



Entwicklung und Umsetzung eines Evaluationskonzeptes für die KMK-Fremdsprachenzertifikatsprüfungen zur Sicherung der Vergleichbarkeit der Standards

BLK-Verbundmodellversuch

Förderkennzeichen: K 0971.00

Laufzeit: 1. November 2003 – 31. Oktober 2006

Handreichung zur

- **Aufgabenstellung**
- **Bewertung und**
- **Prüfungsdurchführung**



Staatsinstitut für
Schulqualität und
Bildungsforschung,
München



Behörde für Bildung und
Sport, Referat
Berufliche Schulen,
Hamburg



Thüringer
Kultusministerium,
Erfurt

Wissenschaftliche Begleitung:

Fachhochschule Rosenheim
University of Applied Sciences



Gefördert durch bmb+f
Bundesministerium
für Bildung
und Forschung

Herausgeber:

ISB, Staatsinstitut für Schulqualität und Bildungsforschung, Bayern

Behörde für Bildung und Sport, Hamburg

Thüringer Kultusministerium, Erfurt

Fachhochschule Rosenheim, Fachbereich Allgemeinwissenschaften,
Professor Micheál Ó Dúill, D. Phil

München 2005 (Teile I und II) / 2006 (Teil III)

Sehr geehrte Damen und Herren, liebe Kolleginnen und Kollegen,

seit 1998 wird in den Bundesländern die Prüfung und Zertifizierung berufsbezogener Fremdsprachenkenntnisse auf der Grundlage europäischer Standards angeboten. Die Grundsätze dieser Prüfung sind in der *KMK-Rahmenvereinbarung über die Zertifizierung von Fremdsprachenkenntnissen in der beruflichen Bildung* festgelegt. Im BLK-Modellversuch „EU-KonZert“ erfolgte eine Evaluation der länderspezifisch unterschiedlich ausgeprägten Prüfungen.

Insbesondere wurden dabei die Aspekte **Aufgabenstellung** (Eignung von Aufgabentypen), **Bewertung** (z. B. kriterienorientierte Bewertung) und **Prüfungsdurchführung** (hauptsächlich mündliche Prüfungen) auch im Hinblick auf die Erfüllung der Rahmenvereinbarung vergleichend untersucht.

Für die Durchführung des BLK-Modellversuches wurden folgende Inhalte und Zeitrahmen festgelegt:

Phase I	Gemeinsame Standardfindung bei der Aufgabenstellung	Nov. 2003 – Okt. 2004
Phase II	Gemeinsame Standardfindung bei der Bewertung	Nov. 2004 – Okt. 2005
Phase III	Gemeinsame Standardfindung bei der Prüfungsdurchführung	Nov. 2005 – Okt. 2006

In der praktischen Modellversuchsarbeit wurde dabei insbesondere Wert auf die Analyse konkreter Prüfungsbeispiele aus verschiedenen Bundesländern gelegt. Dabei wurden auch nicht direkt am Modellversuch beteiligte Länder einbezogen, um in möglichst vielen Fällen das vorhandene Expertenwissen aufzugreifen, zu systematisieren und nachhaltig verfügbar zu machen.

Ziel der Untersuchung war es, aus den erhaltenen Daten *Best-practice*-Beispiele zu extrahieren sowie für die Prüfungsanbieter Konzepte in Form von Kriterienkatalogen und dokumentierten Prozessabläufen für eine weiterentwickelte Zertifizierungsarbeit zu erstellen.

Natürlich ist dieser Prozess „a work in progress“, der auch nach Beendigung des BLK-Modellversuches nicht abgeschlossen sein kann. Eine länderübergreifende Weiterarbeit erscheint allen Beteiligten sinnvoll und notwendig.

Der BLK-Modellversuch verstand sich als ein Beitrag zur Qualitätssicherung des KMK-Fremdsprachenzertifikats. Er wurde als Verbundmodellversuch der Länder Bayern, Hamburg und Thüringen durchgeführt. Diese drei Länder repräsentierten den bundesweiten Querschnitt hinsichtlich verschiedener Formen von Verwaltung und Schulorganisation (Flächenstaat, Stadtstaat, neues Bundesland) und deckten gleichzeitig eine erhebliche Zahl der durchgeführten Zertifikatsprüfungen im Bundesgebiet ab.

Wir möchten uns noch einmal auf diesem Wege bei unseren Teams und den Teilnehmern aus allen Bundesländern bedanken, die aktiv und konstruktiv an den Workshops mitgewirkt haben. Dank gilt auch Frau Maike Engelhardt, der Fachhochschule Rosenheim, der Bund-Länder-Kommission, der KMK als auch dem Bundesministerium für Bildung und Forschung, ohne die dieses unseres Erachtens sehr präsentabile Ergebnis nicht möglich gewesen wäre.

Wir hoffen, mit dieser Handreichung allen Aufgabenerstellern und Prüfern* einen Leitfaden an die Hand zu geben, der als praktische Orientierungshilfe zur erfolgreichen Durchführung von KMK-Fremdsprachenzertifikatsprüfungen dienen kann und gleichzeitig einen Beitrag zur Standardisierung darstellt.

Mit freundlichen Grüßen

die EU-KonZert-Lenkungsgruppenmitglieder

Andrea Nolte (Thüringen), Josef Biro und Jan Mayer (Bayern), Micheál Ó Dúill (Fachhochschule Rosenheim) und Manfred Thönicke (Hamburg)

* Anm.: Die im Text dieser Handreichung verwendeten Personenbezeichnungen gelten für beide Geschlechter.

