

IV Rück- und Ausblick

Micheál Ó Dúill

In diesem Schlusswort möchte ich den Gesamtprozess des BLK-Modellversuchs noch einmal Revue passieren lassen, um abschließend einige Anregungen zu einer möglichen Weiterentwicklung der KMK-Fremdsprachenzertifikate zu geben.

Dieser letzte Teil der Handreichung ist wie folgt aufgebaut:

1. Ausgangsbasis für den BLK-Modellversuchsantrag
2. Anderweitige Forschungsergebnisse im Bereich der Testtheorie seit Antragstellung
3. Gesellschaftspolitische Entwicklung der Wahrnehmung und Wertung bestimmter testtheoretischer Ansätze
4. Ziele des Modellversuchs, wie diese im Antrag formuliert wurden
5. Weiterentwicklung der Ziele und Ergebnisse, die während der Laufzeit des Modellversuchs erarbeitet wurden
6. Zukünftige Entwicklungsmöglichkeiten für die KMK-Fremdsprachenzertifikatsprüfungen

1 Ausgangsbasis für den BLK-Modellversuchsantrag

Der Modellversuchsantrag betont, dass das Vorhaben an die früheren Modellversuche „Fremdsprachen an der Berufsschule – Chancen für den Arbeitnehmer in der EG von morgen“ (1990-1993) sowie „Fremdsprachenpflichtunterricht in der Berufsschule“ (1994-1997) anknüpft.

Eine der Veröffentlichungen zum erstgenannten Modellversuch unterstreicht gleich zwei Mal die Bedeutung von Tests und bringt diese in Verbindung mit - unter anderem - der Möglichkeit einer zukünftigen Zertifizierung:

Folgende Gründe sprechen für die Notwendigkeit von Tests:

(...)

- Motivation zum weiteren Lernen
- Information für den Arbeitgeber (z. B. durch Zertifikate)
- "Backwash Effect": Neben dem Lehrplan bestimmt die Prüfung, die es zu bestehen gilt, Inhalt und Art des Unterrichts
(Baumgartner / Gibson 1992: 6; Vgl. ebenda, S. 9).

Baumgartner und Gibson sprechen von Leistungsmessung als „wünschenswertes Ziel für die Zukunft wegen [...] heterogener Lerngruppen“ (ebenda, S. 9). In einem von einer Reihe von Arbeitsberichten zum zweiten oben erwähnten Modellversuch stellen Weidinger und Träger sechs Jahre später fest:

Der Arbeitskreis einigte sich rasch darauf, dass ein auf unterschiedlichen Niveaus angebotenes freiwilliges Fremdsprachenzertifikat den Rahmenbedingungen des Englischunterrichts an der Berufsschule am Besten gerecht wäre, weil es allen Schülerinnen und Schülern die Möglichkeit gibt, sich im Englischunterricht zu engagieren, ohne durch die Konzentration auf ein einziges Abschlussniveau eine Über- oder Unterforderung erleben zu müssen
(Weidinger / Träger 1998: 18).

Die testtheoretische Grundlage für die vorgesehene Untersuchung der drei Bereiche Aufgabenstellung, Bewertung und Prüfungsdurchführung waren unter

anderem folgende Arbeiten: Buck 2001; Schneider/North 2000: 140; und Krause / Sändig 1999. Schneider und North stellen in ihrer Untersuchung fest, dass

Operationalisation of the scale in appropriate rating instruments [...] training in relation to the overall framework [...], training in the use of particular assessment instruments, and the validation of such practical applications remain essential steps in the development of a coherent assessment framework (Schneider / North 2000: 140).

(Zu Bucks Ausführungen zum Testkonstrukt, vergleiche Abschnitt I. 1.6 oben; zum Thema Wirkung eines Kommunikationsbeitrages, Abschnitt II. 1.5; zum Durchführbarkeitsgrad von Prüfungsformen, siehe Krause / Sändig 1999: 102 und Staatsinstitut 2003: 11.)

Die Antragstellung wurde auch von dem Wunsch nach einer langfristigen Modernisierung des Fremdsprachenunterrichts getragen:

Wird der Ansatz des GERs [= Gemeinsamer europäischer Referenzrahmen, Ó D] nicht nur wegen der Anregungen, die er bezüglich Prüfungsmodi liefert, gelobt, sondern auch deswegen, weil hierdurch ‚eine langfristige Modernisierung von Fremdsprachenunterricht möglich wird‘ (Krumm 2003: 120), so gilt dies mindestens genauso für den Versuch, die Anregungen des GERs im Sektor der beruflichen Schulen Deutschlands umzusetzen. (Staatsinstitut 2003: 12).

Hinzu kam im Antrag die Berücksichtigung der Bedeutung von Authentizität für den berufsbezogenen Fremdsprachenunterricht sowie des Wechselverhältnisses zwischen Sprachfertigkeit und Fachwissen (Alderson / Banerjee 2001: 222; Douglas 2000: 2, 6; Siehe Abschnitte I. 1.4 und I. 1.5 oben).

Der Antrag berücksichtigt den Ruf de Jongs in seiner Evaluierung der Rahmenvereinbarung nach einer detaillierten Evaluation der KMK-Prüfungen und deren Bewertungsinstrumente sowie seine Aufforderungen, Musterprüfungsleistungen zu sammeln und Testergebnisse statistischen Analysen zu unterziehen:

An evaluation of the levels as operationalized in the examinations and as actually attained by vocational students would need to be based on a detailed evaluation of the examination papers, including an analysis of scoring procedures, samples of student work and statistical analyses of examination results (De Jong 2001: 1, zitiert nach Staatsinstitut 2003: 15).

Als von maßgeblicher Bedeutung für das geplante Vorhaben werden entsprechende Arbeiten des Europarates im Antrag erwähnt. Dazu zählen die Arbeiten zu berufsbezogenem Fremdsprachenunterricht. Hier rufen etwa Egloff und Fitzpatrick nach umfassenden Arbeiten auf dem Feld des berufsbezogenen Fremdsprachenunterrichts, deren Ergebnisse frei zugänglich gemacht werden, insbesondere in den Bereichen der Evaluierung und Zertifizierung:

[T]here is a need for a comprehensive and principled approach whose results are well documented and made readily accessible. These needs are most urgent in ... evaluation and certification (Egloff / Fitzpatrick 1997: 21).

Dazu zählen auch die Erfahrungen und Ergebnisse der Arbeitsgruppe Gemeinsamer europäischer Referenzrahmen für Sprachen des Europarates, die zu den Anregungen führen,

dass man Beispiele von Arbeiten in Bezug auf standardisierte Definitionen und Beispiele vergleicht (Benchmarking)

und

dass man in Diskussionen ein gemeinsames Verständnis aufbaut (gemeinsame Standardfindung)

(...) Man ist sich darüber einig, dass die Entwicklung eines standardorientierten Ansatzes zeitaufwendig ist, weil die am Verfahren Beteiligten nur durch einen Prozess der Veranschaulichung und des Austausches von Meinungen ein Gefühl dafür erwerben, was bestimmte Standards bedeuten
(Trim et al. 2001: 177, zitiert nach Staatsinstitut 2003: 15-16).

