

CSP **Center scientifico da cumpetenza per la plurilinguitad** Cogniziun Società Formation Bildung Migration Furmazion Gesellschaft  
CSP **Centro scientifico di competenza per il plurilinguismo** Scuola Arbeit Politique Communitad School Travail Ecole Community  
CSP **Centre scientifique de compétence sur le plurilinguisme** Migrazione Furmazion Societad Cognition Society scola Migration  
KFM **Wissenschaftliches Kompetenzzentrum für Mehrsprachigkeit** Societé Cognizione Migraziun Schule Communauté Kognition  
RCM **Research Centre on Multilingualism** Formazione Lavoro Politics Comunità Work Politik Lavur Politica Formation Gemeinschaft

# Sprachliche Anforderungen und Sprachtests im Migrations- und Integrationsbereich

Eine Zusammenfassung der akademischen Literatur

Evelyne Pochon-Berger  
Peter Lenz

2014

Bericht des Wissenschaftlichen Kompetenzzentrums für Mehrsprachigkeit

**Herausgeberin**

Institut für Mehrsprachigkeit  
[www.institut-mehrsprachigkeit.ch](http://www.institut-mehrsprachigkeit.ch)

**AutorInnen**

Evelyne Pochon-Berger  
Peter Lenz

**Übersetzung**

Katharina Karges

Das vorliegende Projekt wurde im Rahmen des Arbeitsprogramms 2012–2014 des Wissenschaftlichen Kompetenzzentrums für Mehrsprachigkeit durchgeführt und von der Schweizerischen Eidgenossenschaft finanziell unterstützt. Für den Inhalt dieser Veröffentlichung sind die AutorInnen verantwortlich.

**Freiburg, März 2014**

**Layout**

Billy Ben, Graphic Design Studio

# **Sprachliche Anforderungen und Sprachtests im Migrations- und Integrationsbereich**

---

Eine Zusammenfassung der akademischen Literatur

---

**Evelyne Pochon-Berger**  
**Peter Lenz**

2014

---

Bericht des Wissenschaftlichen Kompetenzzentrums für Mehrsprachigkeit

# Abstract

Das vorliegende Dokument ist das Ergebnis einer literaturbasierten Studie der neueren Entwicklungen im Bereich der sprachlichen Anforderungen und Sprachtests für Migrantinnen und Migranten. Die Studie war Teil des Forschungsprogramms 2011-2014 am wissenschaftlichen Kompetenzzentrum für Mehrsprachigkeit (KFM)<sup>1</sup>.

Die Einführung sprachlicher Anforderungen und die Verwendung formeller Sprachtests als Voraussetzungen für die Erteilung von Einreisevisa und Niederlassungsbewilligungen und/oder für Einbürgerungen sind eine relativ neue, aber schnell um sich greifende Praxis in vielen westlichen Ländern, insbesondere in Europa. Diese Anforderungen und die dazugehörigen Beurteilungsinstrumente stehen neu im Zentrum akademischer Debatten und Forschungsprojekte, und sind das Thema mehrerer wissenschaftlicher Publikationen. Der vorliegende Bericht versucht, einen strukturierten Überblick über verschiedene Aspekte dieser sprachlichen Anforderungen und Tests zu liefern, wobei ein Korpus akademischer Publikationen die Grundlage unserer Studie bildet.

Die Zusammenfassung der Literatur ist in zwei Abschnitte gegliedert. Der erste gibt einen Überblick über die neueren Entwicklungen bezüglich der sprachlichen Anforderungen für die Einwanderung, Niederlassung und Einbürgerung, vor allem im letzten Jahrzehnt; er beschreibt auch die Einführung neuer formeller Beurteilungsinstrumente auf verschiedenen Stufen des Immigrationsprozesses. Da die Autorinnen und Autoren, die die neuen Entwicklungen kontrovers beurteilen, führt der erste Abschnitt auch eine Reihe von umstrittenen Themen ein, in erster Linie aus einer kritischen Perspektive. Der zweite Abschnitt beschäftigt sich systematisch mit den Themen der laufenden Debatte über Sprachtests im Migrations- und Integrationsbereich, dies unter Zuhilfenahme von zentralen Kategorien aus der Testvalidierungsforschung. Dabei bildet das *Assessment Use Argument* (AUA) von Bachman & Palmer (2010) den konzeptuellen Rahmen.

Den Abschluss unserer Studie bilden Empfeh-

lungen für die zukünftige Testforschung, Publikationen und Testvalidierungspraktiken.

Insgesamt bestätigen unsere Erkenntnisse die subjektiv wahrgenommene Zunahme von sprachlichen Anforderungen und formellen Sprachtests (sowie gelegentlich kulturellen Wissenstests) im Bereich der Einwanderung und Integration in einigen „westlichen“ Ländern, insbesondere in Europa und während der letzten zehn Jahre. Obwohl ein allgemeiner Trend ausgemacht werden kann, haben wir auch erhebliche Unterschiede bezüglich der genauen Festlegung der Anforderungen und der damit verbundenen Beurteilungen festgestellt, z.B. was das erforderliche Sprachkompetenzniveau angeht.

Die Analyse der Literatur auf der Basis von Bachman und Palmers Validierungsmodell zeigt, dass die einbezogenen Publikationen die verschiedenen Themen unterschiedlich behandeln. Insbesondere zielen nur sehr wenige Texte darauf ab, die Validität eines spezifischen Tests und seiner Verwendung systematisch darzulegen (oder infrage zu stellen). Die meisten Publikationen beschäftigen sich mit der *Aussagekraft* der Testinhalte und Niveaus in Bezug auf das Wissen und die Fähigkeiten, welche zur Beurteilung ausgewählt wurden (üblicherweise „ausreichende Kenntnisse, um sich in die Gesellschaft zu integrieren“), sowie mit der Frage nach den *positiven Auswirkungen* des Tests und der dafür vorgesehen Vorbereitungsphase. Einige Publikationen stellen das Testen der Migrationsbevölkerung *an sich* so grundsätzlich infrage, dass ihre Thematik durch existierende Testvalidierungsmodelle kaum erfasst wird.

In unserer Schlussfolgerung formulieren wir Empfehlungen für zukünftige Arbeiten in Bezug auf die folgenden Punkte: das Einbeziehen von akademischen Disziplinen, welche auf gesellschaftliche, ethische und juristische Themen im Migrations- und Integrationsbereich spezialisiert sind, die Verpflichtung, Verantwortung für den eigenen Test zu übernehmen und diesen mithilfe eines aktuellen Validierungsmodells systematisch zu validieren, die Verwendung von Bedürfnisanalysen als Grundlage für das Testdesign und dessen Verwendung, die Notwendigkeit weiterer Wirkungsstudien zu existierenden Testsystemen und die Wünschbarkeit von mehr Transparenz bezüglich der Testvalidierung und ihrer Ergebnisse.

1 | Die Autoren danken Séverine Beaud Duarte Rodrigues für ihren grossen Beitrag zur Sammlung und Auswahl der akademischen Publikationen zu Beginn des Projektes. Auch die wertvollen Beiträge unseres internen Gutachters, Thomas Studer, zu verschiedenen Zeitpunkten während der Durchführung der Studie waren sehr willkommen.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>4</b>
1.1	Der Schweizer Kontext	4
1.2	Ziel und Struktur des vorliegenden Berichts	4
1.3	Quellen	5
<b>2</b>	<b>Sprachprüfungen im Migrations- und Integrationsbereich in einem sich wandelnden Kontext</b>	<b>7</b>
2.1	Sprachenpolitik und Beurteilungsverfahren im Wandel	7
2.2	Die Umsetzung strengerer Immigrationspolitik im sprachlichen Bereich	8
2.2.1	Die Ausweitung sprachlicher Anforderungen an Migrantinnen und Migranten	9
2.2.2	Zunehmende Verwendung standardisierter Sprachtests	10
2.2.3	Verlangte Sprachkompetenzniveaus	10
2.2.4	Sprachkompetenztests und Landeskundetests	11
<b>3</b>	<b>Validitätsaspekte von Sprachprüfungen im Migrations- und Integrationsbereich</b>	<b>12</b>
3.1	Bachman und Palmers argumentationsbasierter Ansatz zur Validierung von Beurteilungen	12
3.2	Gerechtigkeit und Sprachtests im Migrations- und Integrationsbereich	15
3.3	Positive Auswirkungen des Testens	19
3.4	Wertesensibilität der Entscheidungen	23
3.5	Unparteilichkeit der Entscheidungen	24
3.6	Aussagekraft der Interpretationen bezüglich eines zugrundeliegenden Konstrukts	25
3.7	Unparteilichkeit der Interpretation gegenüber allen Kandidatengruppen	30
3.8	Konsistenz der Prüfungsergebnisse	31
<b>4</b>	<b>Zusammenfassung und Schluss</b>	<b>33</b>
<b>5</b>	<b>Bibliographie</b>	<b>38</b>

# 1

## Einleitung

### 1.1 Der Schweizer Kontext

In vielen Ländern der westlichen Welt, insbesondere in Europa, wird der Aufenthaltsstatus von Staatsangehörigen anderer Länder mehr und mehr von Sprachtests abhängig gemacht. Zwar ist die Schweiz kein Vorreiter in dieser Entwicklung, doch zeigt sich dieser Trend auch hier in legislativen und administrativen Massnahmen auf allen Ebenen des politischen Systems. Auf Bundesebene wird Sprache in erster Linie mit Integration in Verbindung gebracht. So lancierte der Bund 2010 unter dem Motto „fördern und fordern“ einen Integrationsplan, welcher die verstärkte Sprachförderung für alle Migrantinnen und Migranten vorsieht, die keine der Nationalsprachen der Schweiz sprechen (Gerber & Schleiss, 2013). In einer ersten Etappe wurde dafür eine Reihe von Instrumenten entwickelt, die das Lernen und Lehren der nationalen Sprachen fördern sollen. Dazu gehören ein Rahmencurriculum für niederschwellige Sprachkurse (Lenz, Andrey & Lindt-Bangerter, 2009), bedürfnisorientierte Spezifikationen von Lernzielen, die auf Sprachgebrauchsszenarien basieren (Müller & Wertenschlag, 2013) sowie Materialien für bedürfnisorientierte Niveaufeststellungen, Kursplanung, Unterricht und (Selbst) Beurteilung. Ausserdem wurde ein Workshop-Programm entwickelt, welches die Verbreitung der neuen Instrumente unterstützen soll.

Das Ausländergesetz erlaubt seit 2008, die Erteilung der Niederlassungsbewilligung im Rahmen einer sogenannten Integrationsvereinbarung von sprachlichen Voraussetzungen abhängig zu machen (normalerweise vom Besuch eines Sprachkurses) (Skenderovic, 2013). Zwar existieren im Moment noch keine spezifischen Instrumente (darunter Tests) für die Beurteilung<sup>2</sup> der sprachlichen Kompetenzen von Migrantinnen und Migranten, welche eine Niederlassungsbewilli-

2 | Der Begriff „Beurteilung“ umfasst mehr als Sprachtests. Zwar beinhaltet es auch diese, kann aber auch andere Evaluationsformen bezeichnen, welche den Wissens- oder Könnensstand von Lernenden feststellen können. Dort wo die Beurteilung der sprachlichen Kompetenzen die Form eines Tests hat, werden die beiden Begriffe in diesem Dokument synonym verwendet.

gung<sup>3</sup> anstreben. Allerdings hat das Verb „fordern“ im offiziellen Motto der schweizerischen Integrationspolitik in der Periode, in welcher die vorliegende Studie entstand und in welcher auch das Ausländer- und Integrationsgesetz den politischen Prozess durchlief (2013), zunehmend an Bedeutung gewonnen. Es ist davon auszugehen, dass ab 2015 bestimmte Gruppen der Migrationsbevölkerung Kenntnisse einer nationalen Sprache nachweisen müssen, indem sie entweder einen offiziell anerkannten Sprachtest bestehen oder im Rahmen von Sprachkursen behördlich geregelte Nachweise erbringen.

### 1.2 Ziel und Struktur des vorliegenden Berichts

Das Hauptanliegen der vorliegenden Studie ist die Erstellung einer strukturierten Übersicht über verschiedene Aspekte von Sprachkompetenzbeurteilungen im Migrations- und Integrationsbereich auf der Basis von wissenschaftlichen Publikationen. Die untersuchten Aspekte sind

- die neuesten Entwicklungen bezüglich sprachlicher Anforderungen und formeller Sprachtests für Migrantinnen und Migranten,
- die zentralen Kritikpunkte und Argumente in den meist kritischen Diskursen, die sich mit diesen Entwicklungen befassen,
- die Validität der Beurteilung von Sprachkompetenzen und der Sprachtests im Migrations- und Integrationsbereich.

Aufgrund dieses Überblicks hoffen wir, nicht nur aktuelle Themen und Trends herauszuarbeiten, sondern auch auf Schwächen in der laufenden wissenschaftlichen Forschung und Debatte hinzuweisen. Unsere Erkenntnisse bilden dann die Basis für Empfehlungen bezüglich der zukünftigen konzeptuellen Arbeit, Forschung und Entwicklung, die in diesem sensiblen Gebiet zu leisten wären.

3 | Die Verleihung der Staatsbürgerschaft ist in der Schweiz anders geregelt als jene der Niederlassung; die Gemeinden haben (noch) das Recht, sie zu verleihen und die sprachlichen Anforderungen und Beurteilungsverfahren selber festzulegen. Die angewandten Praktiken variieren erheblich von Gemeinde zu Gemeinde.

Der vorliegende Bericht hat folgende Struktur: Kapitel 1 stellt den Hintergrund der Studie dar, darunter den lokalen Kontext, Ziele und Struktur des Berichts, sowie die vorhandenen Quellen. Kapitel 2 informiert über die Entwicklungen des letzten Jahrzehnts, die mit sprachlichen Anforderungen an Migrantinnen und Migranten zu tun haben; es beschreibt auch die Einführung von formellen Sprachtests auf den verschiedenen Ebenen des Immigrationsprozesses (Einwanderung, Niederlassung, Einbürgerung). Da die darin enthaltenen Informationen aus Publikationen stammen, deren Verfasserinnen oder Verfasser überwiegend gegen die Verschärfung der Einwanderungsbedingungen sind, wird in Kapitel 2 auch eine Übersicht der verschiedenen geäußerten Kritikpunkte gegeben. Kapitel 3 beginnt mit der Vorstellung eines mehrheitlich anerkannten Modells für die Validierung von Tests. Die wichtigsten Kategorien dieses Modells werden dann genutzt, um die laufende Diskussion über die Evaluation der Sprachkompetenzen von Migrantinnen und Migranten darzustellen. Kapitel 4 schliesslich fasst die Erkenntnisse der vorherigen Abschnitte zusammen, zieht Bilanz und macht Vorschläge für zukünftige Arbeiten.

### 1.3 Quellen

Einreisevisa, Niederlassungsbewilligungen und Einbürgerung von sprachlichen Kompetenzen und formellen Sprachtests abhängig zu machen, ist eine relativ neue Entwicklung, gewinnt aber zurzeit rasch an Bedeutung. Dementsprechend ist das Sprachentesten in diesem Bereich ein junges, aber wachsendes Gebiet in der wissenschaftlichen Forschung und Debatte, das die Aufmerksamkeit einer ganzen Reihe von Wissenschaftlerinnen und Wissenschaftlern auf sich gezogen hat. Die Studien, die bisher entstanden sind, sind in erster Linie kritische Beschreibungen und Analysen von sprachlichen Aspekten der Einwanderungspolitik und von Testverfahren in den verschiedenen Ländern. Dabei wurden vor allem Fragen der Testfairness und der sozialen Gerechtigkeit aufgeworfen. In den letzten zehn Jahren wurden mehrere Sam-

melbände zu diesen Themen herausgegeben, darunter Adami & Leclercq (2012), Hogan-Brun et al. (2009a), Mar-Molinero & Stevenson (2006), Oers et al. (2010), Slade & Mollering (2010). Ebenso haben einige Zeitschriften diesem Thema Sondernummern gewidmet, z.B. das *International Journal on Multicultural Studies*, Vol. 10 (1) 2008 oder *Language Assessment Quarterly*, Vol. 6 (1), 2009. Ein Handbuchartikel von Kunnan (2012) gibt eine aktuelle Übersicht zum Thema. Die meisten zurzeit erhältlichen Publikationen sind zwar noch hauptsächlich programmatisch, sie haben aber dennoch geholfen, das Forschungsfeld zu begründen und die Aufmerksamkeit von Vertreterinnen und Vertretern der Angewandten Linguistik und des Sprachentestens auf diese neuen Entwicklungen und die damit verbundenen Probleme zu lenken. Bisher wurden nur wenige empirische Arbeiten veröffentlicht, und kaum eine Studie beschäftigt sich mit der Entwicklung und Validierung der eigentlichen Sprachprüfungen. Lediglich die Artikel von De Jong, Lennig, Kerkhoff, & Poelmans (2009), Perlmann-Balme (2011) und Plassmann (2011) bilden hier Ausnahmen. Dabei hat die Publikation von De Jong et al. einen stark forschungsorientierten Fokus, während die anderen beiden Artikel auch einem fachfremden Publikum zugänglich sind. Alle drei Veröffentlichungen haben gemeinsam, dass sie konfirmatorisch angelegt sind und die Sicht und Interessen der Testentwicklerinnen entwickler und vertreten, anstatt diese kritisch zu hinterfragen.

Da Sprachprüfungen im Migrations- und Integrationsbereich für Regierungen und überstaatliche Institutionen (z.B. den Europarat) von beträchtlichem Interesse sind und das Thema auch im öffentlichen Diskurs und der Politik vertreten ist, überrascht es nicht, dass einige Berichte zum Thema von offiziellen Stellen in Auftrag gegeben wurden (z.B. Schneider et al., 2006: Empfehlungen für den Nachweis sprachlicher Kommunikationsfähigkeit im Hinblick auf die Einbürgerung, im Namen der Eidgenössischen Kommission für Migrationsfragen; Balch et al. (2008) und Extramania & Van Avermaet, 2010: Ergebnisse zweier Gutachten des Europarates; Little, 2010: ein Bericht über die sprachliche Integration von erwachsenen Migrantinnen und Migranten, ebenfalls im Auftrag des

Europarates). In unserer Analyse werden wir vor allem wissenschaftliche Artikel und Forschungsberichte aus Zeitschriften oder Büchern rezipieren. Nur eine kleine Auswahl an behördlich in Auftrag gegebenen Berichten, wie die oben erwähnten, soll bei Bedarf konsultiert werden.