Inhaltsverzeichnis

Teil I: Aufgabenstellung

	Seite
1	
	Micheál Ó Dúill,
	Wissenschaftliche Begleitung, Fachhochschule Rosenheim:
	„ Stellenweise klar. “
	Randbemerkungen zu einer zentralen Aufgabe
1.1	Einführung
1.2	Deskriptoren des Referenzrahmens und des KMK- Fremdsprachenzertifikats
1.3	Fremdsprachliche Kompetenzmodelle
1.4	Berufsbezogener Fremdsprachenunterricht und Leistungsmessung
1.5	Bisherige Handreichungen vor und nach Novellierung der Rahmen- vereinbarung
1.6	Testkonstrukt
1.7	Der BLK-Modellversuch EU-KonZert
1.8	Ausblick
2	
	Empfehlungen zur Aufgabenstellung
	Kompilation und Kommentierung von Aufgaben bzw. Aufgabenstel- lungen von Maïke Engelhardt , Fachhochschule Rosenheim
2.1	Anforderungsbereiche
2.2	Rezeption
2.2.1	Dokumentation der Workshops
	Rezeption 1: Hörverstehen
2.2.2	Best-practice-Beispiele Rezeption 1: Hörverstehen
	Stufe I (Waystage, A2)
	Beispiel 1: Gastgewerbe
	Beispiel 2: Gastgewerbe
	Beispiel 3: Bankkaufleute
	Stufe II (Threshold, B1)
	Beispiel 4: Einzelhandelskaufleute
	Beispiel 5: Automobilkaufleute
	Beispiel 6: Gastgewerbe
	Stufe III (Vantage, B2)
	Beispiel 7: Gastgewerbe
	Beispiel 8: Fachangestellte für Medien- und Informationsdienste
	Beispiel 9: Gastgewerbe
	Rezeption 2: Leseverstehen
2.2.3	Best-practice-Beispiele Rezeption 2: Leseverstehen
	Stufe I (Waystage, A2)
	Beispiel 10: Gastgewerbe
	Beispiel 11: Fachangestellte für Medien- und Informationsdienste
	Beispiel 12: Gastgewerbe
	Stufe II (Threshold, B1)
	Beispiel 13: IT-Fachleute
	Beispiel 14: Fachangestellte für Medien- und Informationsdienste
	Beispiel 15: Gastgewerbe

	Seite
	51
	52
	54
	56
2.2.4	57
2.3	58
2.3.1	58
2.3.2	59
	59
	60
	62
	63
	64
	64
	65
	65
	66
	66
	66
	69
2.3.3	70
2.4	71
2.4.1	71
2.4.2	72
	72
	72
	73
	73
	74
	75
	77
	79
	80
	81
	82
	82
2.4.3	83

Teil II: Bewertung

		Seite
1	Micheál Ó Dúill, Wissenschaftliche Begleitung, Fachhochschule Rosenheim „Mark my Words.“	
	Bewertung in der Diskussion	84
1.1	Einführung	84
1.2	Definitionen	85
1.3	Bewertung	87
1.4	Korrekturaufgabe vollständig gelöst?	90
1.5	Einige Empfehlungen der Fachwelt	94
1.6	Envoi	97
2	Bewertung von Prüfungsaufgaben Gemeinsame Standardfindung in der Phase II des BLK- Modellversuches EU-KonZert Dokumentation und Zusammenfassung der Ergebnisse von Maike Engelhardt	99
2.1	Arbeit in den Workshops	99
2.1.1	Arbeitsansatz	100
2.1.2	Durchführung	101
2.1.3	Ergebnissicherung	102
2.1.4	1. Workshop in München (Dezember 2004)	103
2.1.5	2. Workshop in Hamburg (Januar 2005)	104
2.1.6	3. Workshop in Bad Berka	105
2.1.7	Abschlussworkshop in München (Juni 2005)	107
2.2	Zusammenfassung der Ergebnisse der Phase II des BLK- Modellversuches EU-KonZert	110
2.3	Anhang	111
	a. Prüfungsaufgaben mit Schülerlösungen	112
	b. Bewertungsbogen der Workshopteilnehmer, Workshops 1-3	116
	c. Bewertungsbögen der Workshopteilnehmer, Workshop 4	117
	d. Deskriptoren für sprachproduktive Leistungen, zweispaltig, Stufe 1-3	119
	e. Deskriptoren für schriftliche mediative Leistungen, einspaltig, Stufe 1-3	122

Teil III: Prüfungsdurchführung

		Seite
	Micheál Ó Dúill,	
	Wissenschaftliche Begleitung, Fachhochschule Rosenheim	
1	"Now you're talking!"	
	Überlegungen zu mündlichen Sprachprüfungen	125
1.1	"Oh, now come I? Yes, ..."	125
1.2	Einige weitere Definitionen	128
1.3	Aufgabenstellung	131
1.4	Prüfungsdurchführung	134
1.5	Bewertung	137
1.6	"Oh, and another thing I wanted to say was..."	140
2	Götz Reuter, ISB München	
	Eine prozessorientierte Betrachtung der Workshops des BLK-Modellversuchs EU-KonZert zum mündlichen Teil der KMK–Fremdsprachenzertifikatsprüfung	147
2.1	Die ersten drei Workshops	147
2.1.1	Ausgangslage: Die Durchführungspraxis der mündlichen Prüfung in den Ländern	148
2.1.2	Analyse der Aufgabenstellung	150
2.1.2.1	Aspekte der Aufgabenstellung in Bad Berka	150
2.1.2.2	Aspekte der Aufgabenstellung in München	153
2.1.2.3	Aspekte der Aufgabenstellung in Hamburg	154
2.1.2.4	Aspekte der Aufgabenstellung im abschließenden Workshop in Bad Berka	155
2.1.2.5	Überblick: Desiderata zur Standardsicherung mündlicher Prüfungspraxis	156
2.1.3	Erprobung und Weiterentwicklung der Bewertungsinstrumente	159
2.1.3.1	Aspekte der Bewertungsinstrumente in Bad Berka	159
2.1.3.2	Aspekte der Bewertungsinstrumente in München	160
2.1.3.3	Aspekte der Bewertungsinstrumente in Hamburg	162
2.1.3.4	Aspekte der Bewertungsinstrumente im abschließenden Workshop in Bad Berka	163
2.1.4	Entwicklung von Best-Practice Beispielen	164
2.2	Anhang	166
	Anlage 1: Bewertungsinstrumente (Auswertungsbogen)	166
	Anlagen 2, 3 und 4: Einzelauswertung Bewertungsinstrumente Interaktion	167
	Anlage 5: Best-Practice Beispiele	169
	Anlage 6: Beispiele für Operatoren	189
	Anlage 7: Die Bewertungsinstrumente	190
	Hinweise zu den Deskriptoren	190
	Deskriptoren für Stufe I (A2)	191
	Deskriptoren für Stufe II (B1)	192
	Deskriptoren für Stufe III (B2)	193