Konkret werden Anregungen des Referenzrahmens aufgegriffen, in Diskussionen zur gemeinsamen Standardfindung zu gelangen etwa dadurch, dass Beispiele von Prüfungsleistungen unter Berücksichtigung tatsächlich vorgenommener Bewertungsverfahren verglichen werden.

2 Anderweitige Forschungsergebnisse im Bereich der Testtheorie seit Antragstellung

Der Natur eines solchen Projektes entsprechend wurden relevante neue Veröffentlichungen berücksichtigt, die während des Verlaufs des Modellversuchs erschienen sind. Da der Modellversuch in drei einjährige Phasen aufgebaut war, die sich je anderen Fragestellungen gewidmet haben, wurden vor allem diejenigen Neuveröffentlichungen rezipiert, die sich mit den jeweiligen Aufgabenstellungen beschäftigt haben. Aus diesem Grund soll an dieser Stelle eine abschließende Übersicht erstellt werden zu anderen Forschungsarbeiten, die während der Gesamtlaufzeit des Modellversuchs erschienen sind und für dessen Thematik von Bedeutung sind, jedoch wegen der phasenmäßigen Arbeitsweise unseres Projektes noch nicht in dieser Handreichung erwähnt wurden. Dabei geht es nicht mehr darum, wie während der Laufzeit des Modellversuchs, eine detaillierte Untersuchung der Bedeutung dieser Veröffentlichungen für das Projekt zu liefern, sondern darum, die Existenz solcher neueren Arbeiten zu dokumentieren, die für eine weitere wissenschaftliche Beschäftigung mit dem KMK-Fremdsprachenzertifikat wichtig sein könnten.

Im Folgenden versuche ich das Interesse der Forschung für verschiedene Bereiche zu dokumentieren, wie sich dieses in den Veröffentlichungen, die ich habe lesen können, in dieser Zeit widerspiegelt. Ich habe diese Übersicht nach von den Autorinnen und Autoren genannten Schlagwörtern eingeteilt und meine Übersicht nach der Anzahl der Themennennungen gegliedert. Danach ergibt sich folgende Auflistung:

- allgemeine Abhandlungen der Leistungsmessung und des Testens im Bereich des Fremdsprachenerwerbs;
- Diskussion um Bildungsstandards;
- Veröffentlichungen zu gattungsbezogenen Teilleistungen (schriftliche Prüfungen; mündliche Prüfungen).

Ich schließe diesen Abschnitt mit Informationen zu gegenwärtig erscheinenden Fachzeitschriften und Reihen sowie zu den Arbeiten zweier Gesellschaften, die international in der Analyse und Bewertung von Sprachprüfungen tätig sind, ILTA und EALTA.

2.1 Allgemeine Abhandlungen zur Leistungsmessung und zum Testen im Bereich des Fremdsprachenerwerbs

Im Jahr 2003 wurde Lynchs Monograph *Language Assessment and Programme Evaluation* verlegt. Das Jahr 2005 brachte eine Reihe von Übersichtsartikeln, von Hudson; McKay; und Norris und Ortega. Vor allem erschien im groß angelegten, von Hinkel herausgegebenen *Handbook of Research in Second Language Teaching and Learning* der von McNamara und anderen herausgegebene umfangreiche Teil „Second language testing and assessment“ mit sieben Beiträgen (Hinkel 2005: 773-888). Andere Abhandlungen jenes Jahres spiegeln persönliche Erfahrungen wider (Tomlinson); beziehen sich speziell auf einen nationalen Kontext (Zydatið 2005b); oder widmen sich bestimmten bewertungsrelevanten Begriffen (Sorace zu „Near-nativeness“).

Einige andere Veröffentlichungen weisen auf allgemeine Veröffentlichungstrends hin: Savilles Aufsatz (2003) im Sammelband von Weir und Milanovic sowie Hawkeys Monograph (2004) sind charakteristisch für die Reihe „Studies in Language Testing“ vom University of Cambridge Local Examinations Syndicate (UCLES, neuerdings „Cambridge Assessment“ genannt), die sich zu einem großen Teil (aber nicht ausschließlich) der wissenschaftlichen Evaluierung von Prüfungen dieses Großanbieters widmet.

Gottlieb sowie Rivera und Collum (beide 2006) sind repräsentativ für die florierende Beschäftigung in den Vereinten Staaten von Amerika mit den Bedürfnissen von Schülern, die Englisch als Zweitsprache erlernen. Von ähnlicher Bedeutung sind deutsche Arbeiten zu Sprachstandsmessungen bei Schülern, die Deutsch nicht als Muttersprache haben (z.B. Reich 2003).

Furedi (2004) gibt eine subjektive Stellungnahme zum Stellenwert des Testens im universitären Bereich ab, die für die bestehende Skepsis einiger Intellektuellen gegenüber einer vermeintlich zunehmenden Testwut charakteristisch ist.

2.2 Diskussionen um den Gemeinsamen europäischen Referenzrahmen und Bildungsstandards der Kultusministerkonferenz

Es findet eine zunehmende wissenschaftliche Auseinandersetzung mit dem Einfluss des Gemeinsamen europäischen Referenzrahmens sowie mit den Bildungsstandards der Kultusministerkonferenz statt (zu den Bildungsstandards der KMK, Vergleiche unten, Abschnitt IV.3).

Neuere Beiträge zum Referenzrahmen wurden von Quetz (2004), Figueras et al. (2005) sowie Little (2006) verfasst.

Zu den Veröffentlichungen, welche versuchen, Sinn und Auswirkungen der Bildungsstandards auf das deutsche Bildungswesen zu untersuchen gehören Bausch et al. (2005); Zydatið (2005a); sowie Hallet und Müller-Hartmann (2006).

2.3 Einzelveröffentlichungen zu Teilaspekten des Testens

Ein umfangreicher Übersichtsartikel zum Hörverstehen wurde 2005 von Grotjahn vorgelegt.

Zahlreiche Untersuchungen zur mündlichen Produktion wurden unternommen: Yuan und Ellis (2003) zum Einfluss von Vorbereitungszeit; Hasselgreen (2004) zur Rolle von „smallwords“ (ein Beispiel eines Bandes in der Reihe „Studies in Language Testing“, der keine UCLES-Prüfung untersucht); Kormos und Dénes (2004), ebenfalls, wie Hasselgreen, ein Beitrag zum englischen Begriff „fluency“;

Johnson Gerson (2006) zur Bedeutung der Fähigkeit, dialogisch geprägte Unterhaltungen führen zu können; sowie Levis (2006) zur Rolle der Aussprache.

Alderson und andere haben im Auftrag der niederländischen Regierung den einflussreichen Versuch unternommen, vor dem Hintergrund des Gemeinsamen europäischen Referenzrahmens Hilfestellung zur Entwicklung von Testitems in den Bereichen Lesen und Hörverstehen zu geben; dieser unveröffentlichte Bericht wurde bereits berücksichtigt bei der Erstellung des kürzlich erschienenen *Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment. Writing Tasks: Pilot Samples* (s. S. 1).