Die vorhandene Literatur ist in ihrer geographischen Verteilung eher beschränkt. Sie betrifft vor allem die Politik und Praxis in europäischen Ländern (EU- und Nicht-EU-Mitglieder) sowie in wenigen anderen Staaten, vor allem in den USA, Kanada und Australien. Die besprochenen Publikationen kommen unter anderem deshalb aus diesen Ländern, weil sie stark von Zuwanderung betroffen sind. Während einige dieser Staaten bereits eine lange Zuwanderungstradition kennen (z.B. die USA, Kanada und einige westeuropäische Länder), ist es für andere eine neue Erfahrung (vor allem für südeuropäische Länder wie Spanien und andere westliche europäische Länder) (Extra et al. 2009). Van Avermaet (2009) zufolge lassen sich drastische Veränderungen in der Migrationspolitik auf stark erhöhte Zuwanderungsraten zurückführen. Dies gilt v.a. für europäische Länder, da hier Immigration *nach* Europa zu starker Mobilität *innerhalb* Europas hinzukommt. Der europäische Kontext ist somit ein fruchtbares Forschungsfeld, wobei die grosse Spannweite der hiesigen Praktiken und politischen Programme, trotz einiger Gemeinsamkeiten, das Interesse der Forschenden zusätzlich geweckt hat. Es existieren Studien über die Pioniere im Bereich von Sprachprüfungen für Migrantinnen und Migranten, darunter die Niederlande, Deutschland, Dänemark und Grossbritannien (siehe z.B. Blackledge, 2009a und b; Extra & Spotti, 2009a und b; Laversuch, 2008; Michalowski, 2010; Möllering, 2010; Slade, 2010a). Praktiken in traditionellen Einwanderungsländern ausserhalb Europas, besonders den USA, Kanada und Australien, wurden ebenfalls umfassend beschrieben (siehe z.B. Cox, 2010; Farrell, 2010; Holland, 2010; Hargreaves, 2010; McNamara, 2009a). Informationen über Länder wie Österreich, Belgien, Luxemburg, Spanien und die Baltischen Staaten sind dagegen nur spärlich vorhanden, während die Situation in weiteren Ländern nahezu unerforscht ist (z.B. Griechenland, Italien, Polen oder Portugal). Schliesslich wissen wir fast nichts über die

Praktiken in Ländern, die nicht zur westlichen Welt gezählt werden (z.B. asiatische oder afrikanische Staaten, vgl. aber Lee, 2010 zum Fall Korea).

Im vorliegenden Bericht werden wir das Hauptaugenmerk auf Veröffentlichungen über europäische Länder legen.

# 2

## Sprachprüfungen im Migrations- und Integrationsbereich in einem sich wandelnden Kontext

### 2.1 Sprachenpolitik und Beurteilungsverfahren im Wandel

Sprachtests und Sprachkompetenzfeststellung im Migrations- und Integrationsbereich waren in einer Reihe von Ländern Thema hitziger Debatten, welche daraufhin zu Änderungen in den nationalen Gesetzgebungen führten. Die folgenden Publikationen geben dazu einen Überblick: Blackledge (2009a & b) für Grossbritannien, Slade (2010a) für die Niederlande, Möllering (2010) für Deutschland, Farrell (2010) für Australien. Auf supranationaler Ebene formulieren Balch et al. (2008) im Namen der ALTE Empfehlungen für die Verwendung von Sprachtests im Kontext der Einbürgerung, während Extramania & Van Avermaet (2010) eine Zusammenfassung zur Sprachenpolitik im Migrationsbereich in mehreren europäischen Ländern bieten. Ein dritter Bericht, der vom Europarat in Auftrag gegeben wurde und 2014 erscheinen soll, wird auf einer gross angelegten Umfrage zur Politik der Mitgliedsstaaten im Bereich der sprachlichen Integration von Migrantinnen und Migranten basieren. Diese fand 2013 statt.

Einige Studien lassen für die letzten zehn Jahre erhebliche Veränderungen in der nationalen Einwanderungspolitik vieler Länder erkennen (vor allem in Europa, aber auch in Australien, Kanada und den USA). Diese Veränderungen zeigen sich durch eine „erhöhte Anzahl und Reichweite regulatorischer Mechanismen“<sup>4</sup> (Saville, 2009). Mit anderen Worten: Es gibt einen allgemeinen Trend,

strengere Immigrations- und Integrationsmassnahmen zu implementieren (Hogan-Brun et al., 2009b; Van Avermaet, 2009; Slade, 2010a), was einen „assimilationistischen“ Ansatz im Umgang mit Migrantinnen und Migranten verrät (Möllering, 2010; Slade, 2010a). Einige Länder haben sogar verpflichtende Integrationsprogramme eingerichtet, deren Nichtbesuch rechtliche Nachteile mit sich bringt (Kostakopoulo, 2010). Laut Böcker & Strik (2011) steht der Trend hin zu einer strengeren Immigrationspolitik in vielen europäischen Ländern in krassem Gegensatz zur entsprechenden Politik dieser Länder in den 1980er und 1990er Jahren. Damals versuchte man eher, Immigration und Integration durch den einfachen Zugang zu Niederlassungsbewilligung und Staatsbürgerschaft zu fördern. Die Niederlande waren damals ein Beispiel für diese liberale Zuwanderungspolitik, sie waren aber auch das erste europäische Land, welches später strenge Einwanderungs- und Integrationsgesetze erliess (Slade, 2010a). Seit dieser Wende in der niederländischen Gesetzgebung wird die Erteilung der Niederlassungsbewilligung oder der Staatsbürgerschaft gewissermassen als „Belohnung“ für diejenigen verstanden, die eine Reihe von Hindernissen überwunden haben und daher einen Platz in der Gesellschaft verdienen (Böcker & Strik, 2011; Adamo, 2008; Kostakopoulo, 2010). Zuvor wurde dieser Prozess vor allem als Anreiz für die Integration verstanden, weniger als „Belohnung“.

Hogan-Brun et al. (2009b) verbinden die Veränderung der europäischen Sprachenpolitik im Bereich der Einwanderung mit dem Jahr 2004, in dem die ehemals kommunistischen Staaten der Europäischen Union beitraten, was zu massiven Migrationsbewegungen aus den osteuropäischen EU-Ländern in die westeuropäischen EU-Länder führte. Einige Autorinnen und Autoren erklären die verstärkte Einwanderung wiederum mit anderen ökonomischen und politischen Faktoren (z.B. Van Avermaet, 2009; Saville, 2009; Hargreaves, 2010; Slade, 2010a): Globalisierung, allgemeine Arbeitsmigration, Bildung, Tourismus und politische Verfolgung. Die daraus resultierende grössere kulturelle und sprachliche Diversität in den Aufnahmeländern wurde oft als Bedrohung der nationalen Identität wahrgenommen; sogar in Nationen wie Australien, die durch Immigration entstanden sind (Slade, 2010a; Wright, 2008). Hogan-Brun et al. (2009b) zufolge war dieser

4 | Sämtliche in diesem Bericht enthaltenen Zitate wurden ins Deutsche übersetzt. Dort, wo deutschsprachige Übersetzungen bereits vorlagen, wurden diese verwendet und entsprechend zitiert.

Effekt besonders auffällig in jenen EU-Staaten, in denen andauernde politische und wirtschaftliche Vereinheitlichungsprozesse alleine bereits zu einer nationalen Identitätskrise geführt hatten. Um diese nationale Identität und den gesellschaftlichen Zusammenhalt zu bewahren, reagierten viele Nationalstaaten daher mit strengeren Einwanderungsrichtlinien. Die zunehmende Fremdenfeindlichkeit, welche zu Beginn des 21. Jahrhunderts durch mehrere terroristische Attentate verschärft wurde, gilt als weiterer Faktor hin zu einer restriktiveren Migrationspolitik und der Einführung von Bestimmungen, die die „nationale Sicherheit“ verbessern sollen (Blackledge, 2009c; Laversuch, 2008; Van Avermaet, 2009; Wright, 2008).

## 2.2 Die Umsetzung strengerer Immigrationspolitik im sprachlichen Bereich

Die veränderte Sprachenpolitik in den verschiedenen Ländern führte zu einer erstaunlichen Vielfalt der sprachlichen Anforderungen an Migrantinnen und Migranten. Die Unterschiede betreffen vor allem Testmethoden, Testdesign und Testdurchführung, die getesteten Fertigkeiten, das verlangte Sprachkompetenzniveau, das Mindestalter der betroffenen Personen, Kosten, Ausnahmeregelungen für bestimmte Gruppen und den Status der vorbereitenden Sprach- und Integrationskurse (verpflichtend vs. freiwillig). Diese Vielfalt wird in einer Reihe jüngerer Studien zu den Testpraktiken der europäischen Länder hervorgehoben (siehe Van Avermaet, 2009; Böcker & Strik, 2011<sup>5</sup>; Extramania & Van Avermaet, 2010). So stellte Van Avermaet (2009) auf der Basis von Daten aus 17 Ländern fest, dass das verlangte

Sprachniveau zwischen A1 und B1 für den Aufenthalt, A1 und B2 für die Niederlassung und schliesslich zwischen A2 und B2 für die Einbürgerung variieren kann. Diese Daten offenbaren grosse Unterschiede bezüglich der sprachlichen Regelungen in den einzelnen Ländern. In einem Vergleich dreier englischsprachiger Länder (den USA, Kanada und Grossbritannien) stellte Hargreaves (2010) ebenfalls erhebliche Unterschiede zwischen den drei Staaten fest, sowohl bezüglich der Formate der Einbürgerungstests wie auch mit Blick auf ihren Inhalt. So nutzen Grossbritannien und Kanada einen Multiple Choice-Test (in Grossbritannien am Computer, in Kanada auf Papier), während in den USA ein informelles Interview durchgeführt wird. Gemäss einiger Autorinnen und Autoren (Hargreaves, 2010; Hogan-Brun et al., 2009b; McNamara, 2009) ist diese Vielfalt der Sprachprüfungen und ihrer Durchführung zum Teil darauf zurückzuführen, dass diese von nationalen Regierungen ausgewählt und eingeführt werden (anstatt von Expertinnen oder Experten). Die entsprechenden Entscheidungen werden aufgrund des aktuellen gesellschaftspolitischen Kontextes und der wahrgenommenen Einstellungen der Bevölkerung getroffen. Alle Studien belegen klar die rasante und relativ unkoordinierte Entwicklung der Praktiken in den einzelnen Ländern, insbesondere in Europa.

Die nun folgenden Abschnitte beleuchten weitere Massnahmen, welche umgesetzt wurden, um die Anforderungen an die Migrationsbevölkerung bezüglich der im Gastland gesprochenen Sprache(n) strenger zu gestalten: die kürzliche (Wieder-)Einführung von sprachlichen Anforderungen im Migrations- und Integrationsbereich in verschiedenen Ländern, sowie deren Ausweitung auf weitere Etappen des Immigrations- und Integrationsprozesses; die zunehmende Verwendung formeller Sprachtests zur Überprüfung der Anforderungen; das von den Migrantinnen und Migranten verlangte Sprachkompetenzniveau sowie schliesslich landeskundliche Tests und deren implizite Anforderungen an Sprachkenntnisse.

5 | Böcker & Strik (2011) berichten nur über den Spezialfall der Beantragung einer Niederlassungsbewilligung, während die ursprüngliche grössere Studie (die INTEC-Studie, siehe Strik, Böcker, Luiten, & Oers, 2010) die verschiedenen Etappen des Immigrationsprozesses beleuchtet: Einreise und Aufenthalt, Niederlassung und Einbürgerung. Diese Studie bietet nicht nur einen detaillierten Überblick über die Sprachprüfungen in den analysierten Ländern, sondern enthält auch Daten über die Auswirkungen der entsprechenden Prüfungen. Diese wurden mittels informeller Interviews mit verschiedenen Stakeholdern (in den Prozess involvierte und davon betroffene Instanzen und deren Vertreter) gesammelt, um deren Standpunkte zu den lokalen Praktiken und ihren Konsequenzen zu identifizieren.

### 2.2.1 Die Ausweitung sprachlicher Anforderungen an Migrantinnen und Migranten

Sprachliche Anforderungen an Migrantinnen und Migranten sind zu einem zentralen Aspekt der Immigrations- und Integrationspolitik der meisten in diesem Bericht betrachteten Staaten geworden. Diese wurden nicht nur von mehr Ländern als zuvor eingeführt, sie wurden auch auf weitere Schritte auf dem „Weg des Migranten“ ausgeweitet (Saville, 2009).

Mehrere Studien berichten von einer erheblichen Zunahme der sprachlichen Anforderungen an Migrantinnen und Migranten, welche in manchen Fällen auch mit zusätzlichen Kosten verbunden sind und für jene Personen, welche die Anforderungen nicht erfüllen, Sanktionen nach sich ziehen können (Hogan-Brun et al., 2009b; Van Avermaet, 2009; Böcker & Strik, 2011). Aufgrund von Daten aus verschiedenen offiziellen Berichten stellte Extramania (2012) für den Zeitraum von 2003 bis 2008<sup>6</sup> einen Höhepunkt bei der Einführung neuer sprachlicher Anforderungen an Migrantinnen und Migranten fest (in Dänemark, Belgien, Deutschland, Griechenland, Norwegen, Österreich, den Niederlanden, Frankreich und Liechtenstein). Bezüglich der osteuropäischen Länder (die Slowakei, Armenien, Slowenien, Polen und die Türkei), kann ein solcher Höhepunkt zwischen 2007 und 2009 beobachtet werden. Extramania (2012) berichtet weiter, dass die Mehrheit der Länder, welche Einwanderung und Integration neu an sprachliche Bedingungen knüpfen, nun zusätzlich Sprachkurse für die Kandidatinnen und Kandidaten anbietet.

6 | Mar-Molinero (2006) erwähnt Spanien als eine bedeutende Ausnahme von diesem allgemeinen Trend. Hier ist die Staatsbürgerschaft nur schwer zu erhalten. Es gibt nur eine kleine Zahl von Einbürgerungen, keine sprachlichen Vorgaben für die Niederlassung oder Einbürgerung, keine als solche deklarierten Sprachtests. Stattdessen gibt es ein Interview mit einem Richter, in dem die „Anpassung“ an die lokale Kultur indirekt festgestellt wird. Das Fehlen von sprachlichen Anforderungen (Kenntnis der „nationalen“ Sprache) für die Einbürgerung hat mit der Geschichte des Landes zu tun (das Vermeiden von negativen Reaktionen der linguistischen Minderheiten in den *Comunidades autónomas*).

Interessanterweise werden, sobald ein Land strengere Vorschriften und Sprachtests eingeführt hat, diese von anderen Ländern übernommen: Als die niederländische Regierung im Jahre 2006 die Bewilligung von Aufenthalts- und Niederlassungsbewilligungen an die sprachlichen Kompetenzen der einwandernden Personen knüpfte, folgten fünf weitere EU-Staaten dem Beispiel, indem sie ihrerseits für den gleichen Zweck im gleichen Jahr sprachliche Anforderungen an Migrantinnen und Migranten einführten (Böcker & Strik, 2011; Van Avermaet, 2009).

Weiterhin existiert ein allgemeiner Trend, diese sprachlichen Anforderungen nach und nach auf allen drei Stufen des Immigrations- und Integrationsprozesses anzuwenden. Ursprünglich mussten Migrantinnen und Migranten in den meisten Ländern nur für die Einbürgerung gewisse Sprachkenntnisse nachweisen (dies wurde laut Böcker & Strik, 2011, in den 1990er Jahren übliche Praxis). In einem weiteren Schritt wurden diese Kenntnisse auch von jenen Personen verlangt, welche eine dauerhafte Niederlassungsbewilligung anstrebten. Anhand von Zahlen aus zwei Befragungen des Europarates zur mit der Einwanderung verbundenen Sprachenpolitik mehrerer europäischer Länder zwischen 2007 und 2009 stellte Extramania (2012) eine erhebliche Zunahme jener Staaten fest, die auch bei der Erteilung von Niederlassungsbewilligungen sprachliche Kompetenzen verlangen: Waren dies bei der Befragung 2007 noch 57% der teilnehmenden Länder, stieg diese Zahl bis 2009 auf 69%.

Noch neueren Datums, aber nicht so weit verbreitet, sind sprachliche Anforderungen, die bereits bei der Einreise ins Land (d.h. bei der Erteilung eines Visums) gestellt werden. Dies gilt z.B. für die Niederlande (Extra & Spotti, 2009b) und Deutschland (Michalowski, 2010). Extramania (2012) berichtet in diesem Zusammenhang von einer Zunahme von 19% der befragten Länder, die 2007 diese Anforderungen stellten, auf 26% im Jahre 2009. Laut Michalowski (2010) ist diese besonders restriktive Massnahme eine Reaktion der Niederlande und Deutschlands auf den zunehmenden Familiennachzug. Die Verpflichtung, vor der Einreise ins Land einen Sprachtest zu bestehen, wird so zu einer indirekten Massnahme, um diesen zu begrenzen.

Der Nachweis von Sprachkenntnissen als Bedingung für die Einreise wird eher kritisch gesehen (Goodman, 2011; Kostakopoulou, 2010), da davon ausgegangen wird, dass der Integrationsprozess erst beginnen kann, wenn das Individuum im Gastland physisch präsent ist und die Möglichkeit hat, mit der lokalen Gemeinschaft zu interagieren. Goodman (2011) formuliert dies folgendermassen:

Integrationsanforderungen vor der Einreise verlangen einen Grad der Integration in einen Staat, während sich der Kandidat physisch und konzeptuell – bezüglich des legalen Status<sup>4</sup> – *ausserhalb* dieses Staates befindet (S. 237).

Hinzu kommt, wie Extra & Spotti (2009a) und Hogan-Brun et al. (2009b) feststellen, dass der Zugang zur Zielsprache im Ausland zuweilen schwierig sein kann (z.B. weil keine Sprachkurse angeboten werden). In diesem Fall wird von den betroffenen Personen erwartet, dass sie eine Sprache beherrschen, die sie nicht lernen und üben konnten, zumindest nicht unter dafür günstigen Bedingungen.

## 2.2.2 Zunehmende Verwendung standardisierter Sprachtests

Zusammen mit der Ausweitung sprachlicher Anforderungen in der Einwanderungspolitik europäischer Staaten wird auch die Verwendung formeller Sprachtests immer verbreiteter (Van Avermaet, 2009). Alle drei Stufen der Immigration und Integration, Einreise, Niederlassung und Einbürgerung, sind betroffen. Laut Van Avermaet (2009), der über eine Befragung berichtet, die 2007 von der *Association of Language Testers in Europe* (ALTE) in 19 Ländern durchgeführt wurde, nutzten neun der elf Länder, welche für die Einbürgerung sprachliche Kompetenzen verlangen, einen Sprachtest, während sechs der acht Länder, welche bereits für die Niederlassung sprachliche Kompetenzen fordern, auch dafür einen Sprachtest vorschrieben. Ähnliches lässt sich auch bei den Bedingungen für die Einreise beobachten: Sieben der neun Länder, die auch hier Sprachkom-

petenzen erwarten, verwendeten für deren Nachweis einen Sprachtest. Mit anderen Worten: Beinahe alle Länder, die an Migrantinnen und Migranten sprachliche Anforderungen stellen, überprüfen mittels formeller Tests, ob diese das erwartete Niveau erreicht haben.

In einigen Fällen ersetzte ein standardisierter Test andere Formen der Sprachkompetenzfeststellung, die zuvor verwendet wurden (Hogan-Brun et al., 2009b). Böcker & Strik (2011) nennen die Einbürgerungsprozeduren in den Niederlanden und in Dänemark, wo die Sprachkenntnisse zuvor informell in einem Interviewgespräch mit einem oder einer Angestellten des Staates festgestellt wurden. Erst zu Beginn des 21. Jahrhunderts führten diese beiden Länder standardisierte Tests ein, um die Sprachkenntnisse der Anwärterinnen und Anwärter für die Staatsbürgerschaft zu erfassen.

