaus		I <sub>s</sub>	II <sub>s</sub>	III <sub>s</sub>	IV <sub>s</sub>	V <sub>s</sub>		
Kompetenzbereiche	Hörverstehen	→						
	Leseverstehen	→						
	An Gesprächen teilnehmen	→						
	Zusammenhängend sprechen	→						
	Schreiben	→						
	Sprachmittlung	→		→				

**Standard Stufe 8/ 2. Fremdsprache**

Referenzniveaus		A1.1	A1.2	A2.1	A2.2	B1.1	B1.2	B2
Harmo S-Niveaus		I <sub>s</sub>	II <sub>s</sub>	III <sub>s</sub>	IV <sub>s</sub>	V <sub>s</sub>		
Kompetenzbereiche	Hörverstehen	→						
	Leseverstehen	→						
	An Gesprächen teilnehmen	→						
	Zusammenhängend sprechen	→						
	Schreiben	→						
	Sprachmittlung	→						

**Standard Stufe 11: 1. und 2. Fremdsprache**

Referenzniveaus		A1.1	A1.2	A2.1	A2.2	B1.1	B1.2	B2
Harmo S-Niveaus				I <sub>11</sub>	II <sub>11</sub>	III <sub>11</sub>	IV <sub>11</sub>	V <sub>11</sub>
Kompetenzbereiche	Hörverstehen	→						
	Leseverstehen	→						
	An Gesprächen teilnehmen	→						
	Zusammenhängend sprechen	→						
	Schreiben	→						
	Sprachmittlung	→		→				

Abb. 20: Übersicht über die Expertenvorschläge für Basisstandards in den Fremdsprachenfächern

Das Konsortium steht hinter dem Expertenvorschlag, sieht ihn aber gleichzeitig als einen Diskussionsbeitrag, den es offen und unvoreingenommen zu debattieren bereit ist.

## 6 Literatur

Die hier verzeichnete Literatur umfasst einerseits die bereits bei Projekteingabe aufgeführten Titel und andererseits Titel, die bei der Projektarbeit neu hinzu gezogen wurden. Die Literatur zu den überfachlichen Kompetenzen ist – nach interkulturellen und methodischen Kompetenzen getrennt – am Ende von Kap. 4.2.4 verzeichnet. Titel, die über diesen engeren Themenbereich hinaus relevant sind, werden zum Teil im folgenden allgemeinen Literaturverzeichnis noch einmal aufgeführt.

- Adamzik, Kirsten / Antos, Gerd / Jakobs, Eva-Maria (Hg.) (1997): Domänen- und kulturspezifisches Schreiben. Bern: Lang (= Textproduktion und Medium 3).
- Alderson, Charles J. / Banerjee, Jayanti (2001/2002): State-of-the-Art Review: Language testing and assessment. In: Language Teaching 2001 (Part 1) / 2002,4 (Part 2).
- Alderson, Charles J. / Clapham, Caroline / Wall, Dianne (1995): Language Test Construction and Evaluation. Cambridge: Cambridge University Press (= Language Teaching Library).
- Alderson, J. Charles (2000): Assessing Reading. Cambridge: CUP (= Language Assessment).
- Alderson, J. Charles (2004). The Dutch CEF Construct Group: advice on how to develop specifications for tests at each level of the CEF. Online: <http://www.ealta.eu.org/conference/2004/ppt/alderson14may2.ppt> (20.07.08).
- Alderson, J.C. / Figueras, N. / Kuijper, H. / Nold, G. / Takala, S. / Tardieu, C. (2004): The Development of Specifications for Item Development and Classification within The Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Reading and Listening. Online: <http://www.lancs.ac.uk/fss/projects/grid/> (20.07.08).
- ALTE/Association of Language Testers in Europe (o.J. a): The CEFR Grid for Writing Tasks v. 3.1 (analysis) Online: [http://www.coe.int/T/DG4/Portfolio/documents/CEFRWritingGridv3\\_1\\_analysis.doc](http://www.coe.int/T/DG4/Portfolio/documents/CEFRWritingGridv3_1_analysis.doc) (25.01.2007).
- ALTE/Association of Language Testers in Europe (o.J. b): The CEFR Grid for Writing Tasks v. 3.1 (presentation) Online: [http://www.coe.int/T/DG4/Portfolio/documents/CEFRWritingGridv3\\_1\\_presentation.doc](http://www.coe.int/T/DG4/Portfolio/documents/CEFRWritingGridv3_1_presentation.doc) (25.01.2007).
- Altmann, Gerry T. M. (1990): Cognitive models of speech processing. Cambridge, Mass: MIT Press.
- Bachman, Lyle F. (1990): Fundamental considerations in language testing. Oxford: Oxford University Press.
- Bachman, Lyle F. (2000): Modern Language testing at the turn of the century: assuring that what we count counts. In: Language Testing 17 / 1, 1-42.
- Bachman, Lyle F. (2002): Some reflections on task-based language performance assessment. In: Language Testing 19, (3), 453-476.
- Bachman, Lyle F. (2004): Statistical analyses for language assessment. Cambridge: CUP.
- Bachman, Lyle F. / Cohen, Andrew D. (Hg.) (1998): Interfaces Between Second Language Acquisition and Language Testing Research. Cambridge: University Press (= Applied Linguistics).
- Bachman, Lyle F. / Palmer, Adrian S. (1996): Language Testing in Practice. Designing and Developing Useful. Language Tests. Oxford: University Press.
- Baden-Württemberg, Ministerium für Kultus, Jugend und Sport Baden-Württemberg (Hg.) (2004). Bildungspläne. Online: <http://www.bildungsstandards-bw.de/>
- Baer, Matthias / Jurt, Ueli / Fuchs, Michael, Reber-Wyss, Monika, und Nussbaum, Thomas (1997): Wie unterschiedlich gute und verschieden alte Textverfasser/innen Schreibaufgaben bewältigen und Möglichkeiten, Schüler/innen beim Aufbau der Schreibkompetenz zu unterstützen. St. Gallen: UVK Fachverlag für Wissenschaft und Studium.
- Banerjee, J. (2004): Qualitative analysis methods. In: Council of Europe (Hg.): Reference Supplement to the Preliminary Pilot version of the Manual for Relating Language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment, Section D. Strasbourg: Council of Europe (DGIV/EDU/LANG (2004) 13). Online: [www.coe.int/portfolio](http://www.coe.int/portfolio) > Documentation (26.01.2007).
- Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2006): Aufgabenorientierung als Aufgabe. Tübingen: Narr.
- Bausch, Karl-Richard / Burwitz-Melzer, Eva / Königs, Frank G. (Hrsg., 2005): Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand. Tübingen: Narr.
- Bausch, Karl-Richard / Christ, Herbert / Königs, Frank G. / Krumm, Hans-Jürgen (Hrsg.) (2002): Der Gemeinsame europäische Referenzrahmen in der Diskussion. Tübingen: Narr.
- Beacco, Jean-Claude (2004): Une proposition de référentiel pour les compétences culturelles dans les enseignements de langues. In : Beacco, Jean-Claude/ Bouquet, Simon/ Porquier, Rémy (2004): Niveau B2 pour le Français. Textes et Références. Paris: Didier, 251-286.