Hyland (2003) hat einen Monograph zum Bereich Schreiben vorgelegt. Hawkey und Barker haben über eine Skala zum Messen der Schreibkompetenz berichtet (Hawkey / Barker 2004). Auch in den Vereinten Staaten von Amerika werden Forschungsarbeiten zur Entwicklung von Testitems zum Messen von Lesekompetenz vorgenommen (Bailey et al. 2005). Elliot (2005) untersucht auf eindringliche Weise die gesellschaftliche Kontrollfunktion des Messens der Schreibkompetenz.

Dem bisher wenig erforschten Bereich des Messens pragmatischer Kompetenzen haben sich Röver (2005) und Liu (2006) gewidmet. Wall (2005) hat eine umfassende Untersuchung der Rückkoppelungseffekte von Tests auf den Unterricht vorgelegt. Purpura (2004) hat sich dem Themenbereich Grammatik gewidmet.

Für das Projekt der KMK-Zertifikatsprüfungen von besonderer Bedeutung seien zuletzt zwei Arbeiten genannt, die sich mit dem Gebiet des Testens fachbezogener Fremdsprachenkenntnisse beschäftigen: Douglas (2005) untersucht im Sammelband von Hinkel noch einmal diese Thematik, über welche er bereits 2000 seinen einflussreichen Monograph vorlegte. Speziell zu den „Cambridge Business English Certificates“ legt O’Sullivan (2006) eine Untersuchung vor.

2.4 Reihen zum Sprachtesten, Fachzeitschriften und internationale Organisationen

2.4.1 Bücherreihen

Ein Indiz für die Bedeutung, welche einer Teildisziplin im gegenwärtigen Wissenschaftsdiskurs beigemessen wird, ist Fähigkeit des Marktes, das Überlegen bestehender Reihen und Fachzeitschriften zu ermöglichen und sogar das regelmäßige Erscheinen neuer Publikationen zu ermöglichen.

Im vorhergehenden Abschnitt habe ich speziell einige neuere Titel genannt, deren Lektüre ich all denjenigen empfehlen kann, die sich mit der Weiterentwicklung des KMK-Fremdsprachenzertifikats beschäftigen werden.

In diesem Abschnitt möchte ich nun allgemeinen auf das Bestehen und Neuerscheinen einer Reihe von Publikationsorganen hinweisen, welche zusammengenommen die wohl tatsächlich zunehmende Bedeutung einer wissenschaftlich fundierten Testkultur zu belegen scheinen.

Die wiederholt von mir erwähnte Reihe „Studies in Language Testing“ erscheint für das University of Cambridge Local Examinations Syndicate seit 1995 bei der Cambridge University Press und wird zur Zeit von Michael Milanovic und Cyril Weir herausgegeben. In dieser Reihe sind bisher (Stand: November 2006) 24 Titel erschienen; zwei weitere sind für das Jahr 2007 bereits angekündigt.

Eine zweite Reihe erscheint seit 2000, ebenfalls bei Cambridge University Press, die von J. Charles Alderson und Lyle F. Bachman herausgegebene Serie „Cambridge Language Assessment Series“; bisher sind elf Titel erschienen.

Seit 2004 erscheint die Reihe „Language Testing and Evaluation“ beim Verlag Peter Lang in Frankfurt am Main, herausgegeben von Rüdiger Grotjahn und Günther Sigott. Sechs Bände sind bisher erschienen.

2.4.2 Fachzeitschriften

An dieser Stelle zu nennen sei nicht nur die bekannte Vierteljahreszeitschrift *Language Testing*, die sich seit ihrem ersten Erscheinen 1984 sehr gut auf diesem Markt etabliert hat, sondern neuerdings auch *Language Assessment Quarterly*, ebenfalls eine Vierteljahreszeitschrift, die seit 2004 verlegt wird.

2.4.3 Internationale Organisationen

Wenn hier unter der Rubrik „Anderweitige Forschungsergebnisse im Bereich der Testtheorie seit Antragstellung“ über die Arbeit der beiden Organisationen ILTA und EALTA berichtet wird, dann deswegen, weil in einer Zeit zunehmender Benützung des Internets der Netzauftritt solcher Organisationen neben dem Buchhandel eine Informationsquelle ersten Ranges ist (ähnliches gilt seit langem für die Netzseiten der „Language Policy Division“ des Europarates, auf Englisch abrufbar unter http://www.coe.int/T/DG4/Linguistic/Default_en.asp, auf Französisch unter http://www.coe.int/t/dg4/linguistic/default_FR.asp).

Von besonderem Interesse dürften meiner Ansicht nach gegenwärtig der sich noch im Entwurfsstadium befindende „Draft Code of Practice“ der International Language Testing Association ILTA (<http://www.iltaonline.com/ILTA-COP-ver3-21Jun2006.pdf>) sowie die bereits verabschiedeten „Good Practice Guidelines“ der European Association for Language Testing and Assessment EALTA sein (http://www.ealta.eu.org/documents/archive/Good_Practice_Guidelines_21-05-2006.pdf).

Des Weiteren sei etwa auf die seit 2004 erscheinenden Tagungsdokumentationen der EALTA (<http://www.ealta.eu.org/resources.htm>), sowie auf die „Publications“-Seite der ILTA hingewiesen (http://www.iltaonline.com/ILTA_pubs.htm).

3 Gesellschaftspolitische Entwicklung der Wahrnehmung und Wertung bestimmter testtheoretischer Ansätze und Reaktionen der Kultusministerkonferenz auf Ergebnisse internationaler Erhebungen

Dieser Abschnitt soll verdeutlichen, dass es bereits vor und vor allem während der Laufzeit des Modellversuchs eine Intensivierung des Stellenwerts bestimmter Aspekte gegeben hat, die für unser Projekt von zentraler Bedeutung waren, nämlich Konzepte einer möglichen Standardsicherung im Bildungsbereich und deren Evaluierung. Diese Intensivierung beruhte zum Teil auf Veröffentlichungen von wissenschaftlichen Ergebnissen zu Untersuchungen, die von öffentlicher Hand in Auftrag gegeben wurden; zum anderen fanden sie auf Grund des Beschließens neuerer Bildungsmaßnahmen statt, die ihrerseits zum Teil als eine Reaktion auf die Veröffentlichungen der Ergebnisse der in Auftrag gegebenen Untersuchungen waren.

Bereits 1997 beschloss die Kultusministerkonferenz auf ihrer 280. Plenarsitzung am 23./24. Oktober:

Die Durchführung regelmäßiger länderübergreifender Vergleichsuntersuchungen zum Lern- und Leistungsstand von Schülerinnen und Schülern ausgewählter Jahrgangsstufen an allgemeinbildenden Schulen ist [...] eine wichtige Ergänzung der länderbezogenen Qualitätssicherungsmaßnahmen und ermöglicht für jedes Land Rückschlüsse im Hinblick auf die jeweils gewählten Methoden und Maßnahmen zur Qualitätssicherung.