Teil IV: Rück- und Ausblick

		Seite
	Micheál Ó Dúill, Wissenschaftliche Begleitung, Fachhochschule Rosenheim	
	Rück- und Ausblick	194
1	Ausgangsbasis für den BLK-Modellversuchsantrag	194
2	Anderweitige Forschungsergebnisse im Bereich der Testtheorie seit Antragstellung	196
2.1	Allgemeine Abhandlungen zur Leistungsmessung und zum Testen im Bereich des Fremdsprachenerwerbs	197
2.2	Diskussionen um den Gemeinsamen europäischen Referenz- rahmen und Bildungsstandards der Kultusministerkonferenz	197
2.3	Einzelveröffentlichungen zu Teilaspekten des Testens	197
2.4	Reihen zum Sprachtesten, Fachzeitschriften und internationale Organisationen	198
2.4.1	Bücherreihen	198
2.4.2	Fachzeitschriften	199
2.4.3	Internationale Organisationen	199
3	Gesellschaftspolitische Entwicklung der Wahrnehmung und Wertung bestimmter testtheoretischer Ansätze und Reaktionen der Kultusministerkonferenz auf Ergebnisse interna- tionaler Erhebungen	199
4	Ziele des BLK-Modellversuches, wie diese im Antrag formuliert wurden	202
5	Weiterentwicklung der Ziele und Ergebnisse, die während der Laufzeit des BLK-Modellversuchs erarbeitet wurden	203
6	Zukünftige Entwicklungsmöglichkeiten für die KMK- Fremdsprachenzertifikatsprüfungen	206

Anhang

1	Die Rahmenvereinbarung zum KMK-Fremdsprachenzertifikat	210
2	Das Zertifikat	214
	Das Deckblatt (Seite 1)	214
	Angaben zur Person, erreichte Punktzahl (Seite 2)	215
	Die Stufenbeschreibungen in deutscher Sprache (Seite 3)	216
	Die Stufenbeschreibungen in englischer Sprache (Seite 4)	217
3	Angaben zum BLK-Modellversuch, Ansprechpartner, Kon- taktmöglichkeiten	218
4	Die am BLK-Modellversuch beteiligten Personen und Länder	219
5	Literaturverzeichnis	220

I.1 „Stellenweise klar.“

Randbemerkungen zu einer zentralen Aufgabe

Micheál Ó Dúill,

Wissenschaftliche Begleitung, Fachhochschule Rosenheim

Die Kunst schließt sich in ihren einzelnen Werken ab;
die Wissenschaft erscheint uns grenzenlos.

Goethe, *Materialien zur Geschichte der Farbenlehre*
(„Hamburger Ausgabe“ XIV: 40)

1.1 Einführung

Die drei Stufen des Fremdsprachenzertifikates, das an beruflichen Schulen in Deutschland erworben werden kann, orientieren sich, seit 1998 eine entsprechende Rahmenvereinbarung von der Kultusministerkonferenz beschlossen wurde, an Stufen des Gemeinsamen europäischen Referenzrahmens für Sprachen (Trim et al. 2001: *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen* (GeR)):

1. Grundsatz

Berufliche Schulen können auf freiwilliger Basis – unabhängig von einer Benotung im Zeugnis – eine Prüfung anbieten, in der sich Schülerinnen und Schüler ihre Fremdsprachenkenntnisse zertifizieren lassen können.

2. Prüfungsniveau und Berufsbezug

Die Prüfung wird jeweils in einer der drei Stufen I, II oder III durchgeführt. Sie orientieren sich an den Stufen „A2 (Waystage)“ (Stufe I), „B1 Threshold“ (Stufe II) und „B2 (Vantage)“ (Stufe III), die vom Europarat im „Gemeinsamen europäischen Referenzrahmen für Sprachen: Lernen, Lehren, Beurteilen“ aufgeführt werden.

(Zitiert nach der gegenwärtig gültigen Fassung der Rahmenvereinbarung, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland 2002b: 32; Abdruck der ersten, aus dem Jahr 1998 stammenden Fassung in Brandenburg 1999: 6)

Hiermit wurde die bildungspolitische Bedeutung des Gemeinsamen europäischen Referenzrahmens vom deutschen berufsbildenden Schulwesen zu einer Zeit erkannt, in der er zwar bereits von den Mitgliedern des Europarates verabschiedet worden (1997), das Werk und seine deutsche Übersetzung jedoch noch nicht im Verlagswesen erschienen waren (Council of Europe 2001; Trim et al. 2001).

Im ersten Teil dieser Handreichung wird erörtert, wie der BLK-Modellversuch EU-KonZert, der sich der Evaluierung der KMK-Fremdsprachenzertifikatsprüfungen widmet, aus dem gegenwärtigen Diskurs zum Wissenschaftsbereich Sprachtesten entstand und wie die Durchführung seiner ersten Phase innerhalb dieses wissenschaftlichen Diskurses einzuordnen ist. Der zweite Teil dieser Handreichung enthält Empfehlungen zur Aufgabenstellung, die auf der Grundlage der Ergebnisse dieser ersten Arbeitsphase des Modellversuchs entstanden.

Im Jahr 2002 ist die Rahmenvereinbarung der Kultusministerkonferenz in überarbeiteter Fassung erschienen, und der Gemeinsame europäische Referenzrahmen liefert seit seinem Erscheinen in Buchform „durch die erstaunlich große Anerkennung, die er gefunden hat, die Grundlage für eine auch international transparente Zertifizierung von Fremdsprachenkenntnissen“ (Abel 2003: 18-19). Eine schriftliche Betriebsbefragung von 20 niederbayerischen Betrieben ergab 2000, dass 24% der Beauftragten einen Facharbeiter mit KMK-Fremdsprachenzertifikat

„bevorzugt“ einstellen würden (Wagner 2000: 53); noch interessanter ist das Ergebnis der persönlichen Gespräche, die ergaben, dass „dem Kandidaten mit Zertifikat der Vorzug gegeben werden“ würde,

falls die Situation eintrete, dass sich zwei gleichermaßen qualifizierte Facharbeiter für einen Arbeitsplatz bewerben, der eine das Zertifikat hat, der andere aber nicht (ibid.: 54).