- Beacco, Jean-Claude / Byram, Michael (2003): Guide for the development of Language Education Policies in Europe. From Linguistic Diversity to Plurilingual Education. Main Version. Draft 1. Strasbourg: Council of Europe. [http://www.coe.int/T/E/Cultural\\_Co-operation/education/Languages/Language\\_Policy](http://www.coe.int/T/E/Cultural_Co-operation/education/Languages/Language_Policy)
- Beck, Bärbel / Klieme, Eckhard (2003): DESI: Eine Längsschnittstudie zur Untersuchung des Sprachunterrichts in deutschen Schulen. In: Finkbeiner, C. / Fehling, S. (Hg.): Evaluation im Brennpunkt - Fremdsprachen lernen und lehren 17 / 3, 380-395.
- Beck, Bärbel / Klieme, Eckhard (Hrsg., 2007): Sprachliche Kompetenzen. Konzepte und Messung. DESI-Ergebnisse Band 1. Weinheim und Basel: Beltz.
- Bereiter, Carl / Scardamalia, Marlene (1987): The Psychology of Written Composition. Hillsdale: Erlbaum.
- Blue, George M. / Milton, James / Saville, Jane (Hg.) (2000): Assessing English for Academic Purposes. Bern: Lang.
- Bolton, Sibylle (2003): Qualitätsmanagement bei den Prüfungen des Association of Language Testers in Europe (ATE). In: AKS-Rundbrief, Fremdsprachen und Hochschule. Thema: Qualitätssicherung im Sprachtesten 69, 9-22.
- Bond, Trevor G. / Fox, Christine M. (2001): Applying the Rasch Model. Fundamental Measurement in the Human Sciences. Mahwah, N.J.; London: Erlbaum.
- Börner, Wolfgang / Vogel, Klaus (Hg.) (1992): Schreiben in der Fremdsprache. Prozess und Text, Lehren und Lernen. Bochum: AKS (= FLF Fremdsprachen in Lehre und Forschung 10).
- Bregy, Anne-Lore / Revaz, Nadia (2001): Evaluation des enseignements / apprentissages bilingues en classes de 4<sup>ème</sup> primaire de Sierre. Année scolaire 1999 / 2000. Rapport intermédiaire. Neuchâtel: IRDP.
- Brindley, Geoff (2001): Outcomes-based assessment in practice: some examples and emerging insights. In: Language Testing 18, 4.
- Brohy, Claudine (1999): L'enseignement bilingue / l'enseignement précoce des langues secondes en Suisse romande. Neuchâtel: IRDP (= IRDP Regards).
- Brohy, Claudine / Pannatier, Monique (2000): L' évaluation dans l' enseignement immersif - la quadrature du cercle? In: Babylonia. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen. Réfléchir sur les expériences d' apprentissage - en évaluer les résultats 1, 33-35.
- Brown, Gillian / Malmkjaer, Kirsten / Pollitt, Alastair / Williams, John (eds.) (1994): Language and understanding. Oxford: University Press (= Applied Linguistics).
- Brown, Gillian / Malmkjaer, Kirsten / Williams, John (Hg.) (1996): Performance and Competence in Second Language Acquisition. Cambridge: University Press.
- Brown, James Dean / Hudson, Thom (2002): Criterion-referenced Language Testing. Cambridge: University Press.
- Buck, Gary (2001): Assessing Listening. Cambridge: University Press.
- Burwitz-Melzer, Eva (2006): Interkulturelles und sprachliches Lernen mit fremdsprachlichen literarischen Texten: Zwei zentrale Elemente eines neuen Lesekompetenzmodells. In: FLUL 35, 104-120.
- Bygate, Martin / Skehan, Peter / Swain, Merrill (2001): Researching Pedagogic Task. Second Language Learning, Teaching and Testing. Harlow: Pearson Education.
- Byram, Michael (1997): Teaching and Assessing Intercultural Communicative Competence. Clevedon: Multilingual Matters.
- Byram, Michael / Neuner, Gerhard / Zarate, Geneviève (1997): La compétence socioculturelle dans l' apprentissage et l' enseignement des langues. Vers un Cadre Européen de référence pour l'apprentissage et l'enseignement des langues vivantes: études préparatoires. Strasbourg: Editions du Conseil de l' Europe (= Education).
- Byram, Michael / Nichols, Adam / Stevens, David (Hg.) (2001): Developing intercultural competence in practice. Clevedon: Multilingual Matters.
- Byram, Michael / Tost Planet, Manuel (Hg.) (2000): Identité sociale et dimension européenne: la compétence inter-culturelle par l' apprentissage des langues vivantes. Graz / Strasbourg: European Centre for Modern Languages / Conseil de l'Europe.
- Byram, Michael / Tost Planet, Manuel (Hg.) (2000): Sociale Identity and European Dimension: Intercultural Competence Through Foreign Language Learning. Graz / Strasbourg: European Centre for Modern Languages / Conseil de l'Europe.
- Byram, Michael / Zarate, Geneviève (Hg.) (1997): The sociocultural and intercultural dimension of language learning and teaching. Strasbourg: Council of Europe Publishing (= Education).
- Byram, Michael / Zarate, Geneviève (Hg.) (1997): The sociocultural and intercultural dimension of language learning and teaching. Strasbourg: Council of Europe Publishing (= Education).
- Canale, M. / Swain, M. (1980): Theoretical bases of communicative approaches to second language teaching and testing. In: Applied Linguistics 1 (1), 1-47.
- Candlin, Christopher N. / Hyland, Ken (eds.) (1999): Writing: texts, processes and practices. London / New York: Longman (= Applied Linguistics and Language Study).
- Caspari, Daniela (2005): Von Bildungsstandards zu Lehrplänen. In: Bausch, K.-R. / Burwitz-Melzer, E. / Königs, F.G. / Krumm, H.-J. (Hrsg.): Bildungsstandards für den Fremdsprachenunterricht auf dem Prüfstand. Tübingen, 67-77.

- Chapelle, C. (1998): Construct definition and validity inquiry in SLA research. In: Bachman, L. / Cohen, A. (Hg.): Interfaces between second language acquisition and language testing research. Cambridge: CUP, 32-70.
- Chapelle, C. / Grabe, W. / Berns, M. (1997): Communicative language proficiency: definitions and implications for TOEFL 2000. Monograph Series. Princeton, N.J.: ETS.
- Chardenet, Patrick (1999): De l'activité évaluative à l'acte d'évaluation. Approche théorique et pratique communicationnelle. Paris: L'Harmattan (= Savoir et formation).
- Charge, Nick / Williams, Andy (2003): IELTS - Quality Assurance in English Language Testing. In: AKS-Rundbrief, Fremdsprachen und Hochschule. Thema: Qualitätssicherung im Sprachtesten 69, 23-27.
- Charpié, Nathalie (2002): Expérimentation de l'enseignement de l'allemand par immersion dans le canton du Jura. Neuenburg: IRDP (= IRDP document de travail).
- Christ, Ingeborg (2003): Auf dem Wege zu einer neuen Evaluationskultur im Fremdsprachenunterricht. In: Neusprachliche Mitteilungen aus Wissenschaft und Praxis 3, 157-169.
- CIIP (2003). Déclaration de la CIIP relative à la politique de l'enseignement des langues en Suisse romande. Online: [www.ciip.ch/ciip/pdf/cp030403\\_3%20.pdf](http://www.ciip.ch/ciip/pdf/cp030403_3%20.pdf)
- Clapham, Caroline / Corson, David (eds.) (1997): Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment. Dordrecht: Kluwer Academic Publishers.
- Conseil de l'Europe (2001). Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer. Strasbourg: Conseil de l'Europe / Didier (= Apprentissage des langues et citoyenneté européenne).
- Council of Europe (2001). Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: CUP. Online: [www.coe.int/portfolio](http://www.coe.int/portfolio) >Documentation
- Council of Europe (2002a). Common European Framework of Reference for Languages: Learning, teaching, assessment. Guide for Users. Strasbourg: Council of Europe (= Language Policies).
- Council of Europe (2002b). Common European Framework of Reference for Languages: Learning, teaching, assessment. Case studies (Hg. Ch. Alderson). Strasbourg: Council of Europe (= Language Policies).
- Council of Europe (2003). Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEF). Manual, Preliminary Pilot Version. Strasbourg: Language Policies. Online: [http://www.coe.int/T/DG4/Linguistic/Manuel1\\_EN.asp](http://www.coe.int/T/DG4/Linguistic/Manuel1_EN.asp). (20.07.2008).
- Council of Europe (im Druck). Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). Manual. Strasbourg: Language Policies.
- Council of Europe (2004). Relating language examinations to the Common European framework of reference for languages (CEF). Reference supplement to the preliminary pilot version of the manual. Strasbourg: Language Policies. Online: [http://www.coe.int/T/DG4/Linguistic/Manuel1\\_EN.asp](http://www.coe.int/T/DG4/Linguistic/Manuel1_EN.asp). (20.07.2008).
- Cushing Weigle, Sara (2002): Assessing Writing. Cambridge: University Press (= Cambridge Language Assessment Series).
- Davidson, Fred / Lynch, Brian K. (2002): Testcraft: a teacher's guide to writing and using language test specifications. New Haven: Yale University Press.
- Davies, A. (1989): Communicative competence as language use. In: Applied Linguistics, 10 (2), 157-170.
- Dayez, Yves (2001): Guide des sujets du DELF et du DALF. Connaître, concevoir, évaluer. Paris: Didier (= Les cahiers du CIEP).
- de Pietro, Jean-François / Lüdi, George (Hg.) (1991): Modèles d'évaluation du français langue seconde dans l'éducation publique. Dossier établi par le département de linguistique française du Séminaire des langues romanes de l'Université de Bâle. Basel.
- Demierre-Wagner, Andrea / Schwob, Irène / Ducrey, François (2004): L'expérience pilote de l'enseignement bilingue à l'école primaire en Valais romand. In: bulletin vals-asla 79.
- Depover, Christian / Noël, Bernadette (éds) (1999): L'évaluation des compétences et des processus cognitifs. Modèles, pratiques et contextes. Bruxelles: De Boeck (= Pédagogies en développement).
- Douglas, Dan (2000): Assessing Languages for Specific Purposes. Cambridge: CUP (= Language Assessment).
- EDK, Schweizerische Konferenz der kantonalen Erziehungsdirektoren (2004). Sprachenunterricht in der obligatorischen Schule. Strategie der EDK und Arbeitsplan für die gesamtschweizerische Koordination. Online: <http://edudoc.ch/getfile.py?docid=1987&name=7-2d&format=pdf&version=1> (20.07.08).
- EDK, Schweizerische Konferenz der kantonalen Erziehungsdirektoren, Expertengruppe (1998). Sprachenkonzept Schweiz: Welche Sprachen sollen die Schülerinnen und Schüler der Schweiz während der obligatorischen Schulzeit lernen? Online: <http://www.romsem.unibas.ch/sprachenkonzept>
- EDK, Schweizerische Konferenz der kantonalen Erziehungsdirektoren (2004). HARMOS. Zielsetzungen und Konzeption. Bern: EDK Generalsekretariat, Koordinationsbereich „Obligatorische Schule“. Online: <http://www.edk.ch/dyn/11968.php> (20.07.08).
- EDK, Schweizerische Konferenz der kantonalen Erziehungsdirektoren (2001). Empfehlungen zur Koordination des Sprachenunterrichts in der obligatorischen Schule. ENTWURF 1.6.2001