Entsprechend dem Beschluss zu den ‚Standards für den mittleren Schulabschluss‘ sollten vor allem [...] **fremdsprachliche Kompetenzen** Beachtung finden (Sekretariat der Ständigen Konferenz der Kultusminister: 1997).

In dem entsprechenden KMK-Beschluss vom 12.05.1995 hieß es in den allgemeinen Vorbemerkungen interessanterweise: „Die Beschreibung des Anspruchsniveaus erfolgt nicht vorrangig unter dem Gesichtspunkt der Leistungsfeststellung“ (Sekretariat 1995: 14).

In der Mitteilung zu dem KMK-Beschluss von 1997 wurde das Entstehen der Diskussion um Standards eindeutig als Reaktion auf die Ergebnisse einzelner internationaler Studien dargestellt:

Gegenstand der Beratungen der Kultusministerkonferenz waren auch die **Ergebnisse und Auswirkungen der 3. Internationalen Mathematik- und Naturwissenschaftsstudie (TIMMS)** sowie der **OECD-Studie "Education at a Glance"**. (Sekretariat: 1997).

Dabei stellt die Kultusministerkonferenz laut der entsprechenden Pressemitteilung fest,

dass in der künftigen Diskussion der Ergebnisse und Auswirkungen der Studien zu den Leistungen deutscher Schülerinnen und Schüler in den mathematisch-naturwissenschaftlichen Fächern als übergreifender Aspekt die Entwicklung einer neuen **Kultur der Anstrengung** von Bedeutung ist (ebd.).

Von Sprachkenntnissen ist noch nicht die Rede.

Im Jahr 2000 wurde die erste der alle drei Jahre stattfindenden internationalen Untersuchungen des Programme for International Student Assessment (PISA) der Organisation for Economic Co-operation and Development (OECD) unternommen. Diese erste Untersuchung widmete sich schwerpunktmäßig der muttersprachlichen Lesekompetenz von 15-Jährigen (die darauf folgenden Untersuchungen in den Jahren 2003 bzw. 2006 beschäftigten sich hauptsächlich mit Mathematik bzw. Naturwissenschaften, 2009 wird wieder das Lesen Schwerpunkt sein; eine neuere übersichtliche Darstellung dieser Untersuchungen findet sich auf <<http://www.pisa.oecd.org/dataoecd/51/27/37474503.pdf>>).

Als am 4. Dezember 2001 die Ergebnisse der ersten PISA-Untersuchung mittlere Ergebnisse für die 15jährigen in Deutschland deutlich unter dem OECD-Durchschnitt bekannt geben, führt dies zu einer sofortigen Stellungnahme der Präsidentin der Kultusministerkonferenz: "Mit PISA verfügen wir nun über empirische Befunde, die eine Versachlichung der Diskussion geradezu verlangen" (Sekretariat 2001).

Diese Achtung vor empirischen Befunden kennzeichnet auch eines der drei Monate später von der KMK bekannt gegebenen Handlungsfelder, die als Konsequenzen aus der PISA-Studie entstanden sind:

Handlungsfeld 5: Maßnahmen zur konsequenten Weiterentwicklung und Sicherung der Qualität von Unterricht und Schule auf der Grundlage von verbindlichen Standards sowie eine ergebnisorientierte Evaluation

Neufassung von Rahmenlehrplänen; Erstellung von Schulprogrammen; Durchführung und Auswertung von Vergleichsarbeiten (auch schulübergreifend); Intensivierung der externen Evaluation; Erarbeitung von Standards in den Kernfächern; Qualitätsmanagement an Schulen.

(Sekretariat 2002a)

Im folgenden Jahr wurde die Expertise *Zur Entwicklung nationaler Bildungsstandards vom Bundesministerium für Bildung und Forschung* verlegt. Wie bereits im Abschnitt I.1.3 oben berichtet, legte diese Expertise besonderen Wert auf den Gemeinsamen europäischen Referenzrahmen (Klieme et al. 2003: 147).

Im gleichen Jahr wurden die *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss* von der Kultusministerkonferenz verabschiedet (Sekretariat 2004). In der entsprechenden Vereinbarung vom 4. Dezember 2003 heißt es

Die Standards [...] *beziehen* international anerkannte Standardmodelle – u.a. theoretische Grundlagen der PISA-Studie und den Gemeinsamen europäischen Referenzrahmen für Sprachen – *ein*.

(ebd.: 3-4)

Im eigentlichen Beschluss ist dann von einer „*Orientierung*“ dieser „verbindliche[n] Standards“ an den Gemeinsamen europäischen Referenzrahmen die Rede (ebd.: 6); „Die vorliegenden Standards *beziehen sich auf* die Niveaustufe B 1 (ebd.: 7). Die Beschreibung der Standards für Kompetenzbereiche der ersten Fremdsprache „erfolgt *in enger Anlehnung an* den Gemeinsamen europäischen Referenzrahmen“ (ebd.: 11, Hervorhebungen von Ó D).

2004 wurden anschließend die *Bildungsstandards für die erste Fremdsprache Englisch/Französisch, Jahrgangsstufe 9* verabschiedet (Sekretariat 2005b). Auch hier ist die Rede von einer „*Orientierung an dem vom Europarat herausgegebenen [...] Gemeinsamen europäischen Referenzrahmen für Sprachen*“ (ebd.: 6-7); die Standards „*beziehen sich [...] nahezu durchgängig auf* die Niveaustufe A 2“ (ebd.: 7). Auch hier erfolgt die Beschreibung der Standards für Kompetenzbereiche „*in enger Anlehnung an* den Gemeinsamen europäischen Referenzrahmen“ (ebd.: 11; Hervorhebungen von Ó D).

Ergänzend zu diesen KMK-Veröffentlichungen erschien schließlich die Veröffentlichung *Bildungsstandards der Kultusministerkonferenz. Erläuterungen zur Konzeption und Entwicklung. (Am 16.12.2004 von der Kultusministerkonferenz zustimmend zur Kenntnis genommen.)* (Sekretariat 2005a). Hier wird der Referenzrahmen eine „wichtige Grundlage“ genannt (ebd.: 15); die Standards würden diesen als „international anerkannte[r] Standardmodell“ einbeziehen (ebd.: 16), die Aufgabenbeispiele würden sich an ihm „orientieren“ (ebd.: 17).

Im Auftrag der Kultusministerkonferenz fand zeitgleich zu diesen Entwicklungen die Untersuchung DESI statt: „Die Studie ‚Deutsch-Englisch-Schülerleistungen-International‘ (DESI) untersucht die sprachlichen Leistungen von Neuntklässlern und die Unterrichtswirklichkeit in den Fächern Deutsch und Englisch“ (Klieme et al. 2006: 1). (Kompetenzen im Deutschen waren Teil der ersten PISA-Untersuchung, Englischkenntnisse aber nicht.)

Wie sieht es aus mit der zugenommenen Betonung empirischer Forschungsvorhaben, mit der Einbindung der Arbeiten des Europarates bei DESI bzw. mit dem Einfluss der entwickelten Bildungsstandards?