## 2.2.3 Verlangte Sprachkompetenzniveaus

Die erwarteten Sprachniveaus für den Zugang zu einem Land (Einreisevisum), die Niederlassung oder die Erteilung der Staatsbürgerschaft unterscheiden sich von einem Land zum anderen erheblich. Die Übersicht in Van Avermaet (2009) über die entsprechenden Praktiken in 17 europäischen Ländern zeigt, dass die verlangte Sprachkompetenz für den *Zugang* zum Land zwischen A1.1 und B1 auf der Skala des Europarates liegt, jene für die *Niederlassung* zwischen A1 und B2 und für die *Einbürgerung* zwischen A2 und B2. Mit Bezug auf Deutschland, wo für die Staatsbürgerschaft das Sprachniveau B1 verlangt wird, hält Möllering (2010) fest, dass diese relativ hohe erwartete Kompetenz ansonsten interessierte Personen davon abhalten könnte, die Einbürgerung anzustreben, und schlägt vor, dass das Niveau A2, welches in den Niederlanden verlangt wird, besser geeignet wäre.

Die erwarteten sprachlichen Kompetenzen werden nicht immer auf explizit beschriebene Skalen bezogen, wie sie z.B. in den Referenzniveaus des *Gemeinsamen europäischen Referenzrahmens für Sprachen* zu finden sind (GER - Europarat, 2001). So verlangt z.B. die kana-

dische Gesetzgebung „die Fähigkeit, Englisch oder Französisch gut genug zu sprechen, um mit Leuten zu kommunizieren“ (Hargreaves, 2010); der australische Einbürgerungstest verlangt „Grundkenntnisse der englischen Sprache“ (Farrell, 2010; Möllering & Silaghi, 2010). In Bezug auf Letzteren kritisiert der *Woolcott Report* (siehe Möllering & Silaghi, 2010; McNamara & Ryan, 2011) die Ungenauigkeit dieser Sprachkompetenzbeschreibung: Es bleibt unklar, was Migrantinnen und Migranten konkret können müssen. Die fehlende Transparenz macht es für die Kandidatinnen und Kandidaten schwierig zu wissen, wie sie die Bedingungen für eine Verleihung der Staatsbürgerschaft erfüllen können.

Böcker & Strik (2011) verweisen darauf, dass die sprachlichen Voraussetzungen für die Erteilung der Niederlassungserlaubnis oder der Staatsbürgerschaft steigen, wenn auch schon welche für den Zugang zum Land definiert wurden. Mit anderen Worten: Die Einführung von sprachlichen Anforderungen zu einem frühen Zeitpunkt des Immigrationsprozesses hat offenbar einen verschärfenden Effekt auf die folgenden Etappen.

#### 2.2.4 Sprachkompetenztests und Landeskundetests

In einer Reihe von Ländern, darunter Grossbritannien, Dänemark oder die Niederlande, müssen Personen, welche eine Niederlassungserlaubnis oder die Staatsbürgerschaft anstreben, neben dem Sprachtest einen Test über Kultur und Gesellschaft des Gastlandes ablegen (Böcker & Strik, 2011; Van Avermaet, 2009). In dieser Art von Tests werden Fragen zu Religion, Kultur, Recht und Gesetz sowie zur Geschichte des Gastlandes gestellt. In Aufgaben wie dem Beherrschen und Verstehen des australischen Treuegelöbnisses („Australian Pledge of Allegiance“, Slade, 2010a), sieht Möllering (2010) eine auf Assimilierung abzielende Ausrichtung der Staatsbürgerschaft. Laut Slade (2010a) basiert die Verwendung derartiger Tests normalerweise auf der Annahme, dass Migrantinnen und Migranten durch das Bekanntmachen

mit den soziokulturellen Fakten, Normen und Werten der Gastkultur eben diesen Werten der Gesellschaft tatsächlich folgen werden. Dabei handelt es sich um einen Glauben, der kaum aufrechterhalten werden kann, so die Autorin.

Zwei weitere Probleme von landeskundlichen Einbürgerungstests werden in der Literatur aufgeworfen. So erfassen einige der Fragen in derartigen Tests, obwohl sie scheinbar auf reines Faktenwissen abzielen, in Wirklichkeit persönliche Einstellungen und Weltanschauungen der Kandidatinnen oder Kandidaten (z.B. bezüglich Homosexualität und Religion). Dies wird als Ungleichbehandlung (Joppke, 2010a; Michalowski, 2011) und potenzielle Quelle von sozialer Diskriminierung verstanden. Weiterhin hängt das erfolgreiche Ablegen des Tests häufig von den sprachlichen Kompetenzen der betroffenen Individuen ab, da die Tests meist in der oder einer der Amtssprachen des Gastlandes abgelegt werden. McNamara (2009b) stuft daher landeskundliche Tests als *de facto* Sprachkompetenztests ein. Als Beispiel dafür nennen Extra & Spotti (2009a) den niederländischen Sprach- und Kulturtest. McNamara & Ryan (2011) beschreiben ausserdem die Kontroverse um die Einführung des neuen australischen Einbürgerungstests im Detail. Diese entstand, weil die für den Test benötigten Sprachkenntnisse das offiziell angegebene Sprachniveau überstiegen (und auch nach einer Überarbeitung immer noch übersteigen).

Zusammenfassend gesagt nehmen die in Kapitel 2 erwähnten Studien eine generell kritische Haltung gegenüber den aktuellen Entwicklungen im Bereich der sprachlichen Anforderungen für Migrantinnen und Migranten ein. Sie zeigen auf, dass in den westlichen Ländern und insbesondere in Europa in den letzten zehn Jahren eine Tendenz hin zu schärferen sprachlichen Bedingungen auf allen Stufen des Immigrations- und Integrationsprozesses vorherrscht. Ausserdem wird eine Entwicklung von informellen hin zu formellen Sprachtests festgestellt. Während diese allgemeine Tendenz recht klar auszumachen ist, stellen die Studien aber auch grosse Unterschiede zwischen den Staaten fest, insbesondere mit Bezug auf die verlangten Sprachkompetenzen. Landeskundliche Tests werden fast ausschliesslich kritisiert,

nicht nur aufgrund der in ihnen verlangten Arten von Wissen, sondern auch, weil sie implizit höhere Sprachkompetenzen erfordern als die offiziell festgelegten Niveaus.

# 3

## Validitätsaspekte von Sprachprüfungen im Migrations- und Integrationsbereich

Im vorliegenden Kapitel wird gezeigt, wie und in welchem Masse die Literatur, die sich mit Sprachprüfungen im Migrations- und Integrationsbereich befasst, auch Aspekte der Testvalidität miteinbezieht. Die Qualität und Angemessenheit von (formellen) Beurteilungen wird normalerweise mithilfe eines Validitäts- oder Validierungsmodells festgestellt, welches die konfirmatorische oder kritische Diskussion und Analyse anleitet. Die folgende Präsentation der verschiedenen Aspekte von Validität folgt den Kategorien eines Validierungsmodells, das sowohl umfassend angelegt als auch in der internationalen Testforschung und -entwicklung allgemein anerkannt ist: Bachman und Palmers Rahmenmodell für die Testvalidität und Testvalidierung. Sie bezeichnen dieses als *Assessment Use Argument* (AUA). Zunächst werden im Folgenden die wichtigsten Aspekte dieses Modells kurz vorgestellt. Im Anschluss daran werden verschiedene Themen aus der Literatur, die sich mit Validität beschäftigt, hervorgehoben. Dabei wird Bezug genommen auf die Kategorien des AUA-Modells; wo nötig, werden Ergänzungen vorgenommen.

### 3.1

#### Bachman und Palmers argumentationsbasierter Ansatz zur Validierung von Beurteilungen

Die Validität von Beurteilungen ist eines der zentralen wiederkehrenden Themen der psychologischen, pädagogischen und eben auch der Sprachtestforschung der letzten Jahrzehnte. Das oben erwähnte AUA-Modell wurde erst kürzlich in Bachman & Palmer (2010) im Detail beschrieben und operationalisiert. Es basiert

auf zwei Quellen: a) konzeptuellen und praktischen Arbeiten verschiedener Autorinnen und Autoren<sup>7</sup>, darunter Vertreter von führenden Vereinigungen im Fachgebiet<sup>8</sup>; b) Bachmans und Bachman & Palmers früheren Arbeiten zum aufgabenorientierten, kommunikativen Sprachentesten (Bachman, 1990; Bachman & Palmer, 1996; Bachman, 2005). Kapitel 5 in Bachman & Palmer (2010) enthält einen kurzen Überblick über das AUA-Modell.

Hauptsächlicher Beweggrund für die Verwendung eines Validierungsmodells wie das AUA ist die Tatsache, dass Sprachtests üblicherweise Entscheidungen mit sich bringen, die für die beteiligten Parteien (im Folgenden *Stakeholder*) Konsequenzen haben, insbesondere für die Kandidatinnen und Kandidaten selber, für Lehrpersonen, Arbeitgeber und die Gesellschaft als Ganzes). Da die Entscheidungen, die aufgrund von Tests getroffen werden, Auswirkungen auf Individuen haben, sehen Bachman und Palmer Testentwickler und anbieter (z.B. Einwanderungsbehörden) in der Pflicht<sup>9</sup>, ihre Entscheidungen gegenüber diesen Individuen zu rechtfertigen. „Rechtfertigen“ bedeutet in diesem Zusammenhang, dass sie „aufgrund von Argumentation und empirischen Belegen zeigen müssen, dass die Verwendung eines bestimmten Testverfahrens gerechtfertigt ist“ (B&P<sup>10</sup>, S. 85f.). Ein ausgearbeitetes AUA besteht aus einer Reihe von gut gestützten Aussagen (*claims*)

7 | Insbesondere Messick (1989); Messick (1994); Kane et al. (1999); Kane (2004); Kane (2006); Mislevy et al. (2002); Mislevy et al. (2003); Kunnan (2004).

8 | Standards: American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.), (1999); Educational Testing Service (ETS), 2002; Verfahrensregeln: ILTA (International Language Testing Association), 2007; ALTE (Association of Language Testers in Europe), 1994; ALTE (Association of Language Testers in Europe), 2001; Handbücher: Linn, National Council on Measurement in Education, & American Council on Education (1989); Brennan, National Council on Measurement in Education, & American Council on Education (2006).

9 | Bachman & Palmer bezeichnen die Rechtfertigungspflicht den betroffenen Individuen gegenüber sowie die Verpflichtung nachzuweisen, dass die Verwendung eines Tests gerechtfertigt ist, als „zwei grundlegende Axiome“ ihres Ansatzes (Bachman & Palmer, S. 85).

10 | Im Folgenden verwendet für Bachman & Palmer (2010).

über die Validität eines Tests, welche (durch Inferenzen) miteinander verbunden sind und gemeinsam ein umfassendes Validitätsargument für den Gebrauch eines Beurteilungsinstrumentes liefern. Die üblichen Qualitätskriterien für (formelle) Beurteilungen, wie Test- und Interrater-Reliabilität, Unparteilichkeit gegenüber verschiedenen Gruppen von Teilnehmenden oder inhaltliche Validität sind integrale Bestandteile dieser Aussagen und des Validitätsarguments als Ganzem (siehe auch unsere Verweise auf diese Kriterien in eckigen Klammern in der Liste unten). Die Hauptaufgabe der Testentwickler und/oder -anbieter ist es, zu zeigen, dass die verschiedenen validitätsbezogenen Aussagen des AUA für den gegebenen Beurteilungszweck und die damit verbundenen Konsequenzen ausreichend erfüllt werden. Ein gültiges AUA rechtfertigt beides: die verwendete Testmethode und den Gebrauch des Tests für einen bestimmten Zweck innerhalb der Gesellschaft.

Im Folgenden werden die verschiedenen Aussagen, welche durch das AUA gestützt werden sollen, kurz aufgelistet.

#### **Aussage 1 – beabsichtigte Auswirkungen (intended consequences)**

Die Verwendung eines Beurteilungsinstrumentes und die Entscheidungen, welche aufgrund dessen getroffen werden, haben positive *Auswirkungen* auf die verschiedenen Stakeholder (Kandidatinnen und Kandidaten, Lehrpersonen, die Gesellschaft als Ganzes) (B&P, S. 105).

#### **Aussage 2 – Entscheidungen (decisions)**

Die *Entscheidungen*, die aufgrund der Testergebnisse getroffen werden, ...

- berücksichtigen die **Werte** der Gemeinschaft und die relevanten gesetzlichen Vorschriften und
- sind für alle von den Entscheidungen betroffenen Parteien **gerecht** [Unparteilichkeit] (B&P, S. 111).

#### **Aussage 3 – Interpretationen (interpretations)**

Die *Interpretationen* bezüglich der zu testenden Kompetenz sind ...

- **aussagekräftig** (meaningful) mit Bezug auf einen bestimmten Lehrplan, eine Analyse der Kompetenzen, die für die Ausführung einer Aufgabe in der Zielsprache benötigt werden, eine allgemeine Theorie der Sprachkompetenz oder bezüglich einer Kombination dieser Aspekte [Konstruktvalidität],
- **unparteilich** (impartial) gegenüber allen Gruppen von Teilnehmenden [Unparteilichkeit],
- **verallgemeinerbar** (generalizable) auf die ziel-sprachlichen Handlungsfelder<sup>11</sup>, für die die Entscheidung gefällt werden soll [Konstruktvalidität],
- **relevant** (relevant) für die zu fällende Entscheidung [Konstruktvalidität] und
- **ausreichend** (sufficient) für die zu fällende Entscheidung [Konstruktvalidität] (B&P, S. 114).

#### **Aussage 4 – Ergebnisse (assessment records)**

Die *Ergebnisse* (Punkte, Beschreibungen der Leistungen der Teilnehmenden) sind über die einzelnen Aufgaben, über unterschiedliche Aspekte der Beurteilungsmethode (z.B. Testvarianten, Testzeitpunkte, Beurteilende), sowie verschiedene Kandidatengruppen hinweg **konsistent** [Reliabilität, Unparteilichkeit] (B&P, S. 124).

Wie bereits erwähnt, sind eine umfassende Testvalidierung und das Beibringen von überzeugenden Belegen in High-stakes-Kontexten wie die Entscheidung über den Zugang, die Niederlassung und die Einbürgerung äusserst wünschenswert. Allerdings zeigt ein Überblick über die in diesem Artikel besprochene Literatur zum Thema Sprachtests im Migrations- und Integrationsbereich sofort, dass eine umfassende Validitätsargumentation so gut wie nie auch nur versucht wird. Viele Publikationen beschäftigen sich stattdessen nur mit einigen ausgewählten Aspekten von Validität, darunter vor allem Themen, die mit den Aussagen 1 (positive Auswirkungen), 2 (Werte und Unparteilichkeit) und 3 (Aussagekraft) zu tun haben. Diese Diskussionen sind ausserdem meist nur allgemein gehalten, d.h. nicht auf ein bestimmtes Instrument und dessen Verwendung bezogen.

Nur sehr wenige Publikationen zielen tatsächlich darauf ab, Validität umfassend aufzuzeigen (und/oder infrage zu stellen), oder systematisch eine Reihe von entscheidenden Aspekten eines einzelnen Tests und seiner spezifischen Verwendung(en) zu analysieren. Dies ist allerdings nicht überraschend, da Studien, die sich mit einem einzelnen im Migrations- und Integrationsbereich eingesetzten Instrument beschäftigen, generell sehr selten sind.

In den folgenden Abschnitten soll die Diskussion in der Literatur, die sich mit Sprachtests für Migrantinnen und Migranten beschäftigt, mit Bezug auf jede der vier Hauptkategorien (Aussagen zur Testvalidität) aus dem *Assessment Use Argument* (AUA) rekonstruiert werden. Dies bedeutet, dass hier nicht die Validitätsargumente für einzelne Tests untersucht werden, sondern die allgemeinen und kontextübergreifenden Themen in ihrem Bezug zu den Hauptaspekten von Validität im AUA.

Zusätzlich muss das Kategoriensystem des AUA angepasst werden, um Studien miteinbeziehen zu können, die Sprachtests im Migrations- und Integrationsbereich von einem radikaleren Standpunkt aus betrachten und die Frage stellen, ob Sprachtests – egal zu welchem Zweck – *gerecht* sind, also überhaupt eine legitime Massnahme im Einwanderungskontext darstellen.

McNamara und seine Kollegen (z.B. McNamara, 2006; McNamara & Roever, 2006; McNamara & Ryan, 2011) trugen erheblich zu der Diskussion bei, ob der Einsatz von Tests *gerecht* ist. Sie begreifen sich selbst als Teil der Strömung des *Critical Language Testing* (z.B. Shohamy, 2001; Shohamy, 2006), welche sich mit impliziten und expliziten Verbindungen zwischen Macht und Sprache beschäftigt. Gleichzeitig stützen sie sich auf Messicks grundlegende Publikationen zur Validität (Messick, 1989; Messick, 1996; McNamara, 2006), welche betonen, dass die *Verwendung eines Tests* ein integraler Bestandteil seiner Validität sein muss.

In einer neueren Veröffentlichung über *Gerechtigkeit* in Sprachtests stellen McNamara & Ryan (2011) „Unparteilichkeit“ (*fairness*) und „Gerechtigkeit“ (*justice*) als komplementäre Konzepte dar, welche gleichermassen für die Validierung von (Sprach)Tests notwendig sind.

11 | TLU: *Target-Language Use*.

Demnach beinhalten Fragen der Testfairness nicht nur die gleiche Behandlung aller Gruppen von Teilnehmenden und das Vermeiden psychometrischer Ungleichheit, sondern alle Aspekte, die mit der empirischen Validierung der Interpretation von Testresultaten zu tun haben. Fragen zur *Gerechtigkeit* von Tests ergänzen diesen Ansatz, indem auch die Konsequenzen der Interpretation von Testresultaten und deren Gebrauch (oder Missbrauch) in einem spezifischen sozialen Kontext<sup>12</sup> in Betracht gezogen werden, ebenso wie die *impliziten* sozialen und politischen Werte im Testkonstrukt.

Wahrscheinlich werden sämtliche Unparteilichkeits- und einige Gerechtigkeitsaspekte in Bachman und Palmers *Assessment Use Argument* abgedeckt. So zeigen Bachman (2005) und Bachman & Palmer (2010) kurz, wie der Missbrauch von Tests, welchen die Anhänger des *Critical Language Testing* (insbesondere Shohamy, 2001) beanstanden, im Rahmen des AUA behandelt werden kann. Bachman und Palmer argumentieren, dass vielen der identifizierten potenziellen Missbrauchssituationen durch das Qualitätskriterium „positive Auswirkungen“ entgegengewirkt werden könnte.

Allerdings widersprechen dem McNamara und Ryan (2011) in einer Diskussion von Messick (1989) und heben hervor, dass die Testforschung allein nicht mit „Themen von gesellschaftlichem Interesse“ umgehen kann. Ihnen zufolge muss das Testen auch

auf andere Formen der Analyse abgestützt werden, welche eher in den (kritischen) Kultur- und Politikwissenschaften<sup>13</sup> verbreitet sind, oft geschichtliche, ideologische und diskurskontextuelle Fragen miteinbeziehen und dabei vor allem qualitative Analysewerkzeuge nutzen (McNamara & Ryan, 2011, S. 164).