- Ehlers, Swantje (2006): Entwicklung von Lesekompetenz in der Fremdsprache. In: *Babylonia* 3-4, 31-38.
- Ellis, Rod (2000): Task-based research and language pedagogy. In: *Language Teaching Research* 4,3 (2000), 193–220.
- Ellis, Rod (2003): *Task-based Language and Teaching*. Oxford: OUP.
- Ellis, Rod (Hrsg.) (2005): *Planning and Task Performance in a Second Language*. Amsterdam u. a.: John Benjamins.
- Ellis, Rod / Barkhuizen, Gary (2005): *Analysing Learner Language*. Oxford: OUP.
- Engel, Ulrich (1988): *Deutsche Grammatik* (2. Aufl.). Heidelberg: Julius Groos.
- Eriksson, Brigit / Le Pape Racine, Christine / Reutener, Hans / Serra, Cecilia / Stern, Otto (1998): *Français - Allemand: Apprendre en deux langues à l' école secondaire. Rapport de valorisation*. Aarau: CSRE (= L' efficacité de nos systèmes de formation).
- Eriksson, Brigit / Le Pape Racine, Christine / Reutener, Hans / Serra, Cecilia / Stern, Otto (1999): *Französisch - Deutsch: Zweisprachiges Lernen auf der Sekundarstufe I*. Chur / Zürich: Rüegger (= Wirksamkeit unserer Bildungssysteme. NFP 33).
- Europarat (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin u. a.: Langenscheidt. Online: <http://www.goethe.de/referenzrahmen> (30.05.08)
- Finkbeiner, C. / Fehling, S. (Hg.) (2003): *Evaluation im Brennpunkt - Fremdsprachen lernen und lehren*. Landau: Empirische Pädagogik (= Empirische Pädagogik 17 / 3).
- Fuchs, Gabriela / Werlen, Iwar (1995). *Zwischenbericht zum Projekt "Zweisprachige Schule Brig-Glis" an den Schulpräsidenten zuhänden des Gemeinderates der Stadtgemeinde Brig-Glis*. Brig-Glis/ Bern: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Fuchs, Gabriela / Werlen, Iwar (1996). *Zweisprachige Schule Brig-Glis. Zwischenbericht zum Projekt "Zweisprachige Schule Brig-Glis" an den Schulpräsidenten zuhänden des Gemeinderates der Stadtgemeinde Brig-Glis*. Brig-Glis, Universitäres Forschungszentrum für Mehrsprachigkeit.
- Fuchs, Gabriela / Werlen, Iwar (1997): *Zweisprachige Schule Brig-Glis. Zwischenbericht nach zwei Jahren teilimmersivem Unterricht. Schuljahr 1995/96 und 1996/97. Bericht an den Schulpräsidenten zuhänden des Gemeinderates der Stadtgemeinde Brig-Glis*. Brig-Glis: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Fuchs, Gabriela / Werlen, Iwar (1998): *Zweisprachige Schule Brig-Glis. Schuljahr 1997/98 Kurzfassung*. Bern, Brig-Glis: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Fulcher, Glenn (2003): *Testing Second Language Speaking*. London: Longman / Pearson Education (= Applied Linguistics and Language Study).
- Garman, Mike (1990): *Psycholinguistics*. Cambridge: Cambridge University Press.
- Gassner, Otmar / Horak, Angela / Mewald, Claudia / Moser, Wolfgang / Schober, Michael / Stefan, Ferdinand / Valsky, Claudia (2005): *Bildungsstandards in Österreich. Fremdsprachen. Englisch 8. Schulstufe. Version September'05*. Wien: Bundesministerium für Bildung, Wissenschaft und Kultur.
- Gernsbacher, Morton Ann (Hg.) (1994): *Handbook of psycholinguistics*. San Diego: Academic Press.
- Gernsbacher, Morton Ann / Givón, Talmy (Hg.) (1995): *Coherence in spontaneous text*. Amsterdam: John Benjamins (=Typological studies in language; 31).
- Gick, Cornelia / Langner, Michael / Lenz, Peter / Schneider, Günther (1993): *Untersuchungen zur Evaluation der mündlichen und schriftlichen Kommunikationsfähigkeit*. In: Dossier, EDK: Analyse von Evaluationsmodellen für Deutsch und Französisch L2 am Ende der obligatorischen Schulzeit 23/2, 238-322.
- Glaboniat, Manuela (1998): *Kommunikatives Testen im Bereich Deutsch als Fremdsprache. Eine Untersuchung am Beispiel des Österreichischen Sprachdiploms Deutsch*. Innsbruck / Wien: Studien (= Theorie und Praxis - Österreichische Beiträge zu Deutsch als Fremdsprache, Serie B 2).
- Glaboniat, Manuela / Müller, Martin / Rusch, Paul / Schmitz, Helen / Wertenschlag, Lukas (2005): *Profile deutsch. Gemeinsamer europäischer Referenzrahmen. Lernzielbestimmungen - Kannbeschreibungen - Kommunikative Mittel - Niveau A1 - A2 - B1 - B2 - C1 - C2 mit CD-ROM*. München: Langenscheidt.
- Gajo, Laurent / Matthey, Marinette / Py, Bernard (1999): *Die Schule, die Schulkameradinnen und -kameraden, die Familie ... Das Lernen von Französisch und Deutsch in der Westschweiz durch fremdsprachige Kinder*. Bern: NFP33 (= Umsetzungsbericht).
- Grabe, William / Kaplan, Robert (1996): *Theory and practice of writing: an applied linguistic perspective*. London: Longman (= Applied linguistics and language study).
- Gregori, Gian Peder / Todisco, Vincenzo (2003). *Evaluation des Zweitsprachunterrichts (ZSU) Italienisch in den Primarschulen des deutschsprachigen Teils des Kantons Graubünden. Schlussbericht*. Chur, Pädagogische Fachhochschule Graubünden.
- Grotjahn, Rüdiger (1995): *Der C-Test: State of Art*. In: *Zeitschrift für Fremdsprachenforschung (ZFF)* 6/2, 37-60.
- Grotjahn, Rüdiger (1995): *Zweitsprachliches Leseverstehen: Grundlagen und Probleme der Evaluation*. In: *Die Neueren Sprachen* 5, 533-555.
- Grotjahn, Rüdiger (1997): *Der C-Test: Neuere Entwicklungen*. In: Gardenghi, Monica / O'Connell, Mary: *Prüfen, Testen, Bewerten im modernen Fremdsprachenunterricht* 6, 117-128.