Die KMK-Bildungsstandards setzen als Ziel für den Hauptschulabschluss prinzipiell das Niveau A2 des Europäischen Referenzrahmens, für den mittleren Bildungsabschluss nach Klasse 10 das Niveau B1. Diese Zielsetzungen können im Rahmen von DESI beim computergestützten Test für mündliches Sprechen explizit überprüft werden, weil dieser international eingesetzte Test an den Niveaus des Referenzrahmens normiert wurde. Die bei DESI selbst entwickelten Tests haben hingegen nicht den Anspruch, die europäischen Niveaus gleichsam „eins zu eins“ abzubilden, da sie Aufgabenmaterial verwenden, das auf den Englischunterricht und die Lehrpläne in Deutschland abgestimmt ist. Es wird versucht, jene Kompetenzbereiche, die im Zielbereich der Lehrpläne und Bildungsstandards liegen, so differenziert wie möglich zu erfassen. Daher *orientieren* sich die DESI-Kompetenzmodelle am Europäischen Referenzrahmen für Sprachen, enthalten aber eigene, je nach Messbereich unterschiedlich stark ausdifferenzierte Niveaubeschreibungen (Klieme et al. 2006: 11).

Was bedeutet im zweiten Satz dieses Zitats „an den Niveaus des Referenzrahmens normiert“? Hierzu heißt es an anderer Stelle:

Die Fragen dieses Testteils wurden sowohl in ihrer inhaltlichen als auch sprachlichen Komplexität speziell auf die Besonderheiten der Altersgruppe der DESI-Studie und die zu erwartenden Kompetenzniveaus des GER zugeschnitten [...]

Das computergesteuerte Testprogramm errechnete für die Testteile, die eindeutig vorhersagbare Elemente enthalten, Einschätzungen der Testleistungen in den Teilkomponenten Aussprache, Flüssigkeit der Rede, Wortschatz und Satzbau (*subscores*) und erstellte zusätzlich eine Gesamteinschätzung (*overall score*). Diese Einschätzungen sind auf die Niveaus der Sprechfähigkeit des Gemeinsamen Europäischen Referenzrahmens für Sprachen (GER) abgestimmt (ebd.: 17).

Die Arbeit des DESI-Konsortiums ist unter anderem deswegen interessant, weil im Bereich „mündliches Sprechen“ diese Normierung eines bereits vorhandenen Tests an den Niveaus des Referenzrahmens vorgenommen wurde. Wie genau diese Normierung erfolgte, dürfte bei Überlegungen zu einer Weiterentwicklung des KMK-Fremdsprachenzertifikats (Abschnitt IV. 6 unten) von Bedeutung sein, ebenfalls die Art und Weise, in der sich die anderen DESI-Kompetenzmodelle am Europäischen Referenzrahmen „*orientieren*“.

Die DESI-Studie scheint die Akzeptanz eines maßgeblichen Einflusses empirischer Untersuchungen auf die Sekundarstufe des Bildungswesens zu belegen. Welche Konsequenzen hat das für den Unterricht? Für die Evaluation der Lehr- und Lernangebote? Für den berufsbildenden Schulsektor? Sowie für das Testen erreichter oder irgendwie bereits vorhandener oder gar parallel zum Unterricht autonom erworbener Kompetenzen?

Auf diese Fragen werde ich im Abschnitt IV.6 unten nochmals zu sprechen kommen.

4 Ziele des BLK-Modellversuchs, wie diese im Antrag formuliert wurden

Dieser Abschnitt wird kurz die Ziele des Modellversuchs beschreiben, wie diese im Antrag formuliert wurden, bevor im nächsten Abschnitt die Entwicklung dieser Ziele während der Laufzeit des Modellversuchs beschrieben wird (vgl. hierzu auch Abschnitte I.1.7, II.1.1 und II.1.3 oben).

In diesem Modellversuch soll nun eine Evaluation der länderspezifisch unterschiedlich ausgeprägten Prüfungen erfolgen.

Insbesondere sollen dabei die Aspekte

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

auch im Hinblick auf die Erfüllung der Rahmenvereinbarung vergleichend untersucht werden.

In der praktischen Modellversuchsarbeit wird dabei insbesondere Wert auf die Analyse konkreter Prüfungsbeispiele aus verschiedenen Bundesländern gelegt, wobei auch die Einbeziehung nicht direkt am Modellversuch beteiligter Länder vorgesehen ist, um in möglichst vielen Fällen das vorhandene Expertenwissen aufzugreifen, zu systematisieren und nachhaltig verfügbar zu machen.

Ziel der Untersuchung ist 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.

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

(Staatsinstitut 2003: 2; Vgl. auch Abschnitt I. 1.7 oben).

Innovativ am Modellversuch war, auch daran sei an dieser Stelle erinnert, dass er

das erste nationale Projekt [war], welches sich systematisch mit der Frage der Standards im beruflichen Fremdsprachenunterricht auseinandersetzt und dabei insbesondere den Aspekt der Zertifizierung in den Vordergrund stellt. [...]

Das Modellvorhaben folgt dabei bewusst einem sehr **praxisorientierten Ansatz**, in dem in mehreren **länderübergreifend besetzten Landesgruppen** vorhandene Erfahrungen und Ansätze reflektiert, gebündelt und systematisiert werden. [...]

Das Modellvorhaben stärkt durch eine transparente Orientierung an Standards die Wertigkeit der KMK-Fremdsprachenzertifikate und fördert damit deren Akzeptanz im europäischen Raum. [...]

Eine wie [vom Referenzrahmen] vorgeschlagene Datenerhebung von Prüfungsleistungen wird eine erste vertiefte Untersuchung des Verhältnisses Korrekturvorgaben/eigentliche Korrekturpraxis ermöglichen.

Diese Datenmenge lässt eine [vom Referenzrahmen] angeregte Überprüfung der Korrelation zwischen angegebener Kompetenzstufe und tatsächlichem Leistungsniveau der Prüfungsarbeiten zu

(ebd., S. 2-3).

5 Weiterentwicklung der Ziele und Ergebnisse, die während der Laufzeit des BLK-Modellversuchs erarbeitet wurden

5.1 Phase I: Aufgabenstellung schriftlicher Prüfungen

Die Ergebnisse der ersten Phase wurden im ersten Zwischenbericht (Staatsinstitut 2004) protokolliert und im zweiten Modellversuchsflyer zusammengefasst (Staatsinstitut 2005b; S. auch Abschnitt I. 2 oben, ‚Empfehlungen zur Aufgabenstellung‘).

In dieser ersten Phase wurden knapp 100 authentische schriftliche Prüfungen von Vertretern von 15 Bundesländern detailliert gesichtet, um dann auf

Workshops gemeinsam intensiv besprochen zu werden. Ertrag dieser Arbeitsweise waren die Empfehlungen, die auf breiter Basis gefällt und an den im vorherigen Absatz genannten Stellen dokumentiert wurden.