1.2 Deskriptoren des Referenzrahmens und des KMK-Fremdsprachenzertifikats

Basiert der Referenzrahmen auf der Beschreibung von Kompetenzstufen und wurden in Anlehnung hieran bei der Rahmenvereinbarung der Kultusministerkonferenz ähnliche Deskriptoren für die Verwendung von Fremdsprachen im beruflichen Bereich entworfen, so gilt es bei einer Handreichung zur Erstellung von Prüfungsaufgaben, die Rolle der Unterscheidungen zwischen den diversen Kompetenzstufen in Erinnerung zu rufen.

Zunächst soll betont werden, dass die Erarbeitung der Deskriptoren der Referenzniveaus als Kern des Referenzrahmens angesehen wird:

Zweifelsohne machen das Herzstück des Referenzrahmens die empirisch abgesicherten Skalen bzw. Niveaustufenbeschreibungen des Referenzsystems aus [...] Dies schafft Transparenz, Kohärenz und Vergleichbarkeit. (Bausch 2003: 29)

Es ist [...] das Verdienst des Referenzrahmens, [...] Sprachkenntnisse relativ genau und konkret nach sprachlicher Handlungsfähigkeit einschätzen zu können. (Freudenstein 2003: 80)

An den Niveaustufen kommt kein neu zu erstellender Test vorbei (Kleppin 2003: 109).

Die Skalierung fremdsprachlicher Handlungen und ihre Abstufung in sechs Kompetenzniveaus [...] eröffnet der Fremdsprachendidaktik [...] die Möglichkeit, bei der Beurteilung von Lernfortschritt und Leistung von der Fixierung auf den „Fehler“ wegzukommen [...] und auf das einzugehen, **was Lernende schon erreicht haben** und mit der Fremdsprache konkret tun können (Neuner 2003: 142).

Ein [...] Werk [...], dessen Bedeutung [...] für anzustrebende Zielsetzungen sowie für Formen und Inhalte der darauf bezogenen Leistungsmessung (achievement testing) und Fähigkeitsüberprüfung (proficiency testing) nicht hoch genug eingeschätzt werden kann (Vollmer 2003a: 198).

Es gibt zwar kritische Stimmen zum Stellenwert der im Referenzrahmen entwickelten Deskriptoren:

Dieser Versuch einer Eichung, eines auf ein genaues Maß Bringens, kann einem komplexen Phänomen wie dem der Sprachverwendung und des Sprachenlernens nicht gerecht werden (Schocker-v. Dittfurth (2003: 165).

Es überwiegen jedoch die Stimmen, die den Wert des Unternehmens vor solchen Angriffen schützen wollen:

Like any attempt to capture language performance in terms of language, the specifications, or 'descriptors' as they are called, cannot be absolutely precise (Morrow 2004: 8).

Diese Entwicklung von Deskriptoren zum Erfassen der nach Stufen gegliederten Sprachkompetenzen ist als Beitrag zu den Bemühungen um eine Standardsicherung in der Vermittlung von Fremdsprachen anzusehen. Der Beitrag des Referenzrahmens bei dieser Diskussion um Standardsicherung wird allgemein anerkannt:

The meaning of 'standards' as 'levels of proficiency' or 'levels certified by public examinations' [...] has received new impetus [...] with the publication of the Council of Europe's Common European Framework [...]. The ability of a transparent, independent framework like the Common European Framework is crucial to the attempt to establish a common scale of reference and comparison (Alderson / Banerjee 2001: 219-220)

1.3 Fremdsprachliche Kompetenzmodelle

Analog zur vom Europarat unterstützten Diskussion erhoben sich Stimmen in Deutschland, die das Fehlen gemeinsamer Standards bei schulischen Sprachtests bemängelt haben (Vergl. Alderson / Banerjee 2001: 219 zu Kieweg 1999; Zydatiß 2002: 89). Doch nicht nur im Bereich des Fremdsprachenunterrichts gewinnt die Frage der Standards an Brisanz:

Aufgaben der näheren Zukunft sind vor allem die **fachdidaktische Vertiefung in Kompetenzmodellen**, die Festlegung von **Mindeststandards**, die Entwicklung von **Aufgabenpools und Testverfahren** sowie die **Implementation** an den Schulen (Klieme et al. 2003: 10).

In einem Anhang der soeben zitierten, von dem deutschen Bundesministerium für Bildung und Forschung in Auftrag gegebenen Expertise zur Entwicklung nationaler Bildungsstandards wird der Gemeinsame europäische Referenzrahmen als Beispiel für eine weitere Entwicklung von Kompetenzmodellen präsentiert:

Der Referenzrahmen [...] systematisiert Dimensionen und Stufen sprachlicher Handlungsfähigkeit, konkretisiert sie durch verhaltensnahe Deskriptoren und ist so eine gute Ausgangsbasis für die Entwicklung von Aufgaben, Tests und anderen Prüfungsverfahren, mit denen die Sprachkompetenzen einzelner Personen eingeschätzt werden können (ibid.: 147).

In diesen Zusammenhang passen die Bemühungen der Kultusministerkonferenz, Fremdsprachenzertifikatsprüfungen an beruflichen Schulen Deutschlands einzuführen. Diese Bemühungen leiten sich von der Anregung ab, den Referenzrahmen als Versuch anzusehen,

„[...] Lernende und Lehrende, Autoren von Sprachkursen, Prüfungsanbieter und die Bildungsverwaltung dabei zu unterstützen, ihre Bemühungen in diesen Rahmen einzubetten und zu koordinieren“ (Trim et al. 2001: 18).