- Grotjahn, Rüdiger (Hg.) (2002): Der C-Test. Theoretische Grundlagen und praktische Anwendungen. Band 4. Bochum: AKS (= FLF Fremdsprachen in Lehre und Forschung 32).
- Grotjahn, Rüdiger (2000): Determinanten der Schwierigkeit von Leseverstehensaufgaben: Theoretische Grundlagen und Konsequenzen für die Entwicklung des TESTDAF. In: Bolton, Sybille, Hrsg., TESTDAF: Grundlagen für die Entwicklung eines neuen Sprachtests. Beiträge aus einem Expertenseminar. München: Goethe-Institut.
- Habel, Christopher et al. (Hg.) (1996): Perspektiven der kognitiven Linguistik: Modelle und Methoden. Opladen: Westdeutscher Verlag (=Psycholinguistische Studien).
- Hamp-Lyons, Liz (Hg.) (1991): Assessing second language writing in academic contexts. Norwood: Ablex Publishing Corporation.
- Harsch, Claudia (2006): Der Gemeinsame europäische Referenzrahmen: Leistung und Grenzen. Die Bedeutung des Referenzrahmens im Kontext der Beurteilung von Sprachvermögen am Beispiel des semikreativen Schreibens im DESI-Projekt. Inaugural-Dissertation. Online: [http://www.opus-bayern.de/uni-augsburg/frontdoor.php?source\\_opus=368](http://www.opus-bayern.de/uni-augsburg/frontdoor.php?source_opus=368) (20.07.08).
- Hayes, John R. / Flower, Linda S. (1980): Identifying the Organization of Writing Processes. In: Gregg, Lee W./Steinberg, Edwin R. (Hrsg.): Cognitive processes in writing. Hillsdale: Erlbaum, 3-30.
- Henrici, Gert / Zöfgen, Ekkehard (Hg.) (2001): Themenheft: Leistungsmessung und Leistungsevaluation. In: Fremdsprachen lehren und lernen (FLuL) 30.
- Herrmann, Theo / Grabowski, Joachim (1995): Sprechen: Psychologie der Sprachproduktion. Heidelberg: Spektrum.
- Hyland, Ken (2002): Teaching and Researching Writing, Applied Linguistics in Action. Harlow: Pearson Education (= Applied Linguistics in Action).
- Hyland, Ken (2003): Second Language Writing. Cambridge: ELT (= Language Education).
- Ioannou-Georgiou, Sophie / Pavlou, Pavlos (2003): Assessing Young Learners. Oxford: University Press (= Ressource books for Teachers).
- IRDP (1998): Le point sur la recherche. Enseignement bilingue. Bulletin d'information janvier 1998. In: Mehrsprachigkeit.
- Jones, Neil (2002). Relating the ALTE Framework to the Common European Framework of reference. In Council of Europe. Common European Framework of Reference for Languages: Learning, teaching, assessment. Case studies (Hg. Ch. Alderson). Strasbourg: Council of Europe, 167 – 183. Online: [http://www.coe.int/T/DG4/Portfolio/documents/case\\_studies\\_CEF.doc](http://www.coe.int/T/DG4/Portfolio/documents/case_studies_CEF.doc) (20.07.08).
- Kaftandjieva, Felianka / Takala, Sauli (2002). Council of Europe Scales of Language Proficiency: A validation study. In Council of Europe. Common European Framework of Reference for Languages: Learning, teaching, assessment. Case studies. (Hg. Ch. Alderson). Strasbourg: Council of Europe, 106 – 129. Online: [http://www.coe.int/T/DG4/Portfolio/documents/case\\_studies\\_CEF.doc](http://www.coe.int/T/DG4/Portfolio/documents/case_studies_CEF.doc) (20.07.08).
- Kästner, Uwe (1997): Freies Schreiben in der Fremdsprache - Prozesse und ihre Didaktik. Untersuchung zur Textproduktion im fremdsprachlichen Deutschunterricht auf dem Hintergrund neuropsychologischer Theorien. Bochum: AKS (= FLF Fremdsprachen in Lehre und Forschung 19).
- Kintsch, Walter (1998): Comprehension. A paradigm of cognition. Cambridge: Cambridge University Press.
- Klieme, Eckhard et. al.. (2003): Zur Entwicklung nationaler Bildungsstandards. Eine Expertise. Bonn: Bundesministerium für Bildung und Forschung (BMBF). Online: [http://www.bmbf.de/pub/zur\\_entwicklung\\_nationaler\\_bildungsstandards.pdf](http://www.bmbf.de/pub/zur_entwicklung_nationaler_bildungsstandards.pdf) (30.05.08)
- KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hg.) (2004). Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss. Neuwied: Luchterhand. Online: <http://www.kmk.org/doc/beschl/aschulw.htm> (25.06.08)
- Kniffka, Gabriele / Üstünsöz-Beurer, Dörthe (2001): TestDaF: Mündlicher Ausdruck. Zur Entwicklung eines kassetten-gesteuerten Testformats. In: Henrici, Gert / Zöfgen, Ekkehard (Hg.): Fremdsprachen lehren und lernen (FLuL) 30, 127-149.
- Knobloch, Clemens (1994): Sprache und Sprechfähigkeit: Sprachpsychologische Konzepte. Tübingen: Niemeyer.
- Kobayashi, Miyoko (2002): Method effects on reading comprehension test performance: text organization and response format. In: *Language Testing*, 19, 193-220.
- Krause, Wolf-Dieter / Sändig, Uta (2002): Testen und Bewerten kommunikativer Leistungen im Unterricht Deutsch als Fremdsprache. Linguistische Grundlagen und didaktische Angebote. Frankfurt/M. u. a.: Lang (= Sprache - System und Tätigkeit 43).
- Krumm, Hans-Jürgen (Hg.) (1999): Erfahrungen beim Schreiben in der Fremdsprache Deutsch. Untersuchungen zum Schreibprozess und zur Schreibförderung im Unterricht mit Studierenden. Innsbruck / Wien: Studien (= Theorie und Praxis - Österreichische Beiträge zu Deutsch als Fremdsprache, Serie B 4).
- Kunnan, Antony John (ed.) (2000): Fairness and Validation in Language Assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida. Cambridge: University Press (= Studies in Language Testing 9).