12 | McNamara & Ryan (2011) erwähnen Shohamys „verdeckte Motive“ („hidden agendas“) – Sprachtests, die insgeheim im Dienste der Gesellschaftspolitik stehen – als ein Thema, das mit Gerechtigkeit zu tun hat.

13 | engl. „cultural analysis and critical policy analysis“

### 3.2 Gerechtigkeit und Sprachtests im Migrations- und Integrationsbereich

Dieser Abschnitt beschäftigt sich mit der Diskussion um „Themen von gesellschaftlichem Interesse“ (McNamara & Ryan, 2011) in der Literatur über Sprachtests im Migrations- und Integrationsbereich. Er konzentriert sich dabei auf zwei teilweise zusammenhängende Bereiche, welche von Vertretern des *Critical Language Testing* kritisiert werden:

1. gesellschaftliche und politische Werte und Ideologien, die implizit in den Testkonstrukten enthalten sind;
2. eine Regulierung des Zugangs zur Gesellschaft durch Sprachtests, die oft nicht als solche deklariert wird.

Ein erheblicher Teil der uns zugänglichen Studien identifiziert und kritisiert Ideologien und nicht hinterfragte Werte, die als Grundlage und Rechtfertigung verwendet werden, um Anforderungen bezüglich der Sprachkompetenz (und des landeskundlichen Wissens) einzuführen und formell zu testen.

Eine dieser Ideologien ist der Mythos von Sprache als Symbol nationaler Identität und der Zugehörigkeit zu einer Gemeinschaft (Shohamy, 2009; Extra et al., 2009; Van Avermaet, 2012; Piller, 2001). Milani (2008) identifiziert in einer Analyse von öffentlichen Texten und Diskursen zur Einwanderung in Schweden den immer wiederkehrenden Glauben, das fehlende Beherrschen der schwedischen Sprache sei der Hauptgrund für die geringe Beteiligung von Migrantinnen und Migranten am politischen, sozialen und wirtschaftlichen Leben. Kompetenz in der Nationalsprache wird damit als notwendige Bedingung für ein „gutes Funktionieren“ in der Gesellschaft des Gastlandes wahrgenommen und soll es ermöglichen, die kulturellen Dimensionen der Gemeinschaft zu verstehen. Milani stellt fest, dass davon ausgegangen wird,

dass das Beherrschen einer gemeinsamen Sprache (Schwedisch) nicht nur den Zugang zum öffentlichen Raum mit seinen Rechten und Pflichten und politischer Partizipation sowie der Wirtschaft in

Form des Arbeitsmarktes darstellt, sondern in Wirklichkeit der EINZIGE Weg für Immigranten ist, eine Gesellschaft mit ihren Gesetzen, ihrem Alltag und ihren kulturellen Normen zu verstehen (Milani, 2008, S. 44).

Die Beherrschung der Nationalsprache wird in diesem Sinne als Indikator für eine erfolgreiche Integration in die Gemeinschaft des Gastlandes gesehen, während geringe Sprachkompetenzen oft als fehlender Wille des Einwandernden zur Integration gewertet wird (Shohamy, 2009). Einige Interpretationen gehen gar einen Schritt weiter und stufen die unzureichende Beherrschung der Nationalsprache (wegen Unvermögen oder Verweigerung) als Gefahr für den gesellschaftlichen Zusammenhalt und die nationale Identität ein (Blackledge, 2009c).

Einige Autorinnen und Autoren wie Oers (2010) und Mackenzie (2010) weisen darauf hin, dass sich parallel zur Verschärfung der Bedingungen für die Niederlassung und/oder die Einbürgerung auch das Konzept der Staatsbürgerschaft an sich gewandelt hat<sup>14</sup>. Mackenzie (2010) stellt eine Verlagerung vom Konzept der Staatsbürgerschaft als passivem Phänomen, welches auf spezifischen Rechten und Pflichten basiert, hin zu einem aktiven Phänomen fest, welches das Bekenntnis zu den politischen Institutionen voraussetzt. Aus diesem Grund wird Staatsbürgerschaft als Beleg für die Zugehörigkeit des Individuums zur Gemeinschaft verstanden:

Die Staatsbürgerschaft wird als Zeichen verstanden, welches die institutionelle Anerkennung einer Reihe von Voraussetzungen SYMBOLISIERT, die vermeintlich unerlässlich sind, damit ein Individuum die Identität des schwedischen Staatsbürgers annehmen und dadurch ganz Teil der schwedischen Gesellschaft werden kann (Milani, 2008, S. 43).

14 | Oers (2010) zufolge sind in der Literatur drei Staatsbürgerschaftskonzepte zu finden: a) Staatsbürgerschaft als rechtlicher Status (eine privilegierte Beziehung zwischen einer Person und einem Staat); b) Staatsbürgerschaft als Aktivität (Teilnahme am sozialen Leben der politischen Gemeinschaft); c) Staatsbürgerschaft als Identität (Mitgliedschaft in einer politischen Gemeinschaft, Identifikation, persönliche Loyalität, Bekennen zur Kultur der Gesellschaft).

In diesem Verständnis besteht ein direkter Zusammenhang zwischen erfolgreicher Integration und dem Annehmen der Staatsbürgerschaft. Dies ist allerdings offensichtlich eine Vereinfachung, da gezeigt wurde, dass Individuen in die Gesellschaft integriert sein können, ohne eingebürgert worden zu sein (Farrell, 2010), während andere kaum integriert sind, obwohl sie ihr Leben lang die Staatsbürgerschaft innehatten. Staatsbürgerschaft wird heute auch als Belohnung wahrgenommen, als etwas Wertvolles, das nicht nur Rechte und Pflichten, sondern auch Privilegien mit sich bringt (Cox, 2010).

Ein Konzept von Staatsbürgerschaft, welches als Höhepunkt des Integrationsprozesses gesehen wird, in dessen Verlauf der oder die Einzelne sein oder ihr Bekenntnis zur Gesellschaft des Gastlandes beweist, ist eng mit der Vorstellung von Staatsbürgerschaft als Assimilierung verbunden: Von den angehenden Staatsbürgerinnen und Staatsbürgern wird erwartet, dass sie ihre frühere Identität und ihre früheren Werte aufgeben und gänzlich die kulturellen und sozialen Normen des Gastlandes übernehmen (Kostakopoulo, 2010; Slade, 2010a).

Einigen Autorinnen und Autoren zufolge spielt die nationalistische Ideologie „eine Nation – eine Sprache“ in den meisten europäischen Ländern eine Rolle, die ein bestimmtes Kompetenzniveau in der „nationalen“ Sprache erwarten – dies zum Nachteil der regionalen Varietäten (Extra & Spotti, 2009a; Blackledge, 2009a; Shohamy, 2009) und unter Stärkung der existierenden sprachlichen und kulturellen Hegemonie (Hogan-Brun et al., 2009b; Shohamy, 2009). Von Migrantinnen und Migranten wird erwartet, dass sie ihre Integration in die Gesellschaft des Gastlandes durch das Erlernen der Standardsprache und die Beachtung der kulturellen Werte beweisen. Blackledge (2009c) geht sogar so weit zu sagen, dass Sprachtestsysteme für Englisch zum Zwecke der Festlegung standardisierter Niveaus „Englisch“ als eine homogene Menge von sprachlichen Praktiken erst erfinden oder konstruieren müssten, während doch ein Standard-Englisch in Wirklichkeit nicht existiere, da die Praktiken in verschiedenen Kontexten unterschiedlich sind. Van Avermaet (2009) weist darauf hin, dass das Testen von Sprachkompetenzen und landeskundlichem Wissen generell auf

der Annahme basiert, dass Gesellschaften bezüglich ihrer sprachlichen und kulturellen Normen homogen sind, weil simplifizierte Konstrukte einfacher zu testen sind.

Für McNamara steht klar die symbolische Funktion, welche die Beherrschung der Nationalsprache erfüllt, im Vordergrund:

Die Motivation für die Einführung sprachlicher Anforderungen ist nicht in erster Linie kommunikativer Natur, sondern hat mit der symbolischen Funktion von Sprache zu tun. Die Hauptfunktion des Tests ist nicht die Förderung des Wohlergehens der Migranten, sondern das Ausdrücken einer Ideologie, welche Sprachgebrauch und kulturelle Werte miteinander in Verbindung bringt (McNamara, 2009b, S. 158).

Daraus folgt für McNamara, dass das Testkonstrukt neu formuliert werden sollte, da diese Tests eigentlich nicht funktionale Sprachkompetenz testen (d.h. praktische kommunikative Fertigkeiten, die für die Kandidatinnen und Kandidaten im Alltag als niedergelassene oder eingebürgerte Personen relevant sein könnten), sondern implizit die „Anpassung an eine Menge soziokultureller Werte“ oder eine „nationale Ideologie“ überprüfen (McNamara, 2009b). In Wirklichkeit werden die Tests aufgrund der Dichotomie „akzeptabel“ versus „nicht akzeptabel“ bewertet, welche von Entscheidungsträgern statt von Sprachentestern festgelegt wird. Diese Dichotomie stuft die Teilnehmenden schliesslich als berechtigt oder unberechtigt für die Erteilung der Niederlassungserlaubnis oder der Staatsbürgerschaft ein (McNamara, 2010).

Auch für Kostakopoulo (2010) ist die Verbindung von Wissens- oder Fähigkeitstests und der Verbundenheit mit der Gesellschaft des Gastlandes nicht gegeben:

In der Tat findet man kaum eine Antwort auf die Frage, warum angenommen wird, dass „Zusammengehörigkeit“ („shared belonging“) durch das Testen sprachlicher Flüssigkeit in der Landessprache oder das Ansammeln von Fakten über Staatsbürgerkunde, Geschichte oder Leben im Lande, welche wenige Monate nach dem Test leicht wieder vergessen werden können, erreicht werden soll statt aufgrund gemeinsamer Erfahrungen, dem ge-

meinsamen Streben nach einem guten Leben und nach grösserem Wohlstand der Gemeinschaft (Kostakopoulo, 2010, S. 9).

Aus dieser Perspektive scheinen die üblichen Testpraktiken nicht besonders relevant.

Für Milani (2008) sind Sprachtests im Migrationsbereich für die Integration kaum förderlich, sondern erreichen eher das Gegenteil: Sie fördern soziale Ausgrenzung. Für ihn entsteht ein Widerspruch zwischen den im öffentlichen Diskurs deklarierten Zielen von Tests im Migrations- und Integrationsbereich und ihren eigentlichen Konsequenzen. Während Sprachtests die *Eingliederung (inclusion)* von Migrantinnen und Migranten in die Gesellschaft des Gastlandes fördern sollen (weil davon ausgegangen wird, dass soziale Kohäsion und nationale Identität nur durch eine gemeinsame Sprache erreicht werden können), führen diese in Wirklichkeit vielmehr zu *Ausgrenzung (exclusion)*, da es unvermeidlich ist, dass einige Personen durch den Test fallen werden. Milani (2008) zeigt aufgrund von Interviews mit Stakeholdern, dass die Migrantinnen und Migranten selber davon ausgehen, dass diese Form der Sprachpolitik mit dem Ziel umgesetzt wird, bestimmte Gruppen auszuschliessen. Milani hält fest, dass Sprachtests zu gesellschaftlicher Aufspaltung führen, da sie neue Grenzen zwischen sozialen Gruppen ziehen: Sie stellen jene, die den Test bestehen (und die Staatsbürgerschaft erhalten) denen gegenüber, die durchfallen (und ihre Position als ausländische Staatsangehörige behalten). Sprachtests erzeugen auch einen Unterschied zwischen der Gruppe der Migrantinnen und Migranten, die die Einbürgerung anstreben und damit einen Test ablegen müssen, und den Einheimischen, die ihn nicht bestehen müssen. Dadurch werden, anstatt soziale Unterschiede zu verringern, wie offiziell vorgegeben wird, diese bei Einbürgerungen durch Sprachtests reproduziert, indem bestimmte Gruppen von Menschen von einigen staatsbürgerlichen und kulturellen Bereichen ausgeschlossen werden (Milani, 2008).

Einige Autoren des *Critical Language Testing* (Shohamy, 1998, 2001) kritisieren, dass Tests im Migrations- und Integrationsbereich von den entsprechenden Regierungen als Regu-

lierungsinstrument (*gate-keeping*) (Shohamy, 2009; Blackledge, 2009c; Hogan-Brun et al., 2009b; Van Avermaet, 2009; Slade, 2010a; Wright, 2008) und für die Erhaltung von Privilegien der Staatsbürgerinnen und Staatsbürger (Piller, 2001) benutzt bzw. missbraucht werden. Sprachtests in diesem Sinne zu gebrauchen erlaubt es Regierungen nicht nur, den Einwanderungsstrom zu kontrollieren, sondern auch der eigenen Bevölkerung zu zeigen, dass alles unter Kontrolle ist (Böcker & Strik, 2011; Slade, 2010a). Die politischen Akteure können so behaupten, alles unternommen zu haben, um die soziale Kohäsion aufrechtzuerhalten (McNamara, 2009b), indem sie die Nation vor zu viel linguistischer und kultureller Vielfalt bewahrten (Kostakopoulou, 2010; Milani, 2008).

Für die Regulierung (*gate-keeping*) wird Sprache als Indikator für die Mitgliedschaft in einer sozialen Gruppe angenommen, was dazu führt, dass die Nicht-Erfüllung sprachlicher Anforderungen zur Ablehnung des Einbürgerungsantrags und damit zur sozialen Ausgrenzung führt. Milani (2008, S. 53) weist in diesem Zusammenhang auf die Tendenz einiger Nationalstaaten hin, „die MESSLATTE FÜR DIE MITGLIEDSCHAFT, welche den Zugang zur Ingroup regelt, immer höher anzusetzen“, indem sprachliche Anforderungen an Migrantinnen und Migranten gestellt werden, sobald deren Zahl steigt. Der alte Diktattest in Australien (*Australian Dictation Test*; McNamara, 2005) wird als extremes Beispiel für den Missbrauch eines Sprachtests durch eine Regierung angeführt. Ähnlich dem biblischen Schibboleth-Test zielte der Diktattest bewusst darauf ab, weniger willkommene soziale Gruppen zu diskriminieren, indem angehende Bewohnerinnen und Bewohner einen Test in einer Sprache ablegen mussten, die sie nicht beherrschten, was automatisch zum Nicht-Bestehen führte (McNamara, 2005).

Siiner (2006) berichtet von einem neueren Fall gesellschaftlicher Regulierung in Estland, welcher für viele Personen ernsthafte, sogar schädliche Auswirkungen hatte. Mit dem Ziel, die Nation nach dem Ende der sowjetischen Besetzung wiederherzustellen und Estnisch als einzige Nationalsprache zu konsolidieren, führte die Regierung die Beherrschung des Estnischen als Voraussetzung für die Staatsbürger-

schaft ein und überprüfte für deren Erlangung grundlegende mündliche und schriftliche Konversationsfähigkeiten. Viele russischsprachige Einwohnerinnen und Einwohner des Landes konnten diese Bedingung wegen mangelnder Kenntnisse in Estnisch nicht erfüllen. In der Folge wurde vielen Personen die estnische Staatsbürgerschaft vorenthalten, sogar denen, die in Estland geboren wurden und obwohl es an vielen Orten des Landes möglich ist, den Alltag ausschliesslich mit Russisch zu bewältigen. Ihnen wurden als Nicht-Staatsangehörigen einige grundlegende Rechte vorenthalten, darunter die Möglichkeit, sich für bestimmte Arbeitsstellen im öffentlichen Sektor zu bewerben. Dies führte dazu, dass ein grosser Teil der russischsprachigen Minderheit in Estland staatenlos wurde, da diese Leute weder die estnische noch die russische Staatsangehörigkeit erhielten und so schliesslich am Rande der Gesellschaft standen. Die Ausgrenzung von Minderheiten in den baltischen Staaten, welche mit unzureichender Beherrschung der Nationalsprache begründet und durch die nationalen Sprachengesetze legitimiert wurde, wurde von europäischen Institutionen erheblich kritisiert (Ozolins, 2003).

Shohamy diskutiert die Frage, ob Sprachtests für den Erhalt der Niederlassungserlaubnis oder der Staatsbürgerschaft *gerecht* sind, von einem radikaleren Standpunkt aus. Sie stützt sich dafür auf ethische und fundamentale politische Überlegungen. In Shohamy (2006) drückt sie ihren Widerstand gegen sprachliche Anforderungen und Sprachtests im Einwanderungs- und Integrationsbereich aus. Sie begründet dies mit dem Argument, es sei unethisch, Individuen von der vollständigen Teilnahme an der Gesellschaft auszuschliessen, nur weil diese die in einem bestimmten Kontext dominante Sprache nicht beherrschten. Darauf aufbauend argumentiert sie weiter, dass die Verweigerung der Staatsbürgerschaft aufgrund fehlender Sprachkenntnisse eine Verletzung der Persönlichkeitsrechte des Individuums und damit diskriminierend ist. Staatsbürgerliche Rechte und Pflichten werden denen zuteil, die die dominante Sprache beherrschen, nicht aber anderen, die sie nicht können. Shohamy geht davon aus, dass die Verweigerung der Staatsbürgerschaft aufgrund fehlender sprachlicher Kompetenzen

kein gerechter Grund ist, da „kein Beweis vorliegt, dass das Beherrschen der Nationalsprache notwendigerweise bessere Staatsbürger erschafft“ (Shohamy, 2006, S. 148). Weiterhin hinterfragt Shohamy (2009; siehe auch Blackledge, 2009a) die Legitimation eines Staates, durch Sprachgesetze und Sprachtests einem Individuum den Gebrauch und das Erlernen einer Sprache aufzuzwingen. Sie verteidigt ihre Position folgendermassen:

Man fragt sich, ob die Aneignung einer „nationalen“ Sprache nicht eine freie Wahl von Personen sein sollte, die ihre eigenen besten rationalen Entscheidungen treffen können, was die Sprache angeht, die sie in einer mehrsprachigen, grenzübergreifenden Welt können und benutzen müssen (Shohamy, 2009, S. 49).

Mit anderen Worten: Personen zu zwingen, eine Nationalsprache zu erlernen, stellt eine Verletzung ihres Rechts dar, frei jene Sprache zu sprechen, die ihren täglichen Bedürfnissen am besten angepasst ist, und verwandelt die Beherrschung der Nationalsprache in eine „Bürgerpflicht“ (Shohamy, 2009).

Vor diesem Hintergrund verurteilen Shohamy und andere kritische Wissenschaftler (z.B. Blackledge, 2009c; Cox, 2010; McNamara & Shohamy, 2008; Van Avermaet, 2012) die gesellschaftlichen Konsequenzen von sprachlichen Bedingungen für den Zugang zur Niederlassung und Staatsbürgerschaft sowie die politische Instrumentalisierung von Sprachtests zu diesem Zweck. Mit ihren Arbeiten bemühen sie sich, Testanbieter für die Auswirkungen ihrer Prüfungen auf das Leben der Kandidatinnen und Kandidaten zu sensibilisieren und fordern sie auf, für die (missbräuchliche) Verwendung der Tests Verantwortung zu übernehmen.