(Nicht uninteressant in diesem Zusammenhang dürfte sein, dass sich die deutsche Übersetzung des Referenzrahmens dafür entscheidet, hier den Stellenwert des Referenzrahmens nochmals explizit hervorzuheben („in diesen Rahmen einzubetten“), heißt es doch in der englischen Originalfassung an dieser Stelle lediglich: „assist learners, teachers, course designers, examining bodies and educational administrators to situate and co-ordinate their efforts“ [Council of Europe 2001: 6].)

Entscheidend für eine ausgewogene Beurteilung der Möglichkeiten, die sich dem Fremdsprachenunterricht durch Berücksichtigung des Referenzrahmens anbieten, ist die Erkenntnis, dass der Referenzrahmen ausdrücklich nie einengend präskriptiv konzipiert wurde, sondern dass seine Vertreter vielmehr stets eine wissenschaftlich fundierte Pluralität der Ansätze anregen:

Die Erstellung eines umfassenden, transparenten und kohärenten Referenzrahmens für das Lernen und Lehren von Sprachen bedeutet nicht, dass allen Beteiligten ein einziges, einheitliches System aufgezwungen werden soll. Ganz im Gegenteil sollte der *Referenzrahmen* offen und flexibel sein, damit er in allen spezifischen Situationen eingesetzt werden kann – wenn nötig natürlich mit entsprechenden Anpassungen (Trim et al. 2001: 20)

(so that it can be applied, with such adaptations as prove necessary, to particular situations [Council of Europe 2001: 7]).

The Manual [...] does not prescribe any single approach to constructing language tests or examinations. [...] The CEF [...] accepts that different examinations [*sic!*] reflect various goals (“constructs”) (North et al. 2003: 1)

1.4 Berufsbezogener Fremdsprachenunterricht und Leistungsmessung

In den 1990er Jahre wurde an verschiedenen Stellen in Deutschland ange-regt, den Berufsbezug im Fremdsprachenunterricht stärker zu betonen. So argu-mentieren zum einen Forscher des Instituts der deutschen Wirtschaft in Köln auf Grund einer Firmenbefragung:

Eine berufsbezogene Fremdsprachendidaktik, die zur Zeit noch in weiten Teilen fehlt, sollte auf das Ziel der beruflichen Handlungsfähigkeit abgerichtet sein (Schöpfer-Grabe / Weiß 1998: 126).

Zur gleichen Zeit plädieren zwei Modellversuchskoordinatorinnen zum Fremdsprachenpflichtunterricht in der Berufsschule dafür, dass in der Schulrealität eine „Praxisorientierung des Englischunterrichts [...] zu gewährleisten“ sei (Wei-dinger / Träger 1998: 40).

Sobald der Ruf nach einem berufsbezogenen und praxisorientierten Fremd-sprachenunterricht erklingt, ist man der Frage eine Antwort schuldig, wie eine Lei-stungsmessung dieser Anforderungen vorzunehmen sei. Damit befindet man sich gleich mitten in der kontrovers geführten Diskussion, ob und wie in der fremd-sprachlichen Leistungsmessung Sprachkenntnisse und Fachwissen zueinander in Beziehung zu setzen seien. Hierzu bezieht Douglas unter Verwendung des eng-lischsprachigen Begriffes ‘language for specific purposes’ (LSP) klar Stellung: „In LSP testing, [...] background knowledge is a necessary, integral part of the con-cept of specific purpose language ability” (Douglas 2000: 2). Im gleichen Jahr ver-tritt Alderson die Meinung, das Vorhandensein von Fachwissen dürfe höchstens für die Leistung in einer Sprachprüfung förderlich sein, solle aber nicht als Vorausset-zung gelten („every attempt should be made to allow background knowledge to facilitate performance rather than its absence to inhibit“ [Alderson 2000: 121]. Im Jahr darauf scheint Alderson mit seiner Kollegin Banerjee in einem Aufsatz, der den entsprechenden internationalen Forschungsstand zusammenfasst („state-of-the-art review“) mehr auf die Douglas’sche Meinung einzuschwenken:

The interaction between language knowledge and specific content knowledge [...] is perhaps the most crucial difference between general purpose testing and LSP testing, for in the former, any sort of background knowledge is considered to be a confounding variable that contributes construct-irrelevant variance to the test score. However, in the case of LSP testing, background knowledge constitutes an integral part of what is being tested, since it is hypothesised that test takers’ language knowledge has developed within the context of their academic or professional field (Alderson / Banerjee 2001: 222).

Hat man diesen grundsätzlichen Unterschied zwischen Sprachtests, die all-gemeine Sprachkenntnisse und denen, die fachbezogene Sprachkenntnisse prü-fen wollen, akzeptiert, gilt es der Frage nachzugehen, ob diese Unterscheidung eine Auswirkung auf die jeweils geeigneten Prüfungsformen hat. In ihrer Übersicht zu Sprachtesttypen stellen Grotjahn und Klein-Braley fest, dass

kommunikative Performanztests verglichen z. B. mit Wissenstests aufgrund ihrer größeren Wirklichkeitsnähe zumindest in bezug auf bestimmte Zielsetzungen eine Reihe von Vorteilen [haben]. Dies ist insbesondere der Fall, wenn Sprache für ge-

nau spezifizierte berufsbezogene Aufgaben [...] eingesetzt werden soll (Grotjahn / Klein-Braley 2001: 296).

Der beabsichtigte Berufsbezug der KMK-Fremdsprachenzertifikatsprüfungen wird in der entsprechenden Rahmenvereinbarung festgeschrieben:

Je Stufe soll die Prüfung differenziert nach den Erfordernissen der verschiedenen Berufsbereiche, wie zum Beispiel

- kaufmännisch-verwaltende Berufe
- gewerblich-technische Berufe
- gastgewerbliche Berufe
- sozialpflegerische, sozialpädagogische Berufe

durchgeführt werden. Innerhalb der Berufsbereiche können weitere berufsbezogene Konkretisierungen vorgenommen werden (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland 2002b: 32).