- Kurtz, Jürgen (2001): Improvisierendes Sprechen im Fremdsprachenunterricht. Eine Untersuchung zur Entwicklung spontansprachlicher Handlungskompetenz in der Zielsprache, Tübingen: Gunter Narr Verlag (gnv). Labudde, Peter (Hrsg.) (2007): Bildungsstandards am Gymnasium. Korsett oder Katalysator? Bern: hep / Pädagogische Hochschule.
- Labudde, Peter (Hrsg.) (2007): Bildungsstandards am Gymnasium. Korsett oder Katalysator? Bern: hep / Pädagogische Hochschule.
- Language Policy Division / Division des Politiques Linguistiques (eds.) (2004): Relating language examinations to the CEF Reference reading items and writing tasks. Preliminary pilot version of the Manual and related documents. Relier les examens de langues au CECR. Items et tâches de référence pour des tests de lecture et d'écriture. Avant-projet de Manuel et documents corrélatifs. Strasbourg: Conseil de l'Europe. [http://www.coe.int/T/DG4/Portfolio?L=E&M=/main\\_pages/illustrationse.html](http://www.coe.int/T/DG4/Portfolio?L=E&M=/main_pages/illustrationse.html) (20.07.08)
- Language Testing (2000). Special Issue on Assessing Young Language Learners. Vol. 17, iss. 2.
- Lenz, Peter (2000): Erfahrungen mit dem Europäischen Sprachenportfolio in der Schweiz. In: *Babylonia*. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen. Réfléchir sur les expériences d'apprentissage - en évaluer les résultats 1, 23-28.
- Lenz, Peter (2002): Mehrsprachigkeit fördern und sichtbar machen - das Europäische Sprachenportfolio. In: Schneider, Günther / Clalüna, Monika (Hg.): Bulletin suisse de linguistique appliquée. VALS/ASLA. Sonderheft: Mehrsprachigkeit und Deutschunterricht: Thesen, Beiträge und Berichte aus der Sektionsarbeit an der XII. IDT - Internationalen Tagung der Deutschlehrerinnen und Deutschlehrer 2001 in Luzern Sonderheft, 151-159.
- Lenz, Peter (2004a): The European Language Portfolio - A Tool for Comprehensive Assessment and the Promotion of Plurilingualism. In: Bräuer, Gerd / Sanders, Karen (ed): *New Visions in Foreign and Second Language Education*, 123-152.
- Lenz, Peter (2004b): The European Language Portfolio. In: Morrow, Keith (Hg.): *Insights from the Common European Framework*. Oxford: Oxford University Press, S. 22-31.
- Lenz, Peter (2006): Überlegungen zur Sprachkompetenzbeschreibung und Testvalidierung im Projekt HarmoS/Fremdsprachen. In: Mondada, Lorenza / Pekarek Doehler, Simona (Hrsg.): Bulletin suisse de linguistique appliquée. VALS / ASLA. La notion de compétence: études critiques 84, 191-227.
- Lenz, Peter / Schneider, Günther (2002): Developing the Swiss model of the European Language Portfolio. In: Council of Europe / Alderson, J.C. (Hg.): *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Case Studies*. Strasbourg: Council of Europe Publishing 2002, 68-86.
- Lenz, Peter / Schneider, Günther (2004a). Introduction to the bank of descriptors for self-assessment in European Language Portfolios. Strasbourg: Council of Europe, Language Policy Division. Online: [www.coe.int/portfolio](http://www.coe.int/portfolio) (01.02.08)
- Lenz, Peter / Schneider, Günther (2004b). A bank of descriptors for self-assessment in European Language Portfolios. Strasbourg: Council of Europe, Language Policy Division. Online: [www.coe.int/portfolio](http://www.coe.int/portfolio) (01.02.08)
- Lenz, Peter/ Studer, Thomas (2004): Sprachkompetenzen von Jugendlichen einschätzbar machen. In: *Babylonia*. 2, 21-25.
- Lenz, Peter; Studer, Thomas (2007): *lingualevel. Instrumente zur Evaluation von Fremdsprachenkompetenzen*. Hrsg. v. Bildungsdirektoren-Konferenz Zentralschweiz (BKZ) / Nordwestschweizerische Erziehungsdirektoren-Konferenz (NW-EDK) / Erziehungsdirektoren-Konferenz der Ostschweizer Kantone und des Fürstentums Liechtenstein (EDK-Ost). Bern: Schulverlag bmv.
- Levelt, Willem (1989): *Speaking. From Intention to Articulation*. Cambridge MA: Bradford Book.
- Linacre, J. M. (2007). A user's guide to Facets: Rasch-measurement computer program. Chicago. Online: [www.winsteps.com/facets.htm](http://www.winsteps.com/facets.htm) (01.02.08)
- Linacre, J. M. (2007). A user's guide to WINSTEPS-MINISTEP: Rasch-model computer programs. Chicago. Online: [www.winsteps.com/winsteps.htm](http://www.winsteps.com/winsteps.htm) (01.02.08).
- Linacre, John M. (1992). *FACETS: A Computer Program for the Analysis of Multi-Faceted Data*. Chicago: MESA Press.
- Little, David (2003). The Common European Framework: principles, challenges, issues. *Neusprachliche Mitteilungen* 56/3 (2003), 130 – 140.
- Long, Michael H. (2007): *Problems in SLA*. London: Erlbaum.
- Lüdi, Georges (2003). Konzepte für mehrsprachige Repertoires als Zielvorstellungen für die Schule. Überlegungen zu einem „Gesamtsprachenkonzept“ für die Schulen in der Schweiz. In: G. Schneider / M. Clalüna (Hg.). *Mehr Sprache – mehrsprachig – mit Deutsch. Didaktische und politische Perspektiven*. München: iudicium, 276 – 291.
- Lüdi, Georges / Pekarek, Simona / Saudan, Victor (1999): *Apprentissage du français à l'intérieur et à l'extérieur de l'école*. Bern / Aarau: NFP (= PNR 33. Rapport de valorisation).
- Lüdi, Georges / Pekarek-Doehler, Simona / Saudan, Victor (2001): *Französischlernen in der Deutschschweiz. Zur Entwicklung der diskursiven Fähigkeiten innerhalb und ausserhalb der Schule*. Chur / Zürich: Rüegger (= NFP 33 - Wirksamkeit unserer Bildungssysteme).
- Lumley, Tom (2005): *Assessing Second Language Writing*. Frankfurt (Main): Peter Lang.
- Luoma, Sari (2004): *Assessing Speaking*. Cambridge: University Press.