Im Grossen und Ganzen wird der Begriff der Gerechtigkeit als ein Mittel für oft radikale Kritik an sprachlichen Anforderungen und Sprachtests im Migrationsbereich eingesetzt. So wird zum Beispiel hinter sprachlichen Anforderungen, welche dem Anschein nach zum Wohle der Teilnehmenden festgelegt wurden (z.B. für deren Integration), ideologisches Denken wie die Vorstellung „eine Nation – eine (Standard-)Sprache“ aufgedeckt. Ebenso erfassen

Sprach- und Landeskundetests kritischen Stimmen zufolge in erster Linie die Anpassung von Individuen an bestimmte soziokulturelle Werte, da die Behauptung, die getesteten Konstrukte seien für die Teilnahme an der Gesellschaft entscheidend, nicht aufrechterhalten werden kann. Kritische Sprachentestende heben auch hervor, dass sprachliche Anforderungen und Tests in Wirklichkeit zu noch mehr sozialer Benachteiligung und Ausgrenzung führen statt zu mehr Integration, obwohl dies von den zuständigen Behörden meist als Ziel deklariert wird (Testen als gesellschaftliche Regulierung, als *gate-keeping*). Schliesslich existiert eine kritische Strömung, nach der sprachliche Anforderungen und deren Auswirkungen eine Verletzung individueller Grundrechte darstellen und damit unethisch und diskriminierend sind.

### 3.3 Positive Auswirkungen des Testens

Dem AUA-Modell zufolge ist es ein Zeichen von Testvalidität, wenn die Konsequenzen, die sich aus der Verwendung und Interpretation eines Tests ergeben, für alle Stakeholder positive Auswirkungen haben und für niemanden schädlich sind. Dies gilt sowohl für Individuen wie auch für die Gesellschaft als Ganzes.

Einige der damit verbundenen Argumente wurden bereits im vorhergehenden Abschnitt über Gerechtigkeit diskutiert, dies aber eher aus einer allgemeinen (und fundamental kritischen) Perspektive heraus. Im vorliegenden Abschnitt liegt der Fokus mehr auf konkreten Aspekten, mit denen sich Studien beschäftigt haben, welche die Auswirkungen von sprachlichen Anforderungen auf Individuen (z.B. den Zugang zum Arbeitsmarkt oder zum Bildungssystem) und die Gesellschaft (z.B. Folgen für den gesellschaftlichen Zusammenhalt) untersucht haben. Ein erster Abschnitt fasst die verschiedenen vermuteten oder tatsächlichen Effekte von sprachlichen Anforderungen und Sprachtests zusammen, welche in den verschiedensten Publikationen zu finden sind. In einem zweiten Abschnitt werden dann die Ergebnisse von spezifisch ausgerichteten Wirkungsstudien (*impact studies*) vorgestellt.

Üblicherweise werden strengere Bedingungen im Migrations- und Integrationsbereich von politischen Entscheidungsträgern, und gelegentlich auch Testentwicklern, mit praktischen Überlegungen begründet und die potenziell positiven Effekte für die Migrantinnen und Migranten hervorgehoben (Blackledge, 2009c). De Jong et al. (2009) bemühen sich z.B. deutlich, ihren neuen Test für gesprochenes Niederländisch zu rechtfertigen, indem sie den Nutzen von besseren Sprachkompetenzen sowohl für die Kandidatinnen und Kandidaten als auch für die Gesellschaft betonen. Die Autorinnen und Autoren argumentieren, dass fehlende Integration (aufgrund von fehlenden Sprachkenntnissen) zu „einem Teufelskreis [führt], in dem Eltern ihren Kindern nicht helfen können, Kinder die Schule abbrechen, was dann zu Gefühlen der Hilf- und Hoffnungslosigkeit führt“. Demgegenüber soll der Test helfen, die „negativen Konsequenzen der sozialen Ausgrenzung einer grossen Zahl von Immigranten in den Niederlanden“ zu bekämpfen. Auch das Erfassen von Sprachkompetenzen vor der Einreise ins Gastland soll positive Auswirkungen haben, indem es helfe, Zwangsehen und Frauenhandel, „eine neue Form der Sklaverei“, zu bekämpfen. Zusätzlich werden die Ergebnisse einer Umfrage zitiert, welche vor Beginn der Testentwicklung durchgeführt wurde. Diese sollen zeigen, dass

das Gesetz für die Integration neuer Einwohner, trotz seines Potentials für politische Kontroversen, bezüglich seines Zwecks und seiner Bestimmung von einer grossen Mehrheit unterstützt und als vertretbar angesehen wurde, dies sowohl innerhalb als auch ausserhalb des Parlaments (und sogar unter den interviewten Einwanderern) (De Jong et al., 2009, S. 43).

Diese Art der Argumentation ist auch anderswo zu finden. Die übliche Begründung für die Einführung von sprachlichen Anforderungen in Einwanderungsgesetzen ist, dass diese die Integration der Migrantinnen und Migranten in die neue Gesellschaft erleichtern, deren aktive Partizipation am alltäglichen und politischen Leben fördern (Blackledge, 2009c; Oers et al., 2010; Holland, 2010) und ökonomische Unabhängigkeit ermöglichen (Böcker & Strik, 2011).

Angesichts dieser Vorteile sollen sie mit einem ausreichenden Niveau „funktioneller“ Sprachkompetenz ausgestattet werden (Blackledge, 2009c). So wurde z.B. in Deutschland das Niveau B1 des Gemeinsamen Europäischen Referenzrahmens für Sprachen (GER, Europarat, 2001) mit dem Ziel festgelegt, Personen, welche die Niederlassung oder Einbürgerung anstreben, sowohl im privaten wie im beruflichen Leben unabhängig zu machen (Oers et al., 2010). Die Einführung eines recht anspruchsvollen Sprachtests sei somit gerechtfertigt, weil er wirtschaftlich unabhängige und gesellschaftlich aktive neue Staatsbürgerinnen und Staatsbürger garantieren solle. Zusätzliche Kurse und Tests über Geschichte, Kultur und Gesetze des Gastlandes sollen das Repertoire vervollständigen, welches eine Person benötigt, um voll am gesellschaftlichen Leben teilzunehmen (Cooke, 2009; Milani, 2008).

Im Abschnitt über Gerechtigkeit weiter oben wurde gezeigt, dass Argumentationen, welche auf die positiven Auswirkungen der beschlossenen Massnahmen abzielen, in einer Reihe von Studien sehr skeptisch aufgenommen werden. Allerdings existieren auch andere Studien, welche die positiven Auswirkungen (strengerer) sprachlicher Anforderungen als Fakt darstellen. So argumentiert z.B. Kiwan (2008), dass die Vorbereitung auf den britischen Einwanderungstest als eine Möglichkeit für Migrantinnen und Migranten gesehen werden kann, sich mit dem notwendigen Wissen über ihre Rechte und Pflichten als zukünftige Staatsbürgerinnen und Staatsbürger vertraut zu machen. In diesem Sinne kann der Test als Mittel gesehen werden, um das Verständnis für aktive politische Teilnahme zu fördern. Es sei aber betont, dass nicht der Sprachtest selber diesen Beitrag zur Integration leisten soll, sondern die Vorbereitung darauf in einem Integrationskurs. In ähnlicher Weise sieht Yoffe (2010) positive Effekte des Aufnahme- und Integrationsvertrags (*Contrat d'Accueil et d'Intégration, CAI*), welcher 2007 in Frankreich eingeführt wurde. Dieser verpflichtet neu angekommene Migrantinnen und Migranten zu Sprachkursen, wenn deren Französischkenntnisse unter dem Referenzniveau A1.1 des GER liegen. Yoffe sieht in der Verpflichtung, Sprachkurse zu besuchen,

eine Möglichkeit insbesondere für bildungsferne muslimische Frauen, ihr soziales Umfeld zu verlassen und mit Muttersprachlerinnen und Muttersprachlern zu kommunizieren. Anderenfalls, ohne verpflichtende Sprachkurse, wäre es diesen Frauen nicht möglich, mit Einheimischen in Kontakt zu kommen. Ein weiterer positiver Effekt ist Yoffe zufolge die Tatsache, dass Migrantinnen und Migranten, welche ihre Sprachkurse erfolgreich abschließen, „Vertrauen [gewinnen] in ihre Fähigkeit, in der Gesellschaft zu funktionieren, und einen Ansporn [haben], die Sprache weiter zu lernen“ (Yoffe, 2010, S. 77).

Empirische Studien, welche die Konsequenzen der Testverwendung untersuchen, werden in der Validierungsforschung als Wirkungsstudien (*impact studies*) bezeichnet (im Bildungsbereich wird in diesem Zusammenhang auch von *Washback*-Studien gesprochen). Diese Studien identifizieren die Auswirkungen von Sprachtests auf die Gesellschaft als Ganzes (z.B. auf den gesellschaftlichen Zusammenhalt) sowie auf das Leben der Kandidatinnen und Kandidaten (z.B. besserer Zugang zum Arbeitsmarkt oder zu Ausbildungsmöglichkeiten).

Wirkungsstudien über Sprachtests im Migrations- und Integrationsbereich sind eher selten, wohl z.T. deshalb, weil die Einführung formeller Anforderungen und Prüfungen ein relativ neues Phänomen ist (Böcker & Strik, 2011). Van Avermaet & Rocca (2013) stellen fest, dass sich Wirkungsstudien tendenziell auf statistische Angaben beschränken wie die Bestehensraten eines Tests oder die Anzahl der Teilnehmer und Teilnehmerinnen in Vorbereitungskursen. Sie können damit nur wenig über die Auswirkungen eines Sprachtests auf den Integrationsprozess oder die gesellschaftliche Teilhabe der Migrationsbevölkerung aussagen. Aus diesem Grund fordern die beiden Autoren mehr Forschungsarbeiten zu den gesellschaftlichen Auswirkungen von Einwanderungsgesetzen auf den Integrationsprozess von Migrantinnen und Migranten.

Eine wichtige aktuelle Studie, die eben diese Fragestellung untersucht, ist die INTEC-Studie (Strik et al., 2010). Darin werden die tatsächlichen Praktiken untersucht, die sich aus sprachenpolitischen Entscheidungen und aus der Verwendung von Tests auf den ver-

schiedenen Stufen des Immigrationsprozesses (Einwanderung, Niederlassung oder Einbürgerung) ergeben. Es werden auch konkret die Auswirkungen untersucht, die sich aus verschiedenen Testverfahren in neun EU-Mitgliedsstaaten ergeben (Österreich, Belgien, Dänemark, Frankreich, Deutschland, Ungarn, Lettland, die Niederlande und Grossbritannien). Die Studie enthält zwei Arten von Daten: a) statistische Informationen über die Anzahl der ausgestellten und verweigerten Aufenthaltsbewilligungen und b) qualitative Daten aus insgesamt 329 Interviews mit verschiedenen Stakeholdern, darunter vor allem Migrantinnen und Migranten, die von den Anforderungen betroffen waren, aber auch mit Sprachlehrpersonen, Staatsangestellten und Mitarbeitenden von Nichtregierungsorganisationen (NGOs).

Statistische Daten über Einbürgerungsverfahren zeigen, dass in einigen Ländern die Zahl der Migrantinnen und Migranten, welche die Staatsbürgerschaft erhielten, gesunken ist, seit die zusätzlichen Anforderungen eingeführt wurden. So stieg z.B. in Dänemark die Zahl der verweigerten Einbürgerungen zwischen 2002 und 2003, als formelle Sprachtests erstmals eingeführt wurden, und noch einmal zwischen 2007 und 2009, als das geforderte Sprachniveau von B1 auf B2 erhöht wurde. Auch Oers (2008) stellt für 2004 in den Niederlanden, direkt nach der Einführung des Einbürgerungstests, einen signifikanten Rückgang der Anträge auf Einbürgerung fest (70% weniger Anträge 2004 als 2002). Im Falle der beiden Staaten, welche Tests eingeführt haben, die vor der Einreise ins Land abgelegt werden müssen (die Niederlande und Deutschland), verzeichnet die INTEC-Studie einen Rückgang der Zahl der Anträge, vor allem bei Familienzusammenführungen. Dies bedeutet, dass für einige oder möglicherweise die meisten der betroffenen Frauen, welche ihrem Partner – dem (zukünftigen) Ehemann – folgen wollten, die Einführung neuer Testverfahren Nachteile mit sich brachte. Wegen Nichtbestehen des Tests konnten Familien nicht zusammengeführt und Ehen nicht geschlossen werden.

Die Ergebnisse der INTEC-Studie, welche durch die Analyse der qualitativen Daten gewonnen wurden, zeigen ebenfalls, dass die Aus-

wirkungen der neuen Testverfahren von Migrantinnen und Migranten selber häufig als nachteilig wahrgenommen werden. Was die Einbürgerungsverfahren angeht, standen Migrantinnen und Migranten in Österreich, Dänemark, Deutschland, Grossbritannien und Lettland einem positiven Einfluss dieser Tests auf ihren persönlichen Integrationsprozess skeptisch gegenüber, z.B. in den Bereichen Arbeitsmarkt und soziale Kontakte. Stattdessen identifizierten die Befragten andere Faktoren als hilfreich für ihre Integration. Dies ist insofern nicht verwunderlich, als die meisten von ihnen sich zum Zeitpunkt ihres Antrages auf Einbürgerung bereits integriert fühlten. Einige von ihnen wiesen explizit auf die negativen Folgen von Stress und Beklemmung hin, welche sie (und ihre Familien) erlebten, als sie den Test ablegten, der für ihr Leben von erheblicher Wichtigkeit war. Auch die Vorbereitungskurse wurden kritisiert: Sie seien zeitintensiv und mit der Arbeitswelt unvereinbar.

Doch die INTEC-Studie (Strik et al., 2010) berichtet auch von positiven, also vorteilhaften Effekten. So berichteten einige Deutschlehrpersonen von Fällen, in denen die Teilnehmenden den Test vor der Einreise nach Deutschland absichtlich nicht bestanden, um Zwangsehen zu vermeiden. Einige der befragten Anwärtinnen und Anwärter auf die Staatsbürgerschaft waren sich einig, dass der Vorbereitungskurs, den sie besuchen mussten, einen „emanzipatorischen Effekt“ hatte. Verpflichtende Sprachkurse (z.B. in Deutschland oder Österreich) stellen offenbar für einige Frauen eine einzigartige Möglichkeit dar, mit anderen in Kontakt zu kommen, da es ihnen nicht erlaubt wäre, an freiwilligen Integrationsprogrammen teilzunehmen. Die Möglichkeit, soziale Kontakte zu pflegen, zeigt positive Auswirkungen auf das Selbstwertgefühl dieser Frauen mit bescheidenem Bildungshintergrund.

Die Ergebnisse einer kleineren Wirkungsstudie in Flandern, vorgestellt in Van Avermaet & Rocca (2013) und Van Avermaet (2012), zeigen hingegen nur wenige positive Auswirkungen. Die Studie basiert auf 40 halbstrukturierten Interviews mit verschiedenen Stakeholdern (Migrantinnen und Migranten, Lehrpersonen, aber auch Personen, welche die „Mehrheit“ repräsentieren, wie z.B. Arbeitgeberinnen und Arbeitgeber). In Flandern müssen Neuankömmlin-

ge und bereits angesiedelte Migrantinnen und Migranten verpflichtende Integrationskurse besuchen, um das Referenzniveau A1 in Flämisch und einige grundlegende Kenntnisse der Gesellschaft zu erwerben. Am Ende dieser Integrationskurse erhalten erfolgreiche Teilnehmende ein Integrationszertifikat. Die Entscheidung über Bestehen und Nichtbestehen liegt damit bei der Lehrperson.

Migrantinnen und Migranten, welche den Kurs mehr als ein Jahr zuvor besucht hatten und daraufhin keine Arbeitsstelle fanden, stimmten, wenig überraschend, in den Interviews darin überein, dass der Integrationskurs und das Zertifikat für die Arbeitssuche nicht hilfreich waren. Die Interviews mit Arbeitgebern und Arbeitsvermittlungsstellen brachten zum Vorschein, dass das Integrationszertifikat in Bewerbungen keine Rolle spielt und die Sprachkompetenz der potenziellen Angestellten informell in Vorstellungsgesprächen überprüft wird. Ganz allgemein schienen Sprachkenntnisse für den Zugang zum flämischen Arbeitsmarkt kein wichtiger Faktor zu sein. Die Studie schliesst daraus, dass staatliche Integrationsmassnahmen im Ganzen nur begrenzte Auswirkungen auf den tatsächlichen Integrationsprozess der Migrantinnen und Migranten haben.

Insgesamt zeigen die verfügbaren Studien, dass die angeblichen Vorteile von sprachlichen Anforderungen und Sprachtests im Migrationsbereich ein kontroverses Thema bleiben. Politische Entscheidungsträger und Testentwickler tendieren dazu, neue Erlasse mit den potenziellen Vorteilen für die Migrationsbevölkerung und die Gesellschaft zu rechtfertigen: bessere Integration und grössere wirtschaftliche Unabhängigkeit durch bessere Kenntnisse der Sprache und der Gesellschaft. Die wenigen existierenden Studien stützen diese Behauptungen aber nur am Rande: Einige Forschende beobachteten Einzelpersonen, die von den neuen Prüfungsverfahren und den damit verbundenen Vorbereitungskursen profitierten, darunter muslimische Frauen, welche die Möglichkeit erhielten, ausserhalb ihrer Familien Kontakte zu knüpfen. Wirkungsstudien, die auf statistischen Daten und Interviews mit einer ausreichenden Zahl an Betroffenen basieren, zeigen vor allem negative oder überhaupt keine Konsequenzen

auf. Dort, wo Sprachtests für die Erteilung eines Visums vorgeschrieben sind, können weniger Personen einreisen (darunter v.a. Frauen, welche planen zu heiraten); wo ein Sprachzertifikat notwendig ist, um die Einbürgerung zu beantragen, bekommen weniger Personen die Staatsbürgerschaft; und für jene, die die Staatsbürgerschaft anstreben und bereits gut integriert sind, bedeuten Integrationskurse und -prüfungen vor allem höhere zeitliche Belastung und mehr Stress. Eine Studie zeigt, dass Zertifikate, die für das Erfüllen der Anforderungen ausgestellt werden, auf dem Arbeitsmarkt nur einen geringen oder überhaupt keinen Wert haben. Es kann also festgehalten werden, dass in der Literatur insgesamt nur wenige empirische Belege für optimistische Behauptungen vorliegen, strengere Anforderungen an die sprachlichen und kulturellen Kenntnisse würden einen Nutzen mit sich bringen. Allerdings werden dringend weitere Wirkungsstudien benötigt.

### 3.4 Wertesensibilität der Entscheidungen

Um zu erreichen, dass ein Beurteilungsinstrument als wertesesensibel gilt, schlagen Bachman & Palmer (2010) vor, dass sich Testentwickler mit den Werten, die die verschiedenen Stakeholder haben könnten, beschäftigen und sie auf die eine oder andere Weise anerkennen. Sie gehen sogar noch einen Schritt weiter, indem sie dazu auffordern, die existierenden Werte der Gemeinschaft zu hinterfragen, wenn diese die faire und unparteiliche Behandlung aller Teilnehmenden in Gefahr bringen.

Aufgrund der vorliegenden Literatur kann im Grossen und Ganzen festgestellt werden, dass Werte in der Testentwicklungspraxis wenig Beachtung finden, während sie bei gesellschaftskritischen Wissenschaftlerinnen und Wissenschaftlern, die sich mit sprachlichen Anforderungen und Sprachtests beschäftigen, eine Priorität darstellen.