1.5 Bisherige Handreichungen vor und nach Novellierung der Rahmenvereinbarung

Zur Rahmenvereinbarung entstanden in verschiedenen Bundesländern Handreichungen zu den Zertifikatsprüfungen, die als Hilfeleistung für Aufgabenersteller und Korrektoren Empfehlungscharakter besitzen sollten. Der Berufsbezug des KMK-Zertifikats findet entsprechend der Formulierungen der Rahmenvereinbarung seinen Niederschlag in einer der ersten dieser Handreichungen. Hier heißt es zu den schriftlichen Prüfungen:

Die Aufgaben sollen möglichst typische Problemstellungen der beruflichen Praxis repräsentieren und kommunikativen Charakter haben. Das entscheidende Erfüllungskriterium einer Aufgabenlösung ergibt sich damit aus der Frage, ob die Kommunikationsaufgabe realisiert worden ist oder nicht (Brandenburg 1999: 22).

Bei den mündlichen Prüfungen wird angeregt:

Da diese Prüfung vorrangig berufliche bzw. berufsrelevante Qualifikationen in der Fremdsprache attestiert, gehen die Inhalte der mündlichen Prüfung ebenfalls über allgemeinsprachliche Bereiche hinaus. Sie können berufsweltbezogen sein, sind im Regelfall jedoch ausgerichtet auf das Berufsbild des Prüflings. Die Rollenspielen zugrunde liegenden Situationen und die vorgelegten Materialien sind realitätsnah und möglichst authentisch. Sie entsprechen dem Leistungsniveau der entsprechenden Stufe I, II oder III (idem.: 30).

Was es bedeutet, die allgemein gehaltenen Deskriptoren des Referenzrahmens auf berufsspezifische Kontexte anzuwenden, wird in einer neueren Handreichung klar und detailliert erörtert:

Die in der KMK-Vereinbarung vorgenommenen Formulierungen zu den einzelnen Stufen (und Kompetenzbereichen) müssen als Versuch angesehen werden, die sehr ausführlich beschriebenen Kriterien des Referenzrahmens in möglichst knapper und damit handhabbarer Weise auf berufsrelevante Fremdsprachenkompetenz anzuwenden (Hessisches Kultusministerium 2003: 8).

Eine frühe Konsequenz aus der soeben skizzierten Problematik war, dass 2001 in einer Evaluation gegenüber der ersten Fassung der Rahmenvereinbarung eine veränderte Gewichtung der Kompetenzbereiche vorgeschlagen wurde. In der ursprünglichen Fassung von 1998 wurde diese Gewichtung für die schriftliche Prüfung wie folgt festgehalten: „Rezeption ca. 50%, Produktion ca. 20%, Mediation ca. 30%“ (zitiert nach Brandenburg 1999: 7). Die 2001 in der Evaluation neu empfohlene Gewichtung lautet „Rezeption ca. 40%, Produktion ca. 30%, Mediation ca. 30%“ (Brandenburg 2001: 13). Dieser Vorstoß wird damit begründet, dass „die modifizierte Gewichtung [...] eine flexiblere Anpassung an Einzelberufe“ erlaube

(ibid.). In der neuen Fassung des Beschlusses der Rahmenvereinbarung wird die vorgeschlagene Neugewichtung übernommen:

Im Rahmen der schriftlichen Prüfung sollen die Aufgabenteile für die drei Kompetenzbereiche wie folgt gewichtet werden:

- Rezeption 40%
- Produktion 30%
- Mediation 30%

Eine Abweichung von jeweils bis zu 10 Prozent-Punkten ist möglich (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland 2002b: 3).

In der hessischen Handreichung wird festgestellt, dass „nicht immer zufrieden stellende Abgrenzungen zwischen den drei Stufen erreicht werden können“ (Hessisches Kultusministerium 2003: 8). Dass die Anwendung von Deskriptoren im Prozess des Lehrens und Beurteilens von Fremdsprachen eine größere Herausforderung bedeutet, ist schon länger bekannt:

Since the empirical status of the task-skills hierarchies described in rating scale descriptors remains unconfirmed, further research is clearly required to establish the extent to which it is possible to link scale descriptors directly to test performance (Brindley 1998: 126).

Die empfundene Schwierigkeit bei der Zuordnungsschärfe der Deskriptoren der verschiedenen Stufen beruht im Bereich des berufsbezogenen Fremdsprachenunterrichts auf dem Zusammenspiel der Faktoren Authentizität und Fachbezug. Douglas, den wir oben als Verfechter eines fachbezogenen Fremdsprachenunterrichts erlebt haben, setzt eindeutig auf Authentizität als Garant eines Fachbezuges: „The test task [...] must be *authentic* for it to represent a specific purpose field in any measurable way“ (Douglas 2000: 6). Doch es stellt sich die Frage, ob sich das Verhältnis zwischen den Bereichen Authentizität, Fachwissen und Sprachkompetenz so eindeutig beschreiben lässt. Zu dieser Frage wird in der hessischen Handreichung festgestellt:

Die Schwierigkeit besteht darin, dass **berufstypische** Texte – vor allem, wenn es sich um authentische Texte handelt – häufig komplex sind und sich ggf. auch durch eine Häufung von Fachterminologie auszeichnen. [...] Lösen lässt sich dieses Problem häufig nur durch die Aufgabe des Anspruchs an Authentizität und manchmal auch durch die Art der Aufgabenstellung (Hessisches Kultusministerium 2003: 8).

Doch es gibt sogar Stimmen, welche behaupten, dass auf höheren Kompetenzniveaus die Bedeutung von Fachwissen wieder absinke:

Only at certain levels of proficiency does subject knowledge help to compensate for lack of proficiency, and once proficiency advances again, subject knowledge becomes less important“ (Fulcher 1999: 232, zitiert nach Grotjahn 2000: 15).

1.6 Testkonstrukt

Eine zentrale Rolle bei der Erstellung von Testaufgaben spielt das Konzept des Testkonstrukts: „the construct is what our test is measuring, and this determines what our scores mean“ (Buck 2001: 95). Der Begriff des Konstruktes kann Aufgabenersteller dafür sensibilisieren, ob die gestellten Aufgaben Fähigkeiten von den Prüfungsteilnehmern verlangen, die irrelevant sind und zu einer Verzerrung der Testergebnisse führen können:

When we think of construct-irrelevant abilities we need to be aware of whether or not they cause *variance* – which is [...] the extent to which the test-takers are spread out from the average (ibid.: 100).