Als zentral für das Evaluationskonzept zur Aufgabenstellung stellte sich bei Rezeptionsaufgaben vor allem heraus: eine Stufenangemessenheit und berufstypische, situative Einbettung; die Authentizität der verwendeten Hör- und Lesetexte; eine stufenadäquate Informationsdichte, Sprechgeschwindigkeit und Textlänge; und eine nicht aus dem Fachwissen allein lösbare Aufgabenstellung. Auch bei Produktionsaufgaben wurde die Bedeutung der Stufenangemessenheit festgestellt, sowie die Notwendigkeit einer eindeutigen Aufgabenstellung, die eine produktive Eigenleistung der Prüflinge ermöglichen soll. Als besonders wichtig wurde die situative Einbettung bei Mediationsaufgaben befunden. Auf Stufe II und III seien Sprachmittlungen vom Deutschen ins Englische vorzuziehen (vgl. Staatsinstitut 2005: 2).

Zur ersten Modellversuchsphase kann festgestellt werden, dass es durchaus gelungen ist, „das vorhandene Expertenwissen aufzugreifen, zu systematisieren und nachhaltig verfügbar zu machen“ (Staatsinstitut 2003: 2; vgl. Abschnitt IV. 4 oben). Ausdruck dieses Aufgreifens des eingesetzten Expertenpotentials war, dass die anvisierte Extrahierung „Best-practice-Beispiele“ aus den erhaltenen Daten über die Laufzeit der ersten Phase hinaus ging. Dies ist zurückzuführen auf den hohen qualitativen Erkenntnisgewinn der während der ersten Phase bezüglich geeigneter Aufgabenstellungen stattgefunden hat. Hinzu passt allerdings auch der wissenschaftliche Trend, dass neuerdings oft nicht mehr von „best“ sondern von „good practice“ die Rede ist (Saville 2003: 57, zitiert nach Hawkey 2004: 377; Vergl. auch die im Abschnitt IV. 2.4.3 oben erwähnten „Good Practice Guidelines der EALTA, 2006).

5.2 Phase II: Bewertung schriftlicher Prüfungen

Die Ergebnisse der zweiten Modellversuchsphase wurden im zweiten Zwischenbericht festgehalten (Staatsinstitut 2005a; S. auch die Dokumentation und Zusammenfassung im Abschnitt II. 2 oben).

In der zweiten Phase ging es um die gemeinsame Bewertung authentischer schriftlicher Prüfungsleistungen anhand verschiedener Auswertungsverfahren und Instrumente. Dabei wurden diese Verfahren selbst von den Teilnehmern bewertet, so dass eine empirische Untersuchung alternativer Vorgehensweisen dokumentiert werden konnten. Damit konnten bestehende und neuere Vorgehensweisen gemeinsam auf nationaler Ebene evaluiert werden.

Der zweite Modellversuchszwischenbericht berichtet von folgenden Arbeitsergebnissen der Phase II:

Deskriptoren:

Die Bewertung mit Hilfe von **zweispaltigen** Deskriptoren im Bereich **Produktion** wird empfohlen. [...]

Eine Bezeichnung der Spalten mit Begriffen wie „Inhalt“ und „Sprache“ wird nicht empfohlen, da es die derzeitigen Formulierungen ermöglichen, dass der Textzweck miterfasst werden kann. [...]

Eine grundsätzlich paritätische Gewichtung der Aspekte beider Spalten ist angemessen, denn das Ziel ist die Bewertung von Fachsprache.

Da keine strenge Trennung von Inhalt und Sprache vorliegt, sondern eine Mischung von Anforderungsbeschreibungen, kann eine Ausschlussklausel nicht empfohlen werden.

Die Bewertung mit Hilfe von **einspaltigen** Deskriptoren im Bereich **Mediation** erwies sich als ökonomisch und hilfreich.

Punkteschlüssel:

Die erprobte Punkteverteilung der jeweiligen Spalten der Deskriptoren wird empfohlen. [...]

Checkliste:

Sie sollte die inhaltliche Vollständigkeit der Arbeit (z. B. auch bei längeren Arbeiten auf Stufe III) leichter und auch schneller kontrollierbar machen (Staatsinstitut 2005a: 10).

Es kann auch zur zweiten Modellversuchsphase festgestellt werden, dass mehrere länderübergreifend besetzte Landesgruppen „vorhandene Erfahrungen und Ansätze reflektiert, gebündelt und systematisiert“ (Staatsinstitut 2003: 3; S. Abschnitt IV. 4, oben) und – wie bei Phase I – gemeinsam weitergeführt haben.

Die in der zweiten Phase erhobene Datenmenge wurde gemeinsam analysiert und wissenschaftlich ausgewertet, um die Korrelation zwischen angegebener Kompetenzstufe und tatsächlichem Leistungsniveau der Prüfungsarbeiten zu überprüfen. Ob die auf Grund unterschiedlich ausgeprägter Rücklaufquoten von unkorrigierten und korrigierten Prüfungsarbeiten und begrenzter Kapazität der wissenschaftlichen Begleitung zustande gekommene und analysierte Datenmenge ausreicht, um die Akzeptanz der KMK-Zertifikate im europäischen Raum maßgeblich gefördert zu haben lässt sich noch nicht abschließend feststellen (vgl. auch Abschnitt IV. 6 unten); die Grundlage einer effektiven Entwicklung und Umsetzung eines Evaluationskonzeptes bildete diese Datenmenge allemal.

5.3 Phase III: Aufgabenstellung, Durchführung und Bewertung mündlicher Prüfungen

Ein maßgebliches Zeichen des zunehmenden Erkenntnisgewinns während der Laufzeit des Modellversuchs war die Feststellung, dass die Bearbeitung der Aufgaben der dritten Phase mehr Zeit benötigt, als im Antrag vorgesehen, da in dem dritten Jahr, das den mündlichen Prüfungen gewidmet war, sowohl die Bereiche Aufgabenstellung und Bewertung als auch der der Prüfungsdurchführung zu berücksichtigen waren und man in den ersten beiden Phasen sehen konnte, dass es in diesen Phasen keineswegs an Gesprächsbedarf gefehlt hat. Die wegweisenden Ergebnisse der dritten Phase sind in nicht geringem Maße darauf zurück zu führen, dass trotz der damit organisatorischen Schwierigkeiten entgegen der ursprünglichen Planung der Ablauf der Workshops in der dritten Phase des Modellversuchs von der Lenkungsgruppe auf drei statt bisher zwei Tage ausgedehnt wurde und dies von den Ländern mitgetragen wurde.

Darüber hinaus stellte es sich in dieser dritten Phase bald heraus, dass die Herausforderungen im Bereich der Leistungsmessung mündlicher Fertigkeiten bei den KMK-Fremdsprachenzertifikatsprüfungen tatsächlich anderer Natur sind, als die der Beurteilung schriftlicher Leistungen (vgl. hierzu Abschnitt III. 1.6 oben, z. B. die Ausführungen von Luchtenberg (2002) und Ghisla und Kolb (2001)).