- MacWhinney, Brian (1997): Second language acquisition and the Competition Model. In: Annette M. B. de Groot / Judith F. Kroll (Hg.): *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah, NJ: Lawrence Erlbaum, 113-142.
- Maradan, Olivier (2004): Stratégie de la CDIP et programme de travail pour la coordination de l'enseignement des langues à l'école obligatoire en Suisse. In: *bulletin vals-asla* 79.
- Matthey, Marinette (1996): *Apprentissage d'une langue et interaction verbale. Sollicitation, transmission et construction de connaissances linguistiques en situation exolingue*. Bern u. a.: Lang (= *Exploration: Recherches en sciences de l'éducation*).
- McKay, Penny (2006): *Assessing Young Language Learners*. Cambridge: CUP (= *Language Assessment*).
- McNamara, Tim (2000): *Language Testing*. Oxford: University Press (= *Oxford Introductions to Language Study*).
- McNamara, Tim (1996): *Measuring second language performance*. Harlow: Longman.
- Merz-Grötsch, Jasmin (2000): *Schreiben als System, Bd 1: Schreibforschung und Schreibdidaktik. Ein Überblick*. Freiburg / Breisgau: Fillibach.
- Messick, S. (1989): *Validity*. In: Linn, R.L. (Hg.) *Educational measurement*. 3. Aufl. New York: American Council on Education/Macmillan, 13-103.
- Messick, S. (1994): *The interplay of evidence and consequences in the validation of performance assessments*. In: *Educational Researcher* 23(2), 13–23.
- Messick, S. (1996): *Validity and washback in language testing*. Princeton: ETS (Research Report).
- Milanovic, Michael (1996): *Guide for Examiners. Draft 1. (Modern Languages: Learning, Teaching, Assessment. A Common European Framework of reference)*. Strasbourg: Council of Europe (= *Language Learning for European Citizenship CC-LANG (96)10 rev.*). Online: <http://www.coe.int/T/DG4/Portfolio/documents/Guide%20October%202002%20revised%20version1.doc> (20.07.08).
- Mislevy, R.J. / Almond, R.G. / Lukas, J.F. (2004): *A brief introduction to evidence-centered design*. CSE Report 632. Los Angeles: National Center for Research on Evaluation.
- Mislevy, Robert J., Steinberg, Linda S. & Almond, Russell G. (2002): *Design and analysis in task-based language assessment*. In: *Language Testing*, 477-496.
- Molinaro, Agnese (2000): *Evaluation des compétences cognitives des élèves en situation d'apprentissage bilingue*. In: *L'école valdotaine. Cahier pédagogique* 46, 51-53.
- Morrow, K. (1979): *Communicative language testing: revolution or evolution?* In: Brumfit, C.J. and Johnson, K. (Hg.): *The communicative approach to language teaching*. Oxford: Oxford University Press, 143–57.
- Morrow, Keith (Hg.) (2004). *Insights from the Common European Framework*. Oxford: Oxford University Press.
- Moser, Urs (2000): *Evaluation in den 'Fremdsprachen': Chancen und Grenzen*. In: *Babylonia. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen. Réfléchir sur les expériences d'apprentissage - en évaluer les résultats* 1, 10-14.
- Moser, Urs (2003): *Klassencockpit im Kanton Zürich Ergebnisse einer Befragung von Lehrerinnen und Lehrern der 6. Klassen über ihre Erfahrungen im Rahmen der Erprobung von Klassencockpit im Schuljahr 2002/03. Bericht zuhanden der Bildungsdirektion des Kantons Zürich*.
- Moser, Urs / Berweger, Simone (2003): *Bildungsmonitoring Schweiz. Lehrplan und Leistungen. Thematischer Bericht der Erhebung PISA 2000*. Neuchâtel: Bundesamt für Statistik (BFS) und Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).
- Müller-Hartmann, A. & Schocker-v. Difturth, M. (Hrsg.) (2005). *Aufgabenorientierung im Fremdsprachenunterricht. Task Based Language Learning and Teaching. Festschrift für M. Legutke*. Tübingen: Narr.
- Nidegger, Christian (Hg.) (2003): *A la fin de la 6e primaire. Enquête auprès de la volée 2000 des élèves de 6P. Acquis et compétences des élèves, représentations des parents et des enseignants*. Genève: Service de la Recherche en Education.
- Nold, Günter (Hg.) (1992): *Lernbedingungen und Lernstrategien. Welche Rolle spielen kognitive Verstehtensstrukturen*. Tübingen: Narr (= *Tübinger Beiträge zur Linguistik* 366).
- Nold, Günter (2003). *DESI – a language assesment project in Germany and the pros and cons of large scale testing*. In Claudia Finkbeiner / Sylvia Fehling (Hg.). *Evaluation im Brennpunkt – Fremdsprachen lernen und lehren*. *Empirische Pädagogik* 17, 3, 368-379.
- Nold, Günter / Rossa, Henning (2006): *Leseverstehen*. In: Beck, Bärbel / Klieme, Eckhard, Hrsg., *Sprachliche Kompetenzen: Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International)*. Weinheim und Basel: Beltz, 197-211.
- Norris, John M. (2002): *Interpretations, intended uses and designs in task-based language assessment*. In: *Language Testing*, 337-346.
- Norris, John M. / Brown, James D. / Hudson Thom / Yoshioka, Jim (1998): *Designing second language performance assessments*. Honolulu: University of Hawaii Press.
- North, Brian (2000): *The development of a common framework scale of language proficiency*. New York: Lang (= *Theoretical studies in second language acquisition* 8).

- North, Brian / Schneider, Günther (1998): Scaling, descriptors for language proficiency scales. In: *Language Testing* 2, 217-262.
- Nunan, David (1989): *Designing Tasks for the Communicative Classroom*. Cambridge: University Press.
- Nunan, David (2004): *Task-based Language Teaching*. Cambridge: CUP.
- ÖDaF - Mitteilungen. Themenheft: Was soll ich können müssen? Lernziele und Evaluationsverfahren für den DaF / DaZ-Unterricht. Wien: ÖDaF, 2003,1.
- OECD (Hg.) (2003): The PISA 2003 assessment framework. Online: <http://www.oecd.org/dataoecd/46/14/33694881.pdf> (26.01.2007).
- OECD (2003): Reading Literacy. In: OECD, Hrsg., *The PISA 2003 Assessment Framework*, 107-129. Online: <http://www.pisa.oecd.org/dataoecd/38/52/33707212.pdf> (5.02.08)
- Oelkers, Jürgen / Reusser, Kurt (erscheint 2008): Qualität entwickeln, Standards sichern, mit Differenz umgehen. Expertise im Auftrag der vier Länder: Deutschland, Luxemburg, Österreich, Schweiz.
- O'Sullivan, Barry (2002): Learner acquaintanceship and oral proficiency test pair-task performance. In: *Language Testing* 19 / 3, 277-295.
- O'Sullivan, Barry / Weir, Cyril J. / Saville, Nick (2002): Using observation checklists to validate speaking-test tasks. In: *Language Testing* 19 / 1, 33-56.
- Phakiti, Aek (2007): *Strategic Competence and EFL Reading Test Performance: A Structural Equation Modeling Approach*. Frankfurt/M: Lang.
- Plaschy, Stephanie / Iwar Werlen (2000): *Zweisprachige Schule Brig-Glis. Schlussbericht zum Projekt Classe Bilingue 1995-2000 an den Schulpräsidenten zuhanden des Gemeinderates der Stadtgemeinde Brig-Glis*. Bern und Brig-Glis: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Plaschy, Stephanie / Iwar Werlen (2001a): *Schlussbericht über die Ergebnisse der Evaluation des ersten Jahrganges der zweisprachigen Matura. Zuhanden der Rektorate des Deutschen Gymnasiums Biel und des Gymnase Français Bienne*. Bern und Brig-Glis: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Plaschy, Stephanie / Iwar Werlen (2001b): *Zweisprachige Matura Biel-Bienne Kurzbericht über die bisherigen Ergebnisse. Zuhanden der Rektorate des Deutschen Gymnasiums Biel und des Gymnase Français Bienne*. Bern und Brig-Glis: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Plaschy, Stephanie / Werlen, Iwar (2003a). *Echange d'étudiants à la Haute Ecole Pédagogique du Valais. Rapport sur l'expérience et l'avis des étudiants, professeurs et praticiens-formateurs à la fin de la première année d'études*. Brigue-Glis, Berne, Centre universitaires des recherche sur le plurilinguisme.
- Plaschy, Stephanie / Werlen, Iwar (2003b). *Studierendenaustausch an der Pädagogischen Hochschule Wallis. Bericht über die Erfahrungen und Meinungen der beteiligten Studierenden, Dozierenden und Praktikumslehrpersonen am Ende des ersten Studienjahres*. Brig-Glis, Bern, Universitäres Forschungszentrum für Mehrsprachigkeit.
- Pürschel, Heiner / Raatz, Ulrich (Hg.) (2001): *Tests and Translation. Papers in memory of Christine Klein-Braley*. Bochum: AKS (= FLF Fremdsprachen in Lehre und Forschung 27).
- Purpura, James E. (2000): *Learner strategy use and performance on language tests: structural equation modeling Approach*. Cambridge: University Press (= *Studies in Language Testing* 8).
- Quetz, Jürgen (2003): A1 - A2 - B1 - B2 - C1 - C2. Der Gemeinsame europäische Referenzrahmen. In: *Deutsch als Fremdsprache. Zeitschrift zur Theorie und Praxis des Deutschunterrichts für Ausländer* 1, 42-48.
- Rickheit, Gert / Herrmann, Theo / Deutsch, Werner (Hg.) (2003): *Psycholinguistik - Psycholinguistics. Ein internationales Handbuch*. Berlin: de Gruyter (=Handbücher zur Sprach- und Kommunikationswissenschaft, Band 24).
- Rickheit, Gert / Strohner, Hans (1993): *Grundlagen der kognitiven Sprachverarbeitung*. Tübingen: Francke.
- Riemer, Claudia (1997): *Individuelle Unterschiede im Fremdspracherwerb. Eine Longitudinalstudie über die Wechselwirksamkeit ausgewählter Einflussfaktoren*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Riemer, Claudia (2000): *Der interpretative Mehr-Methoden-Ansatz zur Untersuchung individueller Unterschiede*. In: Aguado, Karin (Hrsg.): *Zur Methodologie in der empirischen Fremdsprachenforschung*, Baltmannsweiler: Schneider Verlag Hohengehren, 93-104.
- Robinson, Peter (Hg.) (2001): *Cognition and Second Language Instruction*. Cambridge: University Press (= *Cambridge Applied Linguistics*).
- Robinson, Peter (2001): *Task complexity, cognitive resources, and syllabus design: a triadic framework for examining task influences on SLA*. In: Robinson, Peter (Hg.): *Cognition and Second Language Instruction*, 287-318.
- Robinson, Peter (2005): *Cognitive complexity and task sequencing: Studies in a Componential Framework for second language task design*. In: *International Review of Applied Linguistics in Language Teaching (IRAL)*, 43, (1), 1-32.
- Rost, Michael (2002): *Teaching and Researching Listening, Applied Linguistics in Action*. Harlow: Pearson Education.
- Rupp, A. A., Ferne, T., & Choi, H. (2006). *How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective*. *Language Testing*, 23, 441-474.
- Schneider, Günther (1999): *Funktionen und wissenschaftliche Grundlagen der Kompetenzbeschreibungen im Sprachenportfolio*. In: *Babylonia* 1, 29-33.