Die Art, wie Plassmann, Repräsentantin von telc, einem der Entwickler des Deutsch-Tests für Zuwanderer (DTZ), mit der Tatsache

umgeht, dass Teilnehmende von Integrationskursen an deren Ende das Sprachniveau B1 (das üblicherweise als anspruchsvoll angesehen wird) nachweisen müssen, könnte eine Erklärung für dieses scheinbare Desinteresse liefern. Sie argumentiert schlicht, dass das Niveau B1 vom Gesetzgeber vorgegeben wurde und es daher unnötig sei, darüber zu diskutieren, welches Niveau möglicherweise angemessen wäre (Plassmann, 2011, S. 25). Die niederländischen Entwicklerinnen und Entwickler des Tests für gesprochenes Niederländisch (TGN) befinden sich in einer vergleichbaren Situation wie ihre deutschen Pendanten – auch der TGN wurde von den Behörden in Auftrag gegeben –, sie beschäftigen sich aber kurz mit Wertefragen, bevor sie die technische Validierungsarbeit vorstellen (De Jong et al., 2009). Sie erwähnen dabei mehrere Fakten und Erkenntnisse, die nachweisen sollen, dass sich ihr Test in Bezug auf Werte und Akzeptanz auf sicherem Boden bewegt. Sie heben hervor, dass die Einführung des neuen Tests auf einem Parlamentsbeschluss mit grosser Mehrheit beruht, dass er vom Justizministerium in Auftrag gegeben wurde und dass dieses auch die GER-Skalen als Referenz ausgewählt hat. Sie erwähnen ebenfalls, dass die politische Opposition für ihre Ansichten keine Unterstützung gefunden hat, auch nicht unter den Migrantinnen und Migranten des Landes. Daraufhin erläutern sie den erwarteten gesellschaftlichen Nutzen ihres Tests: Der TGN soll Integration fördern, da fehlende Integration zu einem „Teufelskreis“ führe, der in Hoffnungslosigkeit und Verzweiflung münde; Tests vor der Einreise können bei der Bekämpfung von Frauenhandel helfen; und all jene, welche minimale Anstrengungen unternehmen würden, könnten ohne Weiteres einreisen, da der Test nicht allzu anspruchsvoll sei (De Jong et al., 2009, S. 41-43). Die Autorinnen und Autoren des niederländischen Tests nehmen eine klar affirmative Haltung gegenüber ihrem Test ein und beschäftigen sich nicht ernsthaft mit gegensätzlichen Ansichten<sup>15</sup>.

15 | In einigen Abschnitten ist sogar ein Hauch von Triumph und Spott zu spüren.

### 3.5 Unparteilichkeit der Entscheidungen

Ob eine auf einem Test basierende Entscheidung gerecht ist oder nicht, hängt laut Bachman & Palmer (2010) vor allem von zwei Aspekten ab: Verschiedene Kandidatinnen und Kandidaten mit der gleichen Kompetenz müssen mit der gleichen Wahrscheinlichkeit in die gleiche Gruppe eingeteilt werden (z.B. „bestanden“ oder „nicht bestanden“) und alle Kandidatengruppen müssen die gleichen Möglichkeiten haben, sich die notwendigen Kompetenzen anzueignen. Die Zugehörigkeit zu einer sozialen Gruppe sollte keinerlei Einfluss auf die Entscheidungen haben, die aufgrund des Tests gefällt werden.

Die Vorstellung, dass formelle Tests (zumindest potenziell) gerechter sind als z.B. ein informelles Interview mit einem Staatsangestellten, wird in mehreren von uns gelesenen Studien erwähnt. Joppke (2010; ähnlich Wright, 2008) betrachtet die Einführung von standardisierten Einbürgerungstests als fairer und weniger willkürlich als die frühere informelle Beurteilung, da nun alle Betroffenen gleich behandelt werden. Mehrere Studien berichten von Beispielen aus informellen Prüfungsverfahren, welche zu offensichtlicher Ungleichbehandlung führten, darunter Schneider et al. (2006) zu den lokalen Einbürgerungspraktiken in der Schweiz oder Piller (2001) zu ähnlichen Verfahren in Deutschland, welche in hohem Masse von individuellen Vorlieben von Staatsangestellten abhängig waren. Laversuch (2008) zeigt, wie diese Situation für die Regulierung der Einwanderung missbraucht wurde, insbesondere in Bezug auf Muslime. Die ansonsten eher testskeptische ALTE LAMI-Gruppe erwähnt die Vorteile von „korrekt geplanten, konstruierten und durchgeführten“ Tests, z.B. dass die „Resultate hochgradig standardisiert und reliabel sind“ und die „Kandidaten mit einem hohen Grad an Unabhängigkeit und Objektivität geprüft werden“. Sie verlangen daher nach hohen Qualitätsstandards für die Entwicklung und Durchführung von Tests, wenn diese überhaupt für Entscheidungen im Migrationsbereich verwendet werden sollen (Balch et al., 2008).

Während das Potential von Tests, grössere Gerechtigkeit mit sich zu bringen, von mehre-

ren Autorinnen und Autoren erwähnt wird, werden hochgradig formalisierte Testverfahren auch kritisiert, weil diese neue Hürden erzeugen, welche bestimmte Gruppen der Migrationsbevölkerung benachteiligen. Shohamy weist daraufhin, dass viele Migrantinnen und Migranten, welche nicht lesen und schreiben können oder nur wenig Schulbildung haben, mit den Testverfahren nicht vertraut sind und dementsprechend häufig versagen. Böcker & Strik (2011) bemängeln auch, dass finanziell schlechter gestellte Personen die Gebühren für den Test nicht übernehmen können. Für Michalowski (2010) und Shohamy (2009) sind Sprachtests schliesslich Selektionsinstrumente, da nur hochqualifizierte und gebildete Gruppen alle Hürden überwinden und so den gewünschten Aufenthaltsstatus erreichen können.

Mehreren Studien zufolge beginnt die Ungerechtigkeit im Übrigen lange vor dem Test selber, da verschiedene Personen unterschiedlich guten Zugang zu Informationen, Sprachunterricht und Testvorbereitung haben. Dies betrifft vor allem Tests, die vor der Einreise ins Land abgelegt werden müssen (Deutschland und die Niederlande). Befragte der INTEC-Studie (Strik et al., 2010) in Marokko weisen darauf hin, dass die Beschaffung von Vorbereitungsmaterial mit einer Reise zur Botschaft verbunden ist und dass in den Regionen, in denen viele der Antragstellenden leben, keine Sprachkurse angeboten werden.

Shohamy (2009) schreibt, dass es für Migrantinnen und Migranten sogar schwierig sein kann, im Gastland selbst mit der Sprache in Kontakt zu kommen, und dass sie oft zu wenig Gelegenheit haben, die Zielsprache so gut zu erlernen, wie es für den Einbürgerungstest notwendig ist.

Krumm (2007) äussert weiter Zweifel, dass die Durchführung des gleichen Tests für alle betroffenen Personen zu mehr Gerechtigkeit führt. Er beobachtet grosse Unterschiede in der Migrationsbevölkerung, einerseits bezüglich ihres sprachlichen und kulturellen Hintergrundes und der Vorbildung, andererseits bezüglich der Sprachkompetenzen, über die sie verfügen müssen, um in Gesellschaft und Arbeitswelt zurechtzukommen. Krumm zufolge entsteht ein Problem, wenn diese Personen getestet werden sollen:

Personen, die in allen Aspekten ihrer linguistischen und kulturellen Fähigkeiten und Kompetenzen unterschiedlich sind, mit ein und demselben Test zu prüfen und von ihnen das gleiche Kompetenzniveau in allen Bereichen zu erwarten, kann unmöglich ein Weg sein, um Gleichbehandlung in der Gesellschaft zu demonstrieren (Krumm, 2007, S. 668).

Basierend auf statistischen Daten des Deutsch-Tests für Zuwanderer (DTZ), fand Klein (2013) heraus, dass einige Gruppen von Teilnehmenden grössere Probleme hatten, den DTZ zu bestehen als andere. Sie schlägt daher vor, diesen benachteiligten Gruppen mehr Unterstützung bei der Testvorbereitung zukommen zu lassen: Frauen mit Türkisch als L1, Männer mit den L1 Russisch oder Polnisch sowie ältere Personen, vor allem mit den L1 Italienisch, Chinesisch oder Türkisch. Oers (2008) konnte ebenfalls beobachten, dass bestimmte Gruppen durch die gängigen Testformate besonders benachteiligt werden. Er erwähnt ältere Personen und solche mit niedrigem Bildungsniveau. Für Oers ist das Testen von surinamischen Migrantinnen und Migranten in den Niederlanden ein weiteres Beispiel für ungerechte Behandlung: Obwohl ihre L1 Niederländisch ist, müssen sie den Niederländisch-Test bestehen, um die Staatsbürgerschaft zu erhalten, während die niederländische Bevölkerung den Test natürlich nicht ablegen muss. Stevenson & Schanze (2009) identifizieren eine weitere Quelle von Ungleichbehandlung angesichts der Tatsache, dass viele gut ausgebildete Migrantinnen und Migranten wie Bürgerinnen und Bürger der EU und von Ländern wie den USA oder Australien die sprachlichen Anforderungen nicht erfüllen und keinen Test bestehen müssen, unabhängig davon, wie gut ihre Kenntnisse der Lokalsprache sind.

Insgesamt entsteht zum Thema der Unparteilichkeit der Tests ein gemischtes Bild. Während anerkannt wird, dass Tests oft klar parteiliche informelle Praktiken ersetzt haben, weisen einige Autorinnen und Autoren gleichzeitig auf diskriminatorische Aspekte der neuen Testverfahren hin.

### 3.6 Aussagekraft der Interpretationen bezüglich eines zugrundeliegenden Konstrukts

Damit die Interpretationen von Prüfungsergebnissen im Hinblick auf den Gebrauch der Zielsprache in der „realen“ Welt (*target-language use* – TLU) Aussagekraft haben können, muss das Kompetenzkonstrukt, das dem Instrument zugrunde liegt, auf diesen Zielsprachgebrauch bezogen sein (z.B. aufgrund einer Bedürfnisanalyse), so Bachman & Palmer (2010). Die Aufgaben, aus denen der Test besteht, müssen mit den Aufgaben aus der realen Welt in relevanten Punkten übereinstimmen und die Bewertung der Leistungen der Teilnehmenden muss jene Aspekte widerspiegeln, die in der Konstruktdefinition hervorgehoben werden. Zudem müssen das Konstrukt und die damit verbundenen Anforderungen den verschiedenen Stakeholdern, darunter auch den Kandidatinnen und Kandidaten, klar kommuniziert werden. Der Anspruch, dass Ergebnisse in diesem Sinne aussagekräftig sein sollen, hat mit Aspekten zu tun, welche früher oft unter dem Begriff der Konstruktvalidität diskutiert wurden.

Auch das Testkonstrukt wird in den hier diskutierten Studien thematisiert. So präsentieren Saville & Van Avermaet (2008) in einer Diskussion über den Gebrauch von Sprachtests im Migrations- und Integrationsbereich einige zentrale Themen, welche bei der Entwicklung von angemessenen Sprachprüfungen beachtet werden müssen. Darunter ist auch das Testkonstrukt. Den Autoren zufolge muss das Testkonstrukt so definiert werden, dass es den offiziellen Anspruch an den Test, die Integration der Teilnehmenden zu fördern, widerspiegelt. Der Test wiederum muss Informationen darüber liefern, wie gut das getestete Individuum in der Gesellschaft des Gastlandes funktionieren kann (anstatt z.B. dekontextualisiertes Sprachwissen zu überprüfen). Das erwartete Sprachniveau sollte ebenfalls diesem Ziel des Tests entsprechend gewählt werden. Wenn der Test also die gesellschaftliche Integration des Teilnehmenden überprüfen soll, muss das erwartete Sprachniveau auch den Anforderungen entsprechen, welche der gesellschaftliche Alltag an die

Sprachkompetenzen der Migrantinnen und Migranten stellt. Mit dem Ziel einer konkreten Umsetzung dieser Forderungen verlangt der Leitfaden der ALTE LAMI-Gruppe (Balch et al., 2008) als Erstes, dass die realen Anforderungen an die Migrationsbevölkerung aufgrund einer Bedürfnisanalyse identifiziert werden, welche die praktischen Situationen und Aufgaben untersucht, mit denen sich die Kandidatinnen und Kandidaten (hier: zukünftige niedergelassene Einwohnerinnen und Einwohner oder Staatsangehörige) in ihrem Alltag konfrontiert sehen werden. Dank der Identifizierung dieser Anforderungen des wirklichen Lebens kann dann in einem weiteren Schritt festgelegt werden, welche Kompetenzen und Fähigkeiten in der Zielsprache getestet werden sollen. Die LAMI-Gruppe schlägt ausserdem vor, dass die Bedürfnisanalyse nicht für „Migrantinnen und Migranten“ generell durchgeführt werden soll, sondern für verschiedene Untergruppen (z.B. Arbeitsmigrantinnen und -migranten, Ehepartnerinnen und -partner usw.), da deren reale Bedürfnisse unterschiedlich sein können. Shohamy (2009) weist darauf hin, dass viele Tests häufig auf unrealistische Sprachstandards abzielen (z.B. linguistische Korrektheit), welche die Besonderheiten von Zweitsprachgebrauch und mehrsprachigem Alltag nicht in Betracht ziehen. Sprachtests repräsentieren daher nicht die alltäglichen Bedürfnisse von Migrantinnen und Migranten im Gastland, wodurch ihre Eignung als Grundlage für Entscheidungen infrage gestellt wird.

Fulcher (2004) stellt fest, dass viele der Länder, welche sprachliche Anforderungen als Teil des Immigrations- und Integrationsprozesses eingeführt haben, diese in Bezug auf die Referenzniveaus des GER definiert haben (Europarat, 2001; North, 2009). Dieses Vorgehen wird in den vorliegenden Studien sehr kontrovers behandelt.

Die Kontroverse um die Verwendung der GER-Skalen betrifft in erster Linie, aber nicht

nur<sup>16</sup>, die Definition des Konstrukts. Krumm (2007) macht darauf aufmerksam, dass die Referenzskalen des GER nicht ursprünglich dafür entworfen wurden, die Sprachkompetenzen von Migrantinnen und Migranten zu erfassen, sondern jene klassischer Fremdsprachenlernender. Krumm (2007) nennt als Beispiel einen Deskriptor auf Niveau A1, welcher das Element „einen Toast ausbringen“ enthält, was offensichtlich keine sehr entscheidende Fähigkeit für die Mehrheit der Migrationsbevölkerung ist, für die die sprachlichen Anforderungen eingeführt wurden. In Übereinstimmung mit den Empfehlungen der LAMI-Gruppe schlägt er vor, dass das Testkonstrukt und die damit verbundenen Deskriptoren an jene Sprachgebrauchskontexte angepasst werden, mit denen die Migrantinnen und Migranten tatsächlich konfrontiert sind. Ausserdem sollte beachtet werden, dass diese häufig neben ihrer Erstsprache auch andere Sprachen beherrschen und in ihrer mehrsprachigen Realität einsetzen. Adäquate Deskriptoren sollten daher interkulturelle und mehrsprachige Dimensionen enthalten. Krumm weist auf eine weitere Facette von Mehrsprachigkeit hin, welche beachtet werden sollte, da sie in der Migrationsbevölkerung häufig auftritt: ein ungleichmässiges Fertigungsprofil in ein und derselben Sprache. So entwickeln Migrantinnen und Migranten oft unterschiedliche Kompetenzen in der gleichen Sprache, z.B. geringe Schreib-, aber gute Lese- und Hörkompetenzen. Krumm (2007) kommt daher zu dem Schluss, dass die Skalen des GER auf den spezifischen Kontext „Migration“ angepasst werden sollten, v.a. durch weitere Entwicklungsarbeit in den Bereichen Teilkompetenzen und mehrsprachige Repertoires. Im Hinblick auf die Überprüfung dieser Sprachkompetenzen schlägt er vor, dass das Europäische Sprachenportfolio besser geeignet sein könnte als ein Test, um unterschied-

16 | Extra, Spotti, & Van Avermaet (2009) und auch Yoffe (2010) kritisieren z.B. die Verwendung der GER-Skalen als normatives Instrument im Migrationsbereich, da ihnen zufolge die Beispielskalen des Referenzrahmens ursprünglich entstanden, um *deskriptiv* individuelle Sprachkompetenzen anzuerkennen. Weiterhin entstand der GER, um internationale Mobilität zu fördern und hat in Einwanderungsgesetzen den genau gegenteiligen Effekt: Der Referenzrahmen wird Teil eines Systems sozialer Ausgrenzung und Regulierung.

liche Profile zu erfassen (vgl. auch Papp, 2010).

Shohamy (2007) steht der Eignung des GER als Grundlage für die Konstruktdefinition im Migrationsbereich ähnlich kritisch gegenüber. Ein Kritikpunkt betrifft die höheren Kompetenzniveaus in den GER-Skalen, da diese Beschreibungen spezifische Annahmen bezüglich des Wissens, der Fähigkeiten und des kommunikativen Funktionierens enthalten, welche nicht auf die Migrationsbevölkerung verallgemeinert werden können. Shohamy kritisiert auch, dass die Deskriptoren den Kontext dieser Zielgruppe nicht ausreichend reflektieren: den Zweck der Beurteilung, das Alter der Lernenden, die verschiedenen Lernkontexte oder die funktionale Verteilung der verschiedenen gesprochenen Sprachen.

Alderson (2007) weist auf ein etwas anders gelagertes Problem hin, das sich aus der Verwendung des GER ergibt. Er beobachtet, dass Vertreterinnen und Vertreter der Politik und der Zivilbevölkerung dazu neigen, die von der Migrationsbevölkerung erwarteten Kompetenzstandards allein festzulegen, einzig aufgrund der Skalen im Referenzrahmen, anstatt Sprachexpertinnen und -experten zu Rate zu ziehen. Cooke (2009) nennt als Beispiel den britischen Einbürgerungstest, für welchen das erwartete Kompetenzniveau ("Entry 3" auf der *Adult ESOL Core Curriculum*-Skala, äquivalent zu Niveau B1 des GER) von der Regierung festgelegt wurde. In der Folge wurde diese Festlegung von Sachverständigen aus dem Bereich des Englischen als Zweit- und Fremdsprache (*English for Speakers of Other Languages*, ESOL) als deutlich zu hoch kritisiert. Die Situation in Deutschland ist vergleichbar: Auch dort wurden die sprachlichen Anforderungen für die Niederlassung und Einbürgerung von der Regierung auf Niveau B1 festgelegt (Plassmann, 2011), was wiederum als unrealistisch kritisiert wurde (Laversuch, 2008).