Leicht ist die Wahl einer Aufgabe keineswegs, denn: „we have no way of knowing for certain which competencies are required for any particular task“ (ibid.: 106). Beim Feststellen berufsbezogener Fremdsprachenkenntnisse wie bei den KMK-Zertifikatsprüfungen allerdings besteht die Möglichkeit, mit der Durchführung einer Bedarfsanalyse Konstrukt und Testgestaltung entsprechend zu bestimmen:

In a test of specific purposes [...] a needs analysis [...] can be used as the basis of construct definition and test design (ibid.: 111).

Es wird in der Literatur wiederholt betont, wie wichtig es sei, den Begriff des Testkonstrukts mit dem Testkriterium der Validität in Verbindung zu bringen (vgl. auch Weigle 2002: 51; Vollmer 2003b: 366):

Construct validity is the most important property of a test [...] it is the extent to which the test measures the right construct (ibid.: 195).

In language testing – as in educational measurement generally – construct definition is recognized as having a central role in the validation of tests because at least a working conception of an underlying construct is needed for test results to be meaningfully interpreted (Read / Chapelle 2001: 7).

Validität ist das Konzept, mit dem sich der *Referenzrahmen* besonders befasst. Ein Test oder ein Beurteilungsverfahren kann in dem Maß ‚valide‘ genannt werden, in dem man nachweisen kann, dass das tatsächlich gemessene Konstrukt auch das ist, das in dem betreffenden Kontext gemessen werden soll, und dass die gewonnene Information eine genaue Abbildung der Kompetenz des oder der betreffenden Kandidaten ist (Trim et al. 2001: 172).

Auch wenn für die Messung berufsspezifischer Fremdsprachenkenntnisse Bedarfsanalysen durchgeführt und die zentrale Bedeutung der Verbindung zwischen Testkonstrukt und Validität des Testes im Auge behalten werden, ist es dennoch so, dass auch die Art einzelner Testaufgaben Testteilnehmerleistungen unterschiedlicher Qualität hervorrufen können (vgl. auch Buck 2001: 101):

There seems to be ample documentation that when learner language is elicited using different test methods, learner language can vary systematically in accuracy at all levels (phonological, morphological, syntactic, lexical) in tandem with changes in situational variables such as topic and interlocutor (Tarone 1998: 78).

Zudem muss festgestellt werden, dass nach dem gegenwärtigen Stand der testtheoretischen Forschung der Begriff der Validität gar kein testinhärentes Charakteristikum ist, sondern vielmehr ein Ausdruck der Instrumentalisierung sowohl von Tests als auch von deren Ergebnissen:

Language testers have come to accept that there is no one single answer to the question ‘What does our test measure?’ or ‘Does this test measure what it is supposed to measure’ [...] Validity is not a characteristic of a test, but a feature of the inferences made on the basis of test scores and the uses to which a test is put (Alderson / Banerjee 2002: 79).

Folgerichtig betont auch Vollmer die gesamtgesellschaftlichen Konsequenzen des Nachweises von Konstruktvalidität, denn das Konzept

bezieht sich nicht nur auf die theoretische Plausibilität und Gültigkeit der Operationalisierung des zu messenden Konstrukts, sondern auch auf die Verwendung der Testergebnisse sowie auf Wertimplikationen und soziale Konsequenzen (*impact*) (Vollmer 2003b: 366).

1.7 Der BLK-Modellversuch EU-KonZert

Der im Juli 2003 beantragte, seit November 2003 vom Bundesministerium für Bildung und Forschung geförderte Modellversuch **“Entwicklung und Umsetzung eines Evaluationskonzeptes für die KMK-Fremdsprachenzertifikatsprüfungen zur Sicherung der Vergleichbarkeit der Standards”** (EU-KonZert) evaluiert die länderspezifisch unterschiedlich ausgeprägten Prüfungen.

Insbesondere [werden] dabei die Aspekte

- Aufgabenstellung (Eignung von Aufgabentypen)
- Bewertung (z. B. kriterienorientierte Bewertung)
- Prüfungsdurchführung (insb. mündliche Prüfungen)

auch im Hinblick auf die Erfüllung der Rahmenvereinbarung untersucht [...].

In der praktischen Modellversuchsarbeit wird dabei insbesondere Wert auf die Analyse konkreter Prüfungsbeispiele aus verschiedenen Bundesländern gelegt [...]

Der Modellversuch versteht sich als ein Beitrag zur Qualitätssicherung der KMK-Fremdsprachenzertifikate

(Staatsinstitut für Schulpädagogik und Bildungsforschung 2003: 2).

Der BLK-Modellversuch

ist das erste nationale Projekt, welches sich systematisch mit der Frage der Standards im beruflichen Fremdsprachenunterricht auseinandersetzt und dabei insbesondere den Aspekt der Zertifizierung in den Vordergrund stellt (ibid.).

Präzision und Typologie der Aufgabenstellung sollen untersucht werden, um die Eignung der verwendeten Aufgabentypen zum Testen der zu überprüfenden Kompetenzen auch im Hinblick auf die KMK-Rahmenvereinbarung zu analysieren; Empfehlungen für die konkrete Anwendung der amtlichen Korrekturvorgaben in der Korrekturpraxis sollen erstellt werden; und schließlich sollen aussagekräftige Orientierungshilfen gegeben werden, die es ermöglichen, die spezifisch berufsbezogenen Deskriptoren in der Prüfungssituation sachgemäß anwenden zu können (ibid., 7-8).