- Schneider, Günther (2000): Qualität zeigen. Fremdsprachenkenntnisse, Fremdsprachenlernen, Sprachenportfolio. In: Trier, Uri Peter (ed.): Bildungswirksamkeit zwischen Forschung und Politik. Nationales Forschungsprogramm 33, Wirksamkeit unserer Bildungssysteme. Chur, Zürich: Rüegger, S. 189-199.
- Schneider, Günther (2000) (Hg.): Themenheft Lernerfahrungen reflektieren - Lernerfolge beurteilen. Neue Tendenzen und Perspektiven im Bereich der Leistungsbeurteilung. Babylonia 1, 2000.
- Schneider, Günther (2001): Kompetenzbeschreibungen für das 'Europäische Sprachenportfolio'. In: Henrici, Gert / Zöfgen, Ekkehard (Hg.): Fremdsprachen lehren und lernen (FLuL) 30, 193-214.
- Schneider, Günther (2003): Der Europäische Referenzrahmen und die Mehrsprachigkeit. In: Neuner, Gerhard / Koithan, Ute (Hg.): Tagungsdokumentation 2003. 'Internationales Qualitätsnetz Deutsch als Fremdsprache'. Tagungsbeiträge und Arbeitsberichte zum Thema: 'Mehrsprachigkeit im Bereich Deutsch als Fremdsprache' 27.-29. Juni 2003, Universität Kassel, 87-108.
- Schneider, Günther (2005): Der "Gemeinsame europäische Referenzrahmen für Sprachen" als Grundlage von Bildungsstandards für die Fremdsprachen – Methodologische Probleme der Entwicklung und Adaptierung von Kompetenzbeschreibungen. In: Schweizerische Zeitschrift für Bildungswissenschaften 1, 13-37.
- Schneider, Günther / Lenz, Peter (2001): Guide for Developers of a European Language Portfolio. Strasbourg: Council of Europe.
- Schneider, Günther / Lenz, Peter (2001): Portfolio européen des langues: Guide à l' usage des concepteurs. Strasbourg: Conseil de l' Europe.
- Schneider, Günther / North, Brian (1999): 'In anderen Sprachen kann ich...' Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit. Umsetzungsbericht. Bern: NFP (= NFP 33. Wirksamkeit unserer Bildungssysteme).
- Schneider, Günther / North, Brian (2000): 'Dans d' autres langues, je suis capable de...'. Échelles pour la description, l' évaluation et l' auto-évaluation des compétences en langues étrangères. Rapport de valorisation. Berne/Aarau: PNR 33/CSRE (= PNR 33. L'efficacité de nos systèmes de formation).
- Schneider, Günther / North, Brian (2000): Fremdsprachen können - was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit. Chur / Zürich: Rüegger (= NFP 33. Wirksamkeit unserer Bildungssysteme).
- Schneider, Günther / North, Brian / Koch, Leo / Schweizerische Konferenz der kantonalen Erziehungsdirektoren (Hg.) (2001): Portfolio européen des langues / Europäisches Sprachenportfolio / Portfolio europeo delle lingue / European Language Portfolio. Version pour jeunes et adultes. Version für Jugendliche und Erwachsene. Bern: BLMV.
- Schneider, Günther / Studer, Thomas (2007): Der "Referenzrahmen" und Bildungsstandards für die Fremdsprachen. In: Labudde, Peter (Hrsg.): Bildungsstandards am Gymnasium. Korsett oder Katalysator?, 219-228.
- Schwob, Irène (2002): Evaluation de l' enseignement / apprentissage bilingue en classe de 4e primaire de Monthey, Sierre et Sion. Année scolaire 2000 / 2001. Rapport intermédiaire. Neuchâtel: IRDP.
- Serra, Cecilia (2002). Evaluation des zweisprachigen Unterrichts in Chur. Schulprojekt Zweisprachige Primarschule der Stadtschule Chur. Schuljahr 2001/2002. Zuhanden der Schuldirektion der Stadtschule Chur. Bern und Brig-Glis, Universitäres Forschungszentrum für Mehrsprachigkeit. Übersetzung aus dem Französischen: Stephanie Plaschy.
- Shaw, Stuart D. / Weir, Cyril (2007): Examining Writing. Cambridge: CUP.
- Shohamy, Elana (2003): The Power of Tests: A Critical Perspective on the Uses of Language Tests. Reviewed by Arieh Sherris. In: Chapelle, Carol A.: TESOL Quarterly. Teachers of English to speakers of other Languages 37 / 1, 189-190.
- Sigott, Günther (1993): Zur Lernbarkeit von Englisch und Französisch für deutsche Muttersprachler. Eine exploratorische Pilotstudie. Tübingen: Narr (= AAA - Buchreihe zu den Arbeiten aus Anglistik und Amerikanistik 6).
- Sigott, Günther (1998): Sprachkompetenzmessung mit dem C-Test: Ein Resumé. In: ÖDaF - Mitteilungen. Themenheft: Testen, Prüfen, Leistungsbeurteilung und Evaluation. LernerInnen\_Autonomie 2, 9-20.
- Sigott, Guenther/ Gassner, Otmar/ Mewald, Claudia/ Siller, Klaus (2007): *E8-Standardstests. Entwicklung der Tests für die rezeptiven Fertigkeiten: Überblick*. Wien: Bundesministerium für Bildung, Wissenschaft und Kultur. (=LTC Technical Report 1)
- Skehan, P. & Foster, P. (2001). Cognition and tasks. In P. Robinson (Hrsg.). Cognition and Second Language Instruction. Cambridge: University Press, 183-205.
- Skehan, Peter (1998): A Cognitive Approach to Language Learning. Oxford: University Press.
- Skehan, Peter / Foster, Pauline (2001): Cognition and tasks. In: Robinson, Peter (Hg.): Cognition and Second Language Instruction, 183-205.
- Spolsky, Bernard (o. J.): Assessment and testing: notes for the third millennium. Online: <http://www.russnet.org/fipse/SpolskyWhitePaper.html> (30.05.08)
- Stede, Manfred (2007): *Korpusgestützte Textanalyse. Grundzüge der ebenenorientierten Textlinguistik*. Tübingen: Gunter Narr Verlag.
- Steinberg, Danny D. / Nagata, Hiroshia / Aline, David P. (2001): Psycholinguistics: Language, Mind and World. Harlow: Pearson Education.