Einige Studien stellen die verwendeten Sprachkompetenzkonstrukte generell in Frage. Shohamys radikale Sicht, dass Sprachtests im Migrationsbereich das persönliche Recht verletzen, die eigene Sprache zu verwenden, wurde bereits weiter oben erwähnt. Von einem etwas praktischeren Standpunkt aus stellt Shohamy (2009) die Annahme infrage, dass Kompetenz in der Landessprache notwendig ist, um in der Ge-

sellschaft zu funktionieren. In den Fällen, in denen sie *tatsächlich* gebraucht wird, müssten Umfang und Niveau überdacht werden. Ähnlich wie Shohamy relativieren Saville & Van Avermaet (2008) die Bedeutung der Sprachkompetenz in der Landessprache, indem sie darauf hinweisen, dass die Schwierigkeiten von Migrantinnen und Migranten bei der Arbeitssuche und dem Zugang zu Bildung nicht unbedingt auf fehlende Sprachkenntnisse zurückzuführen sind, sondern ein Ergebnis der Ausgrenzung sind, unter der sie als Fremde eher zu leiden haben als andere. Sie betonen weiter, dass diese Personen, da sie in mehrsprachigen Gesellschaften leben, ihren Alltag eventuell ohne Probleme mit anderen Sprachen als der offiziell anerkannten bewältigen können; dass die Verwendung dieser Sprachen aber als Mittel zur gesellschaftlichen Teilhabe nicht anerkannt ist. Slade (2010a) merkt an, dass das sprachliche und kulturelle Wissen, welches Migrantinnen und Migranten mit der Zeit durch persönliche Erlebnisse erwerben können, sehr vielfältig sein kann, was es schwierig macht, (gut fundierte) Standards festzulegen, die alle erreichen sollten.

Papp (2010) berichtet von einer kleinen Validierungsstudie, die sie über den britischen Niederlassungs- und Einbürgerungstest "Life in the UK" (2005 eingeführt) durchführte. Sie untersuchte dafür den Test selber und das dazugehörige Vorbereitungsmaterial. Die Studie fokussiert in erster Linie auf die Frage, ob die Testmaterialien den erwarteten Sprachgebrauchskontext und das deklarierte Kompetenzniveau widerspiegeln („funktionale Kompetenz, die notwendig ist, um sich in der Gesellschaft erfolgreich als Einwohner/in oder Bürger/in zu bewegen“). Papps Studie ergab zwei hauptsächliche Befunde. Erstens zeigt die Analyse des Tests und seiner Dokumentation, dass die getestete Sprachkompetenz keinen angemessenen Nachweis für die Fähigkeit der Kandidatinnen und Kandidaten liefert, dass sie sich in die Gesellschaft des Gastlandes integrieren können. Die Inhalte des Tests entsprechen auch nicht dem Sprachgebrauchskontext, der für die Teilnahme am gesellschaftlichen Leben relevant ist (z.B. Anstellung, öffentliche Verwaltung, Ausbildung). Zweitens erweist sich das Sprachniveau im Test und den Vorbereitungsmaterialien

höher als das offiziell angegebene Zielniveau. Papp kommt zu diesem Ergebnis aufgrund einer korpuslinguistischen Analyse. Dieses recht negative allgemeine Urteil macht auf ein Problem mit der Aussagekraft (und der Konstruktvalidität) aufmerksam, da das Testdesign und die Interpretation der Ergebnisse nicht mit dem ursprünglichen Testkonstrukt übereinstimmen.

Der Bericht über den Australischen Einbürgerungstest (*Australien Citizenship Test, ACT*) wurde von einem Ausschuss um Woolcott, einem Diplomaten, aufgrund grosser Kritik erstellt. Der Bericht konzentrierte sich auf die Angemessenheit des im Test geforderten Sprachniveaus, sowie einige umstrittene Inhalte und Testverzerrungen. Laut dem Einbürgerungsgesetz des Landes sollte der Test *grundlegende Kenntnisse (basic knowledge)* der englischen Sprache überprüfen. In den Stellungnahmen, die Sprachexperten für den Woolcott-Ausschuss erstellten, wurde kritisiert, dass die sprachlichen Anforderungen in der Vorbereitungsbrochüre und in den Formulierungen der Testitems deutlich über „grundlegende Kenntnisse“ der englischen Sprache hinausgingen. Der Woolcott-Bericht versuchte, die Forderung „grundlegender Kenntnisse“ konkreter als „ausreichende Kenntnisse des Englischen, um in der australischen Gesellschaft zu existieren“ zu definieren, sowie diese in der Nähe der GER-Niveaus A1/A2 zu platzieren (McNamara & Ryan, 2011). Offenbar war dieser Klärungsversuch aber von geringer praktischer Bedeutung: McNamara & Ryan (2011) zufolge blieb das Sprachniveau im überarbeiteten Test und seiner Brochüre nahezu unverändert.

Kunnan (2009) stellt einen kritischen Vergleich an zwischen dem angegebenen Zweck (und damit dem impliziten Konstrukt) des neuen US-Einbürgerungstests, eingeführt 2008, und seiner Umsetzung. Der offizielle Zweck des Tests ist die Förderung von politischer Teilhabe und gesellschaftlicher Integration. Der Autor fand in diesem Zusammenhang zwei hauptsächliche Probleme. Eines betrifft die Art des durch den Test abgefragten staatsbürgerlichen Wissens, welches nicht über auswendig zu lernende Fakten hinausgeht, während doch Aufgaben eingesetzt werden sollten, die ein aktives Verständnis für die amerikanische Geschichte und das politische System der USA fördern und verlangen. Ein

zweiter problematischer Punkt betrifft Art und Umfang der englischen Sprachkenntnisse, die nötig sind, um den Test zu bestehen. Kunnan (2009) zufolge ist das Niveau wahrscheinlich nicht hoch genug, um nachzuweisen, dass sich die Migrantinnen und Migranten am politischen und gesellschaftlichen Leben der US-Gesellschaft beteiligen können. Angesichts dieser und anderer Schwächen kommt Kunnan (2009) zu dem Schluss, dass der US-Einbürgerungstest seinen angegebenen Zweck, gesellschaftliche und politische Integration zu fördern, nicht erfüllen kann, wodurch sein Sinn grundsätzlich infrage gestellt wird.

Der Einbürgerungstest in seiner aktuellen Form kann staatsbürgerliche Zugehörigkeit und gesellschaftliche Integration nicht durch indirekte Messungen von englischen Sprachkenntnissen und Faktenwissen über US-Geschichte und Staat erfassen, da es sich dabei um Fertigkeiten und Wissen handelt und nicht um Masse für gesellschaftliche Teilhabe und Engagement (Kunnan, 2009, S. 95).

Kunnan erwähnt einen weiteren Punkt, der nicht für den neuen Einbürgerungstest spricht: Die niedrigen Einbürgerungsraten, welche darauf hindeuten, dass der Test wohl interessierte Personen abschreckt.

Über den offiziellen deutschen Einbürgerungstest DTZ existieren nur wenige Artikel von Personen, die an seiner Erstellung beteiligt waren. Perlmann-Balme (2011) und Plassmann (2011) versuchen beide zu zeigen, dass ihre Arbeit den Qualitätsstandards der ALTE entspricht, einer Organisation, der die meisten europäischen Testanbieter angehören. Die ALTE-Standards 1 und 2 beschäftigen sich mit der Aussagekraft von testbasierten Interpretationen mit Bezug auf den Zielsprachgebrauch:

- 1) Stellen Sie sicher, dass Ihre Prüfung sich auf ein theoretisches Konstrukt bezieht, z.B. auf ein Modell der kommunikativen Kompetenz.
- 2) Beschreiben Sie die Ziele der Prüfung sowie die Verwendungssituationen und die Adressaten, für die die Prüfung geeignet ist (ALTE, 2007).

Die Entwickler des DTZ bemühen sich, diese Standards einzuhalten, indem sie einen generell

handlungsorientierten Ansatz, wie er im GER (Europarat, 2001) separat für die vier „klassischen“ Fertigkeiten beschrieben wird, mit einer Orientierung an den spezifischen Bedürfnissen der Zielgruppe(n) kombinieren. Das deutsche Rahmencurriculum für Integrationskurse (Buhlmann et al., 2007) bildet für Letzteres die Grundlage. Für das DTZ-Konstrukt wurden aus den 12 Domänen im Ursprungsdokument jene ausgewählt, die den verschiedenen Gruppen von Teilnehmenden gemeinsam sind. Das Konstrukt und die Ziele des DTZ werden in einem DTZ-Handbuch der Öffentlichkeit zugänglich gemacht (Perlmann-Balme et al., 2009). In diesem Dokument befindet sich auch ein Beispieltest.

Perlmann-Balme und Plassmann beschreiben in ihren Artikeln die Schritte, welche unternommen wurden, um einen Test zu entwickeln, der das Konstrukt aussagekräftig umsetzt und gleichzeitig die ALTE-Standards beachtet. Leider führen sie nicht ausreichend Belege an, um ihre Lesenden zu einem unabhängigen Schluss kommen zu lassen.

De Jong et al. (2009) unternehmen einen deutlich weiter gehenden Versuch, um zu belegen, dass ihr telefon- und computerbasierter Test des gesprochenen Niederländisch sinnvolle Aussagen im Kontext des Testens von Migrantinnen und Migranten erlaubt – dies obwohl der Test vornehmlich Itemformate umfasst, welche intuitiv nicht mit kommunikativen Aufgaben aus dem realen Leben in Verbindung gebracht würden. Für den TGN werden drei Itemtypen verwendet: Nachsprechen von Sätzen, Kurzantworten (was einem authentischen Aufgabenformat am nächsten kommt) und Gegensatzpaare von Einzelwörtern<sup>17</sup>. Es gibt keinen menschlichen Gesprächspartner; der gesamte Test wird von einem Computer abgenommen und ausgewertet. Alle Itemformate beinhalten gesprochenen Input und einigen gesprochenen Output; damit sind Hören und Sprechen involviert. Den Autorinnen und Autoren zufolge,

misst der Test die Fähigkeit der Kandidat/innen zu verfolgen, was gesagt wird, Bedeutungen in Echt-

zeit zu erfassen und relevante, verständliche Antworten in normaler Gesprächsgeschwindigkeit zu formulieren und zu produzieren (De Jong et al., 2009, S. 43).

Durch eine Darstellung der Anforderungen der Items in dieser Weise werden die Parallelen zu kognitiven und linguistischen Prozessen sichtbar gemacht, welche auch in der mündlichen Interaktion eine grosse Rolle spielen.

Im Artikel werden noch weitere Argumente genannt, um die Verbindung zwischen Sprachgebrauch im Test und tatsächlichem Zielsprachgebrauch aufzuzeigen:

- Die Stimuli in den Aufgaben zum Nachsprechen wurden aus einem Korpus des gesprochenen Niederländisch ausgewählt, welcher alltägliche spontane Sprache aus verschiedenen Regionen enthält.
- Das automatische Bewertungssystem wurde an einer grossen Stichprobe muttersprachlicher und nicht-muttersprachlicher Sprecher des Niederländischen trainiert, um die gesamte Palette des Niederländischen zu erfassen, wie es im realen Leben gebraucht wird.
- Der TGN bewertet Wortschatz, Grammatik, Aussprache und Flüssigkeit als Indikatoren für die mündliche Sprachkompetenz der Teilnehmenden. Dies entspricht der Praxis bei den üblichen kommunikativen Tests zum Sprechen, bei denen dieselben sprachlichen Merkmale bewertet werden, weil sie als fundamental angesehen werden.

Ein grosser Unterschied zwischen den beiden Arten von Tests und auch zwischen dem Sprachgebrauch im TGN und dem im alltäglichen Leben bleibt, nämlich der Grad an Offenheit der Aufgaben: Während die drei TGN-Formate die (Re)Produktion von Sprache eng steuern, verlangt der Sprachgebrauch im wahren Leben oft nach aktiver Konstruktion und Aushandlung des Diskurses.

Offensichtlich erwarteten die Entwicklerinnen und Entwickler diese potenzielle Schwäche in der Validitätsargumentation. Sie liessen daher von einer grossen Zahl Personen zusätzliche Aufgaben ausführen, darunter Nacherzählungen,

17 | Am Ende jedes Tests werden zwei Aufgaben zum Nacherzählen von Geschichten gestellt. Diese dienen aber lediglich der Testvalidierung und zu Forschungszwecken.

das Beantworten offener Fragen und mündliche Interviews. Die dabei entstandenen Leistungen wurden von menschlichen Beurteilern bewertet, einige davon mit Bezug auf die Kompetenzskalen des GER, so dass alle Bewertungen zusammen in einer Analyse zusammengefasst werden konnten und zu einer einzelnen GER-bezogenen Skala führten. In der Folge konnte diese Skala genutzt werden, um die TGN-Skala, welche aus maschinell erstellten Bewertungen von Vokabular, Grammatik, Aussprache und Flüssigkeit besteht, ausreichend genau an den GER anzubinden. Dies gilt insbesondere für den unteren Skalenbereich, wo die wichtigen Entscheidungen über Bestehen und Nichtbestehen gefällt werden.

Es überrascht nicht, dass das Thema der Aussagekraft von testbasierten Interpretationen – welches den grössten Teil dessen abdeckt, was früher als Konstruktvalidität behandelt wurde – in den von uns diskutierten Veröffentlichungen grosse Beachtung findet. Die verschiedenen Texte beschäftigen sich u.a. mit der Frage, wie ein angemessenes Testkonstrukt aussehen sollte, ob ein Konstrukt über eine einzelne Sprache hinaus gültig sein sollte, ob die GER-Skalen in diesem konkreten Kontext sinnvoll sind oder ob sprachliche Anforderungen überhaupt gerechtfertigt werden können. Andere untersuchen existierende Tests kritisch; wieder andere präsentieren und rechtfertigen die Konstrukte eigener Tests. Um diesen letzten Fall darzustellen, wurde die Argumentation der Autorinnen und Autoren zum Konstrukt für den Test des gesprochenen Niederländisch hier genauer besprochen, da diese in ihrer Gründlichkeit und Transparenz aussergewöhnlich ist. Es sollte aber dabei nicht vergessen werden, dass diese Studie bei all ihren Qualitäten im Grossen und Ganzen dennoch ein affirmativer Text ist, welcher verfasst wurde, um zu zeigen, dass der Test für seinen Zweck geeignet ist.

### **3.7 Unparteilichkeit der Interpretation gegenüber allen Kandidatengruppen**

Bachman & Palmer (2010) zufolge können die Interpretationen von Prüfungsergebnissen nur dann unparteilich sein, wenn alle Aspekte der

Testdurchführung, darunter die Formate und Inhalte des Tests sowie der Zugang zu den Informationen zum Test keine Gruppe von Teilnehmenden spezifisch benachteiligen.

De Jong et al. (2009) sowie Perlmann-Balme (2011) und Plassmann (2011), also die Publikationen in unserer Sammlung, welche tatsächlich Details über ihre Entwicklungsarbeit offenlegen, beschäftigen sich mit Aspekten der Unparteilichkeit.

So vermuten De Jong et al. (2009) offenbar in der Durchführung des Tests in niederländischen Botschaften weltweit ein potenzielles Problemfeld, welches zu Ungleichbehandlung führen könnte. In ihrem Artikel rechtfertigen oder verteidigen sie diese Praxis mit folgenden Argumenten: 1) Die Erklärungen zur Testdurchführung sind für alle Kandidatinnen und Kandidaten ausreichend, weil sie üblicherweise von ausgebildeten Mitarbeiterinnen oder Mitarbeitern in deren Herkunftssprache gegeben werden. Wo solch eine Person nicht verfügbar ist, haben die Kandidatinnen und Kandidaten die Möglichkeit, einen Übersetzer oder eine Übersetzerin mitzubringen. 2) Während es generell bei hochgradig dezentralisierten Testsystemen leicht zu Betrug und damit zu Ungleichbehandlung kommen kann, nehmen die Entwicklerinnen und Entwickler des TGN für sich in Anspruch, dass dies nahezu unmöglich ist, da der eigentliche Test zentral von einer grossen Item-Datenbank zur Verfügung gestellt wird, für jeden Kandidaten einzigartig ist und von einem Computersystem maschinell bewertet wird.

Weiter bemüht man sich im Text auch, zu zeigen, dass die Itemformate und inhalte nicht zu Ungleichbehandlung führen. Eines der im TGN verwendeten Itemformate sind Kurzantworten. Diese sind notwendigerweise bis zu einem gewissen Grad offen und verlangen Vorwissen. Dies kann zu Verzerrungen führen, wenn der Test international einem sehr gemischten Publikum vorgelegt wird. Die TGN-Entwicklerinnen und Entwickler begegnen diesem Verdacht, indem sie zeigen, dass diese Möglichkeit durch Erprobungen mit drei verschiedenen Gruppen minimiert wurde: potenziellen Testkandidatinnen und kandidaten aus Integrationskursen, Personen mit Migrationshintergrund von ausserhalb dieser Programme sowie Niederländischsprachigen mit

unterschiedlichem sozioökonomischem Hintergrund und aus verschiedenen Altersgruppen, darunter sowohl Männer als auch Frauen. Darüber hinaus wurden auch andere Hypothesen über verschiedene Hintergrundvariablen getestet, welche das Bestehen des Tests übermässig beeinflussen könnten, z.B. das erreichte Bildungsniveau, Alter, Geschlecht, Alphabetisierungsgrad sowie eine übermässige Benachteiligung starker Akzente durch das automatische Bewertungssystem.

Mit Bezug auf den ALTE-Standard 2 – „Beschreiben Sie die [...] Adressaten, für die die Prüfung geeignet ist“ – erwähnen die Autorinnen und Autoren des deutschen DTZ Massnahmen, welche unternommen wurden, um potenzielle Ungleichbehandlungen bei den verschiedenen Zielgruppen zu minimieren, die den Test absolvieren sollen. Diese Gruppen umfassen einerseits Personen mit einem bescheidenen Bildungshintergrund, welche vor allem in der Familie aktiv sein oder eine niedrig qualifizierte Arbeitsstelle finden wollen und andererseits Migrantinnen und Migranten mit erheblicher Lernerfahrung und sogar Diplomen, welche eine Karriere verfolgen wollen. Eine der ergriffenen Massnahmen bezieht sich auf den Testinhalt und das Konstrukt: Es wurden nur Aufgaben aus „Kerndomänen“ des Rahmencurriculums ausgewählt, die vermutlich für *alle* Gruppen relevant sind. Eine andere wichtige Massnahme betrifft die Wahl der Fachpersonen, die bei der Testentwicklung konsultiert wurden, nämlich vor allem Lehrpersonen, Lehrerbildnerinnen und -bildner sowie Lehrwerkautorinnen und -autoren aus dem Bereich Deutsch als Zweitsprache. Ihre Aufgabe war es, darauf zu achten, dass keine der genannten Gruppen aufgrund von unangemessenen Themen und Inhalten, Itemformaten oder Elementen der Testdurchführung benachteiligt werden. Die Hinweise dieser Fachleute wurden zusätzlich mit den Ergebnissen abgeglichen, welche aus statistischen Analysen der Pilotierungen des DTZ hervorgingen.

### 3.8 Konsistenz der Prüfungsergebnisse

Der Grad der Beständigkeit oder Konsistenz

der aus einem Test gewonnenen Informationen über die Fähigkeiten der Teilnehmenden entscheidet über die Zuverlässigkeit der Interpretationen, die darauf basieren. Im Idealfall sollte keine systematische Variation der Daten vorliegen, welche auf Unterschiede in der Testdurchführung, -auswertung oder der Gruppenzugehörigkeit der Kandidatinnen und Kandidaten zurückgeführt werden können. Die früher unter dem Titel der Reliabilität behandelten Problemquellen gehören zu diesem Themenbereich. Es ist daher für Testentwickler unerlässlich zu zeigen, dass ihr Test konsistente Ergebnisse erzeugt. Formelle Tests sind normalerweise zuverlässiger und objektiver als informelle Beurteilungen und, wie in Kapitel 3.5 erwähnt wurde, wissen sogar testskeptische Autorinnen und Autoren deren grössere Fairness zu schätzen (z.B. Balch et al., 2008).