Ergebnisse der Phase III werden im obigen Abschnitt III. 2 von Götz Reuter erläutert. Hierzu gehören bei der Frage der Aufgabenstellung Empfehlungen und Gedanken zum Gesprächsablauf, zur Berufsrelevanz, zur Praktikabilität, zum Auf-

gabentyp und zur Eignung für die Stufe (siehe vor allem IV.2.1.2.5 **Überblick**, oben).

Beim Gesprächsablauf wurde unter anderem empfohlen: je höher die Stufe, desto geringer die Lenkung; Handlungsspielräume auf allen Stufen bieten; situativen Rahmen immer vorgeben; Reserve bereithalten.

Berufsrelevanz und Authentizität sollte in allen Aufgaben weitgehend gegeben sein. Bei der Praktikabilität wurde unter anderem betont, dass Realgegenstände bzw. situatives Ambiente geliefert werden sollten.

Zum Thema Aufgabentyp und Struktur wurde man sich unter anderem darüber einig, dass die Anteile kognitiver und kommunikativer Leistungen zu beachten sind und dass es zu ausgewogenen Redeanteilen aller Prüflinge kommt. Da sich die Stufenrelevanz der Aufgaben nach wie vor als schwierig erweist, wurde die Bedeutung einer Abstimmung unter den Aufgabenerstellern betont.

Das andere wesentliche Merkmal der dritten Modellversuchsphase war die konzentrierte Auseinandersetzung mit verschiedenen Bewertungsinstrumenten, die gerade wegen der Intensität der Beiträge als besonders produktiv gekennzeichnet werden darf. Dabei erwies es sich als unmöglich, trotz der Verlängerung der vier Workshops der dritten Phase von je zwei auf drei Tage zu einer Empfehlung für einen bestimmten Deskriptorensatz zu gelangen. Auf Grund dessen wurde am Ende des letzten Workshops des Modellversuchs die Bitte an die Lenkungsgruppe gerichtet, die Erarbeitung eines auf der Basis der Ergebnisse der Workshops weiterentwickelten Deskriptorensatzes in Auftrag zu geben. Dieser neue Deskriptorensatz wird im Anhang dieser Handreichung abgedruckt.

Schließlich dokumentiert Götz Reuter im Abschnitt III. 2 oben, in welchen Bereichen der Modalitäten mündlicher Prüfungen noch Redebedarf besteht. Bedenkt man, dass eine auf mündliche Fertigkeiten bezogene, wissenschaftlich fundierte Prüfungskultur verglichen mit den fest etablierten Traditionen der Bewertung schriftlicher Leistungen in Deutschland im großen und ganzen erst im Entstehen ist, sind die Ergebnisse der dritten Modellversuchsphase, zu denen auch die Erarbeitung der im Anhang abgedruckten best-practice-Beispiele zählen, allerdings alles andere als gering.

6 Zukünftige Entwicklungsmöglichkeiten für die KMK-Fremdsprachenzertifikatsprüfungen

Nachdem wir in den beiden letzten Abschnitten gesehen haben, was im Modellversuchsantrag geplant war und sich hat realisieren lassen, gilt es abschließend zu beraten, welche Entwicklungsmöglichkeiten die KMK-Zertifikatsprüfungen haben.

Sieht man sich noch einmal die Ziele des Modellversuchs an, wie diese im Abschnitt IV. 4 oben wieder zitiert wurden, entdeckt man wieder folgende Zielvorgabe:

Das Modellvorhaben stärkt durch eine transparente Orientierung an Standards die Wertigkeit der KMK-Fremdsprachenzertifikate und fördert damit deren Akzeptanz im europäischen Raum.
(Staatsinstitut 2003: 2-3)

Wie wir im Abschnitt IV.3.1 oben ausführlich gesehen haben, sprechen die Bildungsstandards und die DESI-Studie von einer „Orientierung“ an den Gemeinsamen europäischen Referenzrahmen. Wie wir sehen konnten, ist die Arbeit des DESI-Konsortiums unter anderem deswegen interessant, weil im Bereich Spre-

chen eine Normierung eines bereits vorhandenen Tests an den Niveaus des Referenzrahmens vorgenommen wurde. Ist diese Vorgehensweise wegweisend für die Zukunft der KMK-Fremdsprachenzertifikatsprüfungen?

Im Abschnitt III.1.6 oben habe ich davon gesprochen, wie der *Reference Supplement* zum *Manual for Relating Language Examinations to the Common European Framework* von “empirical validation” spricht als

the collection and analysis of test data and ratings from assessments to provide evidence that both the examination itself and the linking to the CEFR are sound“ (Takala (ed.) 2004: 1).

Ferner habe ich an jener Stelle eine Vorgehensweise vorgeschlagen, die quantitative und qualitative Verfahren verbindet bei dem Versuch, die KMK-Fremdsprachenzertifikatsprüfungen noch genauer und enger mit dem Referenzrahmen zu verzahnen. Im Modellversuchsantrag hieß es, die Akzeptanz der KMK-Fremdsprachenzertifikatsprüfungen im europäischen Raum zu fördern. Im Laufe des Modellversuchs bemühten sich die neuen Veröffentlichungen des Europarates immer mehr um Hilfestellungen bei einer Verzahnung von bereits existierenden Prüfungen mit dem Referenzrahmen. Auf Grund dieser Entwicklungen wurde während unserer Projektarbeit immer deutlicher, dass die Förderung der Akzeptanz der Fremdsprachenzertifikatsprüfungen nun das Unternehmen einer solchen Verzahnung bedeuten muss.

Die Frage ist, wer eine solche Verzahnung unternehmen soll und kann. Vieles ist mit dem BLK-Modellversuch EU-KonZert in dieser Richtung bereits geschehen. Deskriptoren, die sich an die des Referenzrahmens beziehen und für den berufsbezogenen Bereich weiterentwickelt wurden, wurden vor dem Hintergrund des Referenzrahmens noch einmal kritisch unter die Lupe genommen; Aufgabenstellungen wurden auf ihren Berufsbezug und ihre Nähe zu den Vorgaben des Referenzrahmens hin untersucht und best-practice-Beispiele erstellt; alternative Bewertungsinstrumente wurden an authentischen Prüfungsleistungen vergleichend auf ihre Eignung hin detailliert untersucht.

Doch vieles bliebe jetzt noch zu tun, wenn das KMK-Fremdsprachenzertifikat ein Produkt anbieten möchte, das sich mit den Produkten der großen Testanbieter, die gegenwärtig bei dem Europarat den größten Einfluss haben, wie etwas UCLES / Cambridge Assessment, vergleichen ließe. Noch wurden beim KMK-Fremdsprachenzertifikat Vorgehensweisen wie die Vorerprobung von Items oder die Bildung von Aufgabenbanken nicht realisiert; noch blieb die statistische Auswertung der mittlerweile mehr als 25.000 Prüfungen im Jahr unterentwickelt. Dies hat natürlich mit dem deutschen Föderalismus zu tun, ist zweifelsohne auch eine Frage der Ressourcen. Aber wenn an das Ziel festgehalten wird, die Akzeptanz der KMK-Fremdsprachenprüfungen im europäischen Raum zu fördern, wird man nicht darum herum kommen können, sich nach der Relevanz dieser hier unerprobten oder unterentwickelten, doch in der Testentwicklung international gängigen Vorgehensweisen zu fragen.