- Stotz, Daniel (2000): Evaluation im frühen teilimmersiven Fremdsprachenunterricht. In: *Babylonia*. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen. Réfléchir sur les expériences d' apprentissage - en évaluer les résultats 1, 19-22.
- Studer, Thomas (2000): Internationale Sprachenzertifikate aus der Sicht von Schweizer TestautorInnen. Aspekte der Testentstehung am Beispiel des Zertifikats Deutsch, der Zentralen Mittelstufenprüfung und der DELF/DALF-Diplome. In: *Babylonia*. Themenheft: Lernerfahrungen reflektieren - Lernerfolge beurteilen. Réfléchir sur les expériences d' apprentissage - en évaluer les résultats 1, 62-66.
- Studer, Thomas (2004): Internationale Zertifikate für Deutsch als Fremdsprache in der Schweiz. In: *bulletin vals-asla* 79.
- Studer, Thomas / Mettler, Monika / Lenz, Peter (2004). Das Projekt IEF: Instrumente für die Evaluation von Fremdsprachenkompetenzen (Englisch/Französisch). *Schweizerische Zeitschrift für Bildungswissenschaften*, 419-434.
- Studer, Thomas u. a. (1999): Zertifikat Deutsch. Lernziele und Testformat. Frankfurt a.M. WBT/ ÖSD/ Goethe-Institut/ EDK.
- Tschirmer, Erwin (2001): Die Evaluation fremdsprachlicher mündlicher Handlungskompetenz: Ein Problemaufriss. In: Henrici, Gert / Zöfgen, Ekkehard (Hg.): *Fremdsprachen lehren und lernen (FLuL)* 30, 87-115.
- Tunger, Verena / Werlen, Iwar (2003a). Échange intersite à la Haute École Pédagogique du Valais. Rapport sur l'expérience et l'avis des étudiants, professeurs et praticiens-formateurs à la fin de la première et de la deuxième année d'études. Brigue-Glis, Berne, Centre universitaires des recherche sur le plurilinguisme.
- Tunger, Verena / Werlen, Iwar (2003b). Studierendenaustausch an der Pädagogischen Hochschule Wallis. Bericht über die Erfahrungen und Meinungen der beteiligten Studierenden, Dozierenden und Praktikumslehrpersonen am Ende des ersten und des zweiten Studienjahres. Brig-Glis, Bern, Universitäres Forschungszentrum für Mehrsprachigkeit.
- Urquhart, Sandy A.H. / Weir, Cyril J. (1998): *Reading in a second language: process, product and practice*. London / New York: Longman (= Applied Linguistics and Language Study).
- Veltcheff, Caroline / Hilton, Stanley (2003): *Evaluation en FLE*. Paris: Hachette.
- Vollmer, Hellmut. J. (2003). Ein gemeinsamer europäischer Referenzrahmen für Sprachen: Nicht mehr, nicht weniger. In Bausch / Christ / Königs / Krumm (Hg.). *Der Gemeinsame europäische Referenzrahmen in der Diskussion*. Tübingen: Narr, 192-206.
- Weidinger, Walter et. al. (2003): Erprobung und Evaluierung von Bildungsstandards an Wiener Pflichtschulen der Sekundarstufe I. Studie Bildungsstandards. Projektinformaton Mai 2003. Wien: Stadtschulrat.
- Weinert, F. E. (2001): Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In: F. E. Weinert (Hrsg.): *Leistungsmessungen in Schulen*. Weinheim und Basel: Beltz Verlag, S. 17–31.
- Weir, Cyril J. (1993): *Understanding and Developing Language Tests*. London: Prentice Hall.
- Weir, Cyril J. (2005a): *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave/Macmillan.
- Weir, Cyril J. (2005b): Limitations of the Common European framework for developing comparable examinations and tests. In: *Language Testing*, 3, 281-300.
- Weiss, Jacques (1994): Evaluation interne de l' expérience d' apprentissage bilingue de Sierre. Rapport à l' intention du groupe de coordination. Neuchâtel: IRDP (= IRDP Recherches).
- Weiss, Jacques (1998): *Une évaluation informative pour la régulation des apprentissages et des formations*. Neuenburg: IRDP (= IRDP Recherches).
- Werlen, Iwar (2003). Flüssiges Sprechen vortäuschen: Wiederholungen als Technik zur Erzeugung flüssigen Sprechens bei Sprecherinnen von Deutsch als L 2. In: Mondada, Lorenza / Pekarek Doehler, Simona (éds.), *Plurilinguisme Mehrsprachigkeit Plurilingualism. Enjeux identitaires, socio-culturels et éducatifs*. Festschrift pour Georges Lüdi. Tübingen, Basel, Francke, 179-195.
- Werlen, Iwar / Fuchs, Gabriela (1999): Zweisprachige Klassen der Bieler Gymnasien, Projektverlauf 1998/99. Befragung der Schülerinnen und Schüler, der Eltern und der Lehrerschaft. Brig-Glis/ Bern: Universitäres Forschungszentrum für Mehrsprachigkeit.
- Widdowson, H.G. (1989): Knowledge of language and ability for use. In: *Applied Linguistics*, 10 (2), 128-137.
- Wilson, Mark (2004): *Constructing measures. An item response modeling approach*. Mahwah, N.J.: Erlbaum.
- Wolfe-Quintero, Kate / Inagaki, Shunji / Kim, Hae-Young (1998): *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. Honolulu: University Press.
- Wolff, Dieter (2002): *Fremdsprachenlernen als Konstruktion. Grundlagen für eine konstruktivistische Fremdsprachendidaktik*. Frankfurt/Main etc.: Lang.
- Yuan, Fangyuan / Ellis, Rod (2003): The Effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. In: *Applied Linguistics* 24 / 1, 1-27.
- Zarate, Geneviève / Gohard-Radenkovic, Aline / Lussier, Denise / Penz, Hermine (2003): *Médiation culturelle et didactique des langues*. Strasbourg: European Centre for Modern Languages / Conseil de l'Europe.
- Ziberi-Luginbühl, Johanna (1999): *Zweitsprachunterricht im obligatorischen Schulsystem*. Bern: NFP (= NFP 33. Umsetzungsbericht).

- Zimmermann, Hansmartin / Iwar Werlen (1996). Das Projekt "Zweitsprachunterricht im obligatorischen Schulsystem": Konzept, empirische Daten und erste Ergebnisse. In : Anne-Claude Berthoud (éd), Acquisition des compétences discursives dans un contexte plurilingue. Bulletin Suisse de linguistique appliquée (VALS/ASLA) 64, octobre, 35-59.
- Zydati, Wolfgang (2002): Leistungsentwicklung und Sprachstandserhebungen im Englischunterricht. Methoden und Ergebnisse der Evaluierung eines Schulversuchs zur Begabtenfderung: Gymnasiale Regel- und Expressklassen im Vergleich. Frankfurt/M. u. a.: Lang (= KFU - Kolloquium Fremdsprachenunterricht 8).
- Zydati, Wolfgang (2005): Bildungsstandards und Kompetenzniveau im Englischunterricht. Konzepte, Empirie, Kritik und Konsequenzen. Frankfurt u. a.: Lang (= Kolloquium Fremdsprachenunterricht 22).

**Statistik-Software:**

- ConQuest Generalised Item Response Modelling Software, Build Jan. 10, 2006. Wu, Margaret / Adams, Ray / Wilson, Mark / Haldane, Sam. Australian Council for Educational Research; University of California, Berkeley.
- Facets, v. 3.63 (2007). Linacre, John Mike. Chicago.
- JMP - The Statistical Discovery Software. SAS Institute Inc. Cary, N.C.
- R. The R Project for Statistical Computing. Online: <http://www.r-project.org>.
- SPSS, v. 16 (2007). SPSS, Inc. Chicago.
- Winsteps, v. 3.64 (2007). Linacre, John Mike. Chicago.