Angesichts der langen Tradition von Reliabilitätsanalysen ist es nicht verwunderlich, dass die mit der Konsistenz der Ergebnisse verbundenen Punkte in De Jong et al. (2009), Perlmann-Balme (2011) und Plassmann (2011) recht umfassend diskutiert werden.

De Jong et al. (2009) stellen ihr automatisiertes Durchführungs- und Bewertungssystem und die dahinterliegende Itemdatenbank als einen wichtigen Faktor bei der Erzeugung konsistenter Testdaten dar. Dieses verhindert im Prinzip, dass Menschen Störfaktoren oder Verzerrungen hineinbringen. Allerdings besteht für die Autorinnen und Autoren des TGN die grosse Herausforderung darin, zu zeigen, dass Testbenutzerinnen und benutzer sich tatsächlich darauf verlassen können, dass das maschinelle Verfahren zuverlässig arbeitet. Ein erheblicher Teil des Artikels beschäftigt sich nur damit. De Jong et al. (2009) zeigen, dass zwischen menschlichen und maschinellen Bewertungen hohe Korrelationen bestehen. Sie zeigen auch, dass die Messung der entscheidenden Fähigkeit auf den niedrigen Niveaus, in denen die Bestehensgrenze angesiedelt ist, mit einem nur kleinen statistischen Fehler möglich ist. Ihre Argumentation wird durch eine beeindruckende Zahl von eingesetzten Testpersonen und items untermauert, welche die Grundlage der Analyse bildeten.

Die deutschen DTZ-Entwicklerinnen und Entwickler ihrerseits folgen in der Präsentation ih-

rer Arbeit zum „Reliabilitätsargument“ (Perlmann-Balme, 2011) den ALTE-Standards.

Der ALTE-Standard 3 zielt auf bewährte Praktiken bei der Testkonstruktion ab:

- 3) Definieren Sie die Anforderungen, die Sie an die Auswahl und das Training von Testkonstruktoren stellen. Beziehen Sie das Urteil von Experten in Form von Gutachten sowohl in die Prüfungsentwicklung als auch in die Revision ein (ALTE, 2007).

Im Bericht über die DTZ-Entwicklung wird das Training der Itemautorinnen und -autoren, sowie der Item-Gutachterinnen und -Gutachter beschrieben, weiter auch die interaktiven und schrittweisen Entwicklungsmethoden, welche Erprobungen (Pilotierungen), Feedback von Lehrpersonen und statistische Analysen umfassten, bevor die eigentlichen Tests mit gutem Gewissen veröffentlicht werden konnten.

Der ALTE-Standard 4 verlangt die Gleichwertigkeit von Tests, die zu verschiedenen Zeitpunkten durchgeführt wurden:

- 4) Stellen Sie sicher, dass die verschiedenen Testsätze einer Prüfung, die an unterschiedlichen Terminen [eingesetzt] werden, vergleichbar sind. Dies betrifft sowohl den Prüfungsinhalt und die Durchführung als auch die Notengebung, die Festsetzung der Bestehensgrenze und die statistischen Werte der Aufgaben (ALTE, 2007).

Perlmann-Balme (2011) nennt die annotierte DTZ-Itemdatenbank als die technische Grundlage für die Vergleichbarkeit des Tests über verschiedene Versionen hinweg, während Prüfungsordnungen und Durchführungsrichtlinien über die verschiedenen Testdurchführungen hinweg die organisatorische Vergleichbarkeit gewährleisten. Das Prüfungshandbuch (Perlmann-Balme et al., 2009) enthält die Transkription einer mündlichen Prüfung und spielt damit eine wichtige Rolle bei der Standardisierung von Durchführung und Bewertung dieses Testteils. Sowohl Perlmann-Balme als auch Plassmann betonen die hohe Qualität des Vorgehens, welches für die Bewertung der schriftlichen Texte entwickelt wurde: Sie werden zentral ausgewertet, was Plassmann zufolge den grossen Vorteil hat, dass Probleme von den zustän-

digen Personen gelöst werden können, sobald sie auftreten. Zur Gewährleistung vergleichbarer Bewertungen wurden drei konkrete Massnahmen ergriffen: Beurteilertraining, Doppelbeurteilung sowie nachträgliche Analyse der Punktevergabe zur Identifizierung stark abweichender Bewertungen.

Der ALTE-Standard 14 verlangt die statistische Analyse von Items:

- 14) Stellen Sie sicher, dass die Erprobungsdaten für die statistischen Analysen (z.B. um die Schwierigkeit und die Trennschärfe der einzelnen Items und die Reliabilität sowie den Messfehler der gesamten Prüfung zu ermitteln) mithilfe von repräsentativen und angemessenen [Kandidatenstichproben] gewonnen werden (ALTE, 2007).

Perlmann-Balme (2011) erwähnt nur kurz, dass die in Standard Nr. 14 erwähnten, üblichen psychometrischen Analysen tatsächlich durchgeführt und auffällige Items aufgrund der Resultate eliminiert werden.

Dieser kurze Abschnitt über die Konsistenz der Prüfungsergebnisse zeigt einmal mehr, dass sich die Publikationen, welche aus den *High-stakes*-Projekten in den Niederlanden und Deutschland hervorgingen, tatsächlich mit zentralen Elementen des üblichen Vorgehens bei der Testvalidierung auseinandersetzen. Der stärker forschungsorientierte Artikel von De Jong et al. (2009) geht dabei anders vor als die Autorinnen, die über den DTZ geschrieben haben. Während De Jong et al. (2009) sich klar bemüht, die Massnahmen zur Sicherstellung der Validität darzustellen, haben Perlmann-Balme (2011) und Plassmann (2011) die Tendenz, lediglich zu deklarieren, dass sie hohe Standards eingehalten und die richtigen Massnahmen angewendet haben. Dieser Unterschied kann möglicherweise vollständig dem Stil und dem Zielpublikum der jeweiligen Zeitschriften zu tun haben, in denen diese Artikel veröffentlicht wurden. Aus Expertensicht wären gründlichere Darstellungen zum DTZ über die existierenden Artikel hinaus allerdings äusserst wünschenswert.

Kapitel 3 untersuchte, wie die in dieser Studie betrachtete Literatur mit Fragen der Testvalidität und Validierung umgeht. Die Validitätsaspekte, die wir uns näher angesehen ha-

# 4

## Zusammenfassung und Schluss

ben, sind Teil des Validierungsmodells *Assessment Use Argument* (AUA) von Bachman & Palmer (2010). Unsere Analysen zeigen, dass die meisten Autorinnen und Autoren nur einen oder sehr wenige Aspekte von Validität und Validierung behandeln. Dabei scheinen vor allem grundlegende Fragen sowie gesellschaftspolitische Betrachtungen mit Bezug auf das Testen im Migrations- und Integrationsbereich von Interesse zu sein. Diese Themen werden aber sogar von modernen Validierungsmodellen, welche, wie das AUA, Fragen der Testverwendung und ihrer Auswirkungen einbeziehen, nur teilweise abgedeckt. Diskussionen über die impliziten Ideologien in Testkonstrukten (z.B. „eine Nation – eine Sprache“) oder Überlegungen zur „Gerechtigkeit“ und Ethik des Testens von Migrantinnen und Migranten gehen über den konzeptuellen Rahmen von Testvalidierungsmodellen – und auch über die Kompetenzen und Erfahrungen von Sprachentestern – hinaus.

Ein Validitätskriterium, die „Aussagekraft der Interpretationen mit Bezug auf ein Konstrukt“, sonst häufig auch als Konstruktvalidität behandelt, hat die Aufmerksamkeit von Testentwicklerinnen und entwicklern, sowie testkritischen Autorinnen und Autoren auf sich gezogen. Während letztere die Verbindung von Testinhalt und (tatsächlichem) Zielsprachengebrauch infrage stellen, bemühen sich erstere, die Aussagekraft ihrer Tests zu zeigen und zu erläutern.

Wir waren überrascht, wie wenige Publikationen aus dem Bereich des Sprachentestens für Migrantinnen und Migranten effektiv als Validierungsstudien gesehen werden können. Wirkliche Validierungsstudien sind entweder an sich selten oder werden der Gruppe der unabhängigen Testexpertinnen und -experten nicht zugänglich gemacht.

Diese Studie gibt einen Überblick über die Entwicklungen der letzten Jahre im Bereich des Sprachentestens im Migrations- und Integrationsbereich, die damit verbundenen Untersuchungen und Diskussionen zur Testvalidität und stellt in diesem Zusammenhang auch die umstrittenen Themen und die vorgebrachten Argumente vor. Auf diesen Informationen basieren die folgenden abschliessenden Bemerkungen und Vorschläge.

Unsere Untersuchung zeigt einige rasante und recht drastische Entwicklungen im Bereich des Sprachentestens für Migrantinnen und Migranten in Europa und anderen, v.a. westlichen, Ländern während der letzten zehn Jahre auf. Der allgemeine Trend geht dabei hin zu sprachlichen Anforderungen auf immer mehr Stufen des Immigrations- und Integrationsprozesses – in einigen Ländern sogar vor Beginn der Immigration – und hin zu umfassenderen Ansprüchen, mit einer sichtbaren Tendenz zu formellen Beurteilungen, oft standardisierten Tests. In einigen Ländern erfassen diese Tests nicht nur sprachliches, sondern auch kulturelles Wissen. In Europa ist es in der Regel so, dass die neuen Ansprüche aufgrund von EU-Regelungen nicht für alle Migrantinnen und Migranten gleichermaßen gelten.

Abgesehen von diesen allgemeinen Tendenzen konnten wir bei der konkreten Konzeptualisierung und Festlegung der Anforderungen und der damit verbundenen Tests grosse Unterschiede feststellen. Die veränderlichen Elemente betreffen den Status von vorbereitenden Sprach- und Integrationskursen, Ausnahmeregelungen für bestimmte Gruppen von Migrantinnen und Migranten, das Mindestalter der betroffenen Personen, Testdesign und Testdurchführung, Testmethoden, die getesteten Fertigkeiten und das erwartete Sprachkompetenzniveau. So bewegen sich die Kompetenzniveaus für die Einreiseerlaubnis zwischen A1 und B1 auf der europäischen Referenzskala, während die erwarteten Niveaus für die Niederlassung oder Einbürgerung

von A1 bis B2 reichen. Dabei ähneln sich allerdings in vielen Fällen die Begründungen für die Wahl der Niveaus.

Das umfassende und weithin bekannte Modell des *Assessment Use Argument* (AUA) von Bachman und Palmer (2010) wurde verwendet, um die Diskussion um Validitätsaspekte in der Literatur zu Sprachtests für Migrantinnen und Migranten zusammenzufassen. Dieses Validierungsmodell ist umfassend insofern, als es alle Aspekte eines einsatzfähigen Tests betrachtet und konzeptuell miteinander verbindet, die mit seiner Validität zu tun haben könnten. Diese reichen von den Testaufgaben und ihrem Einsatz bis zu den Auswirkungen des Tests auf die betroffenen Individuen und die Gesellschaft als Ganzes.

Die verschiedenen Themen, die mit der validen Verwendung von Tests zu tun haben, werden von den für die vorliegende Studie berücksichtigten Publikationen in sehr unterschiedlichem Masse behandelt. Nur sehr wenige Texte zielen explizit darauf ab, die Validität eines spezifischen Tests für Migrantinnen und Migranten umfassend aufzuzeigen (oder systematisch zu hinterfragen). Jene, die sich mit einem spezifischen Instrument beschäftigen, wurden von den jeweiligen Entwicklerinnen und Entwicklern verfasst, um seine Verwendung zu rechtfertigen. Eine dieser Veröffentlichungen (De Jong et al., 2009) unternimmt sehr systematisch den Versuch, die Validität des Tests für gesprochenes Niederländisch (TGN) ausreichend zu belegen, um damit seinen Einsatz zu rechtfertigen. Die Autorinnen und Autoren folgen offensichtlich einem argumentationsbasierten Validierungsmodell wie es in Bachman & Palmer (2010) vorgestellt wird, jedoch ohne dies explizit anzugeben. Die meisten Arbeiten in unserer Auswahl beschäftigen sich entweder mit der *Aussagekraft* der Testinhalte und -anforderungen, den *Auswirkungen* eines Tests auf die betroffenen Personen, oder sie richten so grundsätzliche Fragen an das Testen von Migrantinnen und Migranten *an sich*, dass diese kaum oder gar nicht mehr mithilfe existierender Validierungsmodelle angegangen werden können.

In mehreren Texten wird die *Aussagekraft* der Interpretationen bezweifelt, die auf den Resultaten von Sprach- (und/oder Landeskunde-)Tests basieren. Dies wird begründet mit einer festgestellten oder vermuteten Diskrepanz zwischen

- den Testinhalten und den tatsächlichen Bedürfnissen und Aufgaben,
- dem angegebenen Testkonstrukt (z.B. „ausreichende Sprachkenntnisse für die soziale Integration“) und den tatsächlich getesteten Fähigkeiten und Kompetenzen (welche mehr oder weniger zufällig ermittelt werden),
- der erwarteten Sprachkompetenz für den Test und für das Handeln im Alltag,
- der getesteten Sprache und der oder den Sprachen, welche die Migrantinnen und Migranten in den mehrsprachigen sozialen Kontexten, in denen sie leben und arbeiten, brauchen,
- oder, noch grundlegender, dem Erfolg in Sprach- und/oder Landeskundetests und „guter“ Staatsbürgerschaft.

Wenn (strengere) Testverfahren eingeführt werden, machen die zuständigen Behörden oft geltend, dass die Auswirkungen des Tests weitgehend positiv sind, nicht zuletzt für die Migrantinnen und Migranten selber. Diese Behauptung wird in einer Reihe von Studien behandelt, oft eher anekdotisch, in einigen Fällen aber auch basierend auf tatsächlichen Wirkungsstudien wie der INTEC-Studie (Strik et al., 2010). Dabei dominiert die negative Wahrnehmung des Nutzens von Tests in den untersuchten Arbeiten. Die Pflicht, (strengere) Tests zu bestehen, um den Immigrations- und Integrationsprozess voranzutreiben, wird häufig für nutzlos gehalten, für Migrantinnen und Migranten aufreibend und abschreckend, und wird zudem für das selektive Ausschließen bestimmter Gruppen (z.B. Individuen mit geringer Schulbildung) kritisiert, weil diese dadurch von besserer Integration abgehalten werden. Hingegen erfahren die Vorbereitungskurse, welche in einigen Ländern die Sprach- und Landeskundetests ergänzen, einige Anerkennung: Der hauptsächliche Nutzen dieser Kurse soll in der (verpflichtenden) Gelegenheit liegen, die engen Grenzen der eigenen Gemeinschaft zu verlassen und mit anderen Personen in Kontakt zu treten. Die INTEC-Studie, welche auf statistischen Daten und Interviews aus neun europäischen Ländern basiert, bekräftigt andere, mehr oder weniger informelle Erkenntnisse, wonach die neuen Anforderungen abschreckend wirken und für die

betroffenen Individuen nur wenige Vorteile mit sich bringen.

Kritische Sprachtesterinnen und -tester wie Shohamy und McNamara tragen zur Diskussion über Sprachtests im Migrations- und Integrationsbereich eine radikalere Perspektive bei, indem sie *fundamentale Fragen* stellen (Fragen der „Gerechtigkeit“ gemäss McNamara und seinen Kollegen). Diese berühren mehrere Gebiete ausserhalb des Sprachentestens, darunter (kritische) Sozial- und Politikwissenschaft, Philosophie (Ethik) und Verfassungsrecht. Durch das Aufdecken von impliziten Ideologien („eine Nation – eine Sprache“), verborgenen Absichten („hidden agendas“, Sprachtests als Instrumente gesellschaftlicher Regulierung), politischen Beweggründen (die Erhaltung des gesellschaftlichen Zusammenhalts durch Ausgrenzung) und ethisch problematischen Konsequenzen von Tests (Ausschluss von gesellschaftlicher Teilhabe) stossen diese kritischen Forschenden eine weitreichende Debatte an, bei der verschiedene Disziplinen miteinbezogen werden, welche bisher bei der Entwicklung und Anwendung von Sprach- und Landeskundetests kaum eine Rolle gespielt haben. Nun aber müssten Expertinnen und Experten aus diesen Bereichen lange vor dem eigentlichen Einsatz eines Tests beteiligt werden – sofern überhaupt ein Test eingeführt werden muss.

Welche Schlüsse können aus unseren Beobachtungen gezogen werden? Was kann für die Zukunft empfohlen werden?

Es erscheint wünschenswert, dass die Bedenken der testkritischen Autorinnen und Autoren in die Überlegungen zu Sprachenpolitik und Testverfahren miteinbezogen werden, da ihre Einwände zweifellos fundamentaler Natur sind. Wir glauben allerdings, dass Veränderungen nur angestossen werden können, wenn sich Forschende massgeblich beteiligen, die Disziplinen angehören, in denen man sich normalerweise nicht mit sprachlichen Anforderungen und Tests beschäftigt. Entscheidungen über sprachliche Anforderungen an Migrantinnen und Migranten sowie Immigrations- bzw. Integrationstests werden oft vom Gesetzgeber getroffen und deren Umsetzung von staatlichen Behörden in Auftrag gegeben und überwacht. Diese Instanzen unterliegen nicht zwingend der Pflicht, kritische Wissenschaftlerinnen und

Wissenschaftler hinzuzuziehen, um mit ihnen „Themen von gesellschaftlichem Interesse“ zu diskutieren und zu klären, wie dies von McNamara & Ryan (2011) vorgeschlagen wird. Staaten sind naturgemäss nur dann an diesen Themen interessiert, wenn einklagbare Rechte (potenziell) verletzt werden. Anstelle von staatlichem Handeln ist es also an der akademischen Gemeinschaft, die notwendigen Schritte zu unternehmen, um selbst die Diskussion auf andere Bereiche auszudehnen und die zentralen Probleme herauszukristallisieren. Um genügend Einfluss zu gewinnen, scheint es ausserdem notwendig, die Disziplinen miteinzubeziehen, die sich üblicherweise mit gesellschaftlichen, rechtlichen und ethischen Fragen beschäftigen.

Ein Punkt, welcher sehr deutlich geworden ist, ist der potenzielle Nutzen eines Validierungsmodells wie das *Assessment Use Argument* (AUA) von Bachman & Palmer. Aus professioneller Sicht können Entwicklerinnen und Entwickler von Sprachtests nichts anderes tun als die Rechenschaftspflicht („accountability“) für ihre Arbeit als Richtlinie oder „Axiom“ anzuerkennen (vgl. Bachman & Palmer, 2010). Ein Validierungsmodell operationalisiert diese Rechenschaftspflicht, um eine zeitgemässe Testentwicklung und -verwendung sicherzustellen. Behörden, welche (potenziell) Tests in Auftrag geben, sollten auf die Existenz und Tragweite von modernen Validierungsmodellen wie dem AUA aufmerksam gemacht werden: Dies erlaubt es