Bleibt immer noch die Frage, wer eine solche Ausbreitung der test-theoretischen Basis der KMK-Fremdsprachenzertifikatsprüfungen unternehmen sollte und könnte.

Eine Möglichkeit ist die Vergabe eines solchen Projektes an ein oder mehrere Landesinstitute oder Behörden.

Traditionell besteht auch die Möglichkeit eines Auftrages an eine Universitätsprofessur, zu deren Dienstverpflichtungen die aktive Forschung gehört. Ein be-

kannter wissenschaftlicher Mehrwert dieser Variante ist, dass etwaige Assistentinnen und Assistenten ihre Tätigkeiten im Drittmittelprojekt als Grundlage einer Promotion benutzen können.

Wie oben wiederholt bemerkt, sind hier die Erfahrungen des DESI-Konsortiums bei der Normierung eines bereits vorhandenen Tests an den Niveaus des Referenzrahmens sicherlich von Interesse.

Von großem Interesse ist in diesem Zusammenhang ferner die Arbeit des 2004 von der Kultusministerkonferenz und der Humboldt-Universität gegründeten Instituts zur Qualitätsentwicklung im Bildungswesen (IQB), dessen wesentliches Ziel es ist, „die nationalen Bildungsstandards zu normieren, zu präzisieren, und weiterzuentwickeln sowie deren Implementierung in den Schulen wissenschaftlich zu begleiten“ (Institut ohne Jahr: 4). Leiter des Instituts ist der beurlaubte Nürnberger Professor für Pädagogische Psychologie Olaf Köller. Für die Koordinierung der Aufgabenentwicklung im Fach Englisch (Sekundarstufe I) verantwortlich ist eine abgeordnete Realschullehrerin. Ein studierter Anglist und promovierter Testtheoretiker arbeitet beim IQB gegenwärtig über statistische Modellen mit latenten Variablen. Zusätzlich arbeitet dort im Bereich Normierung der Bildungsstandards im Fach Englisch auch noch ein Doktorand. Damit bildet sich ein Team an dem Institut, der sinnvoll ergänzt werden könnte durch die Abordnung einer im Bereich der beruflichen Schulen tätige Lehrkraft oder eines einschlägig ausgewiesenen Wissenschaftlers, um an der Weiterentwicklung der KMK-Fremdsprachenzertifikate zu arbeiten.

Schließlich bietet sich die Möglichkeit an, sich bei einer Weiterentwicklung der Zertifikatsprüfungen und einer anvisierten Stärkung der europäischen Akzeptanz dieser Prüfungen von einem der „experts“ der EALTA beraten zu lassen. Diese „experts“ müssen viel strengere Kriterien erfüllen als die „individual members“:

Qualifications include:

- a doctorate in language testing or a related field
- at least four years of teaching experience in language testing or a related field
- publications and conference presentations on language testing and assessment
- consultancies to organisations, institutions or projects
- documented participation in test design, development and validation of significant tests or public examinations
- documented participation in relevant research
- references from at least two other Expert Members of EALTA (EALTA ohne Jahr)

Ob und, falls ja, welche Möglichkeit für eine wissenschaftliche Weiterentwicklung der KMK-Fremdsprachenzertifikate gewählt wird, das Fazit von Charles Alderson auf der Abschlusstagung des Modellversuchs in Kloster Banz am 25. Oktober 2006 wird wohl auf alle Fälle stimmen: "Good tests cost time and money".

Wofür und für wen sollen diese Investitionen vorgenommen werden?

Wir haben bei IV.1 oben gesehen, wie in einer Veröffentlichung über einen frühen Modellversuch zu Fremdsprachen an der Berufsschule bereits 1992 neben der pädagogisch zentralen Überlegung der Schülermotivation die für die beruflichen Schulen charakteristische Frage der Heterogenität der Lerngruppen sowie die Erwartungen der Arbeitgeber bezüglich aussagekräftiger, berufsrelevanter Informationen gestellt wurden (Baumgartner / Gibson 1992: 6). Wenn man mitver-

folgt, wie manche Länder eine KMK-Fremdsprachenzertifizierung auf einer höheren Ebene, die C1 des Referenzrahmens entsprechen würde (was faktisch die Einführung einer KMK-Ebene IV bedeuten würde) oder heute bereits das freiwillige Ablegen der KMK-Fremdsprachenzertifikatsprüfungen durch Besucher der Fachoberschule zulassen, dann sieht man, dass der Natur des föderalistischen Bildungswesen entsprechend vieles im Fluss ist.

1996 sprach der Staatssekretär des Ministeriums für Schule und Weiterbildung des Landes Nordrhein-Westfalen anlässlich einer Fachtagung zum Thema Englisch in der Berufsschule folgende sehr interessanten Worte:

Hinsichtlich der Standardisierung der verschiedenen Niveaustufen streben wir gemeinsam mit den anderen Bundesländern und mit den Kooperationspartnern aus Wirtschaft und Handwerk eine Einbettung in ein nationales oder besser noch in ein europäisches Zertifizierungssystem, das ein vom jeweiligen Schulabschluss unabhängiges Zertifikat ermöglicht
(Besch 1997: 13).

Die Heterogenität der Schüler der beruflichen Schulen stellt ohne Zweifel eine der entscheidenden Herausforderungen dieser Schulformen dar. Gleichzeitig jedoch bietet unter Umständen gerade diese Heterogenität die Chance, Unterrichts- und Zertifizierungsangebote weiter auszubauen, die auf die individuellen Bedürfnisse der Unterrichtsteilnehmer aufbauen. Die beruflichen Schulen haben mit der Entwicklung des KMK-Fremdsprachenzertifikats die Heterogenität ihrer Schüler als Chance begriffen, während die allgemeinbildenden Schulen Deutschlands gegenwärtig Stufen des Referenzrahmens zunehmend als allgemein geltend vorschreiben.

Welche Ressourcen die berufsbildenden Schulen für die Weiterverfolgung dieses pädagogisch sehr innovativen Ansatzes benötigen und wie organisatorisch die Bewahrung eines breiten Prüfangebots mit einer Unterrichtsgestaltung, die Rücksicht nimmt auf die berühmten Rückkopplungseffekte von Prüfungen auf Unterricht berücksichtigt werden kann sind freilich neben der Förderung der europäischen Akzeptanz der KMK-Fremdsprachenzertifikatsprüfungen Fragen, die bei der Erörterung der zukünftigen Weiterentwicklungsmöglichkeiten dieser Prüfungen mitberücksichtigt werden sollen. Denn wie der Bundespräsident in diesen Tagen seiner Nation zu bedenken gegeben hat: „Gute Bildung [...] bleibt für den Einzelnen [...] die wichtigste Voraussetzung für [...] berufliches Fortkommen“ (Köhler 2006: 3).