Nach Beschreibung der Problemlage und des empfohlenen Lösungsansatzes wurden im Modellversuchsantrag geplante Arbeitsschritte und erwartete Ergebnisse entworfen (ibid., 13-14). Diese sahen vor, in der ersten Phase der Modellversuchslaufzeit (November 2003 – Oktober 2004) als Arbeitsschritt zu einer gemeinsamen Standardfindung bei der Erstellung von schriftlichen Prüfungsaufgaben zu kommen; Ergebnis dieser ersten Phase soll die nun hier vorliegende Veröffentlichung länderübergreifender Muster von schriftlichen Prüfungen sein. Für den weiteren Verlauf des Modellversuchs wurden als Arbeitsschritte und Ergebnisse vorgesehen: In der Phase II (November 2004 – Oktober 2005) die gemeinsame Standardfindung bei der Bewertung schriftlicher Prüfungsleistungen und die Erarbeitung einer Veröffentlichung zur Bewertung solcher schriftlichen Prüfungen sowie in der Phase III (November 2005 – Oktober 2006) die gemeinsame Standardfindung bei der Bewertung mündlicher Prüfungsleistungen und die Erarbeitung einer Veröffentlichung zur Beurteilung dieser Leistungen.

Der Modellversuch wurde von drei Ländern koordiniert (Bayern, Hamburg und Thüringen). Sie repräsentieren einen Querschnitt hinsichtlich verschiedener Formen von Verwaltung und Schulorganisation (Flächenstaat, Stadtstaat, neues Bundesland) und decken gleichzeitig eine erhebliche Zahl der durchgeführten Zertifikatsprüfungen im Bundesgebiet ab. Ihnen wurden jeweils vier angrenzende Länder zugeordnet.

Gemäß dem oben zitierten Zeitplan widmete sich der Modellversuch im ersten Jahr seiner Laufzeit der Frage der Aufgabenstellung. Dazu haben die koordinierenden Länder gemeinsam mit den ihnen zugeordneten Ländern Workshops durchgeführt, um anhand der von den Ländern zur Verfügung gestellten Prüfungsaufgabensätze zunächst den Ist-Stand zu analysieren, Empfehlungen für die Prüfungsgestaltung abzuleiten und Best-practice-Beispiele zu identifizieren als Grundlage für den zweiten Teil der Handreichung.

Der folgende, zweite Teil dieser Handreichung dokumentiert die Ergebnisse der ersten Phase des Modellversuchs und legt länderübergreifende Muster von schriftlichen Prüfungen vor, die für diejenigen eine Unterstützung sein sollten, die Aufgaben für die schriftlichen KMK-Zertifikatsprüfungen erstellen.

1.8 Ausblick

Protokolle und Stellungnahmen lassen erkennen, dass auch nach Abschluss dieser Phase des Modellversuchs eine Reihe von Fragen noch der intensiveren Beschäftigung bedürfen. So kann zwar festgestellt werden, dass sowohl die Prüfungssätze, die in den jeweiligen Arbeitsmappen und Protokollen der Workshops dokumentiert wurden, als auch die Anzahl an neu erstellten ‚Best-practice-Beispielen‘, die nach Abschluss der Phase I der wissenschaftlichen Begleitung von Prüfungserstellern zugesandt wurden, durchaus einen Variantenreichtum der Aufgabentypen belegen. Dennoch scheint auch am Ende der ersten Phase dieses Modellversuchs eine intensivere Beschäftigung mit den Möglichkeiten unterschiedlicher Aufgabentypen ein bleibendes Desideratum der wissenschaftlichen Fundierung des Prüfungsverfahrens des KMK-Zertifikates zu bleiben. An dieser Stelle sei es deswegen vielleicht hilfreich, auf eine Reihe von teils älteren, aber noch relevanten, teils neueren Veröffentlichungen sowohl zu Übungs- als auch zu Prüfungsformen zu verweisen, die bezüglich einer Aufgabenvielfalt mit Gewinn zur Hand genommen werden können: Doyé 1986; Bundesarbeitsgemeinschaft Englisch an Gesamtschulen 1996; Willis 1996; Staatsinstitut für Schulpädagogik und Bildungsforschung 1998; Brandenburg 1999; Zydatið 2002, insbesondere Abschnitt 2.2.3; Weir 2005, insbesondere 119-176 (Kapitel 8: „Response formats“).

Der BLK-Modellversuch EU-KonZert wurde konzipiert als Forschungs- und Fortbildungsprojekt zugleich. Dieser Ansatz wurde gewählt, da der erfolgreiche Einsatz eines Bewertungsverfahrens, das auf Standards beruht, mit konzentriertem gemeinsamen Arbeiten und vertieftem Meinungsaustausch einhergehen muss:

Ein [...] wichtiges Verfahren zur Objektivierung ist die Schulung der Beurteiler und Lehrenden, die mit den festgelegten Kriterien arbeiten sollen (Glaboniat 1998: 20).

It is generally acknowledged that the development of a standards-oriented approach takes time, as partners acquire a feel for the meaning of the standards through the process of exemplification and exchange of opinions (North et al. 2003: 21).

The crucial part in every standard setting procedure is the training of the judges (Kaftandjieva 2004: B5).

If new training is not provided, we should hardly find it surprising that old methodologies persist (Weir 2005: 38).

Die Konzeption des Modellversuchs, die für das Abhalten der Workshops sowie die Genese dieser Handreichung entscheidend war, entspricht dem iterativen Modell für die Erstellung von Sprachtests, die zunehmend an Bedeutung gewinnt:

We advocate tests that are developed in an iterative manner: there are cycles of feedback-laden improvement over time as the test grows and evolves (Davidson / Lynch 2002: 7; vgl. zur iterativen Testerstellungsmethodik jetzt auch Luoma 2004: 128).

Ihre Darstellung der Methodik des Erstellens und der Anwendung von Standards fasst die oben zitierte Beauftragte des Europarates in Anlehnung an das Buch *Le petit prince* von Saint-Exupéry mit der Bemerkung zusammen: „Nothing is perfect“ (Kaftandjeva 2004: B31). Zu Goethes Zeiten sprach man weniger von Perfektion denn von Vollkommenheit und so befindet der Dichter: „Vollkommenheit ist die Norm des Himmels, Vollkommenes wollen die Norm des Menschen“ (Hamburger Ausgabe' XII: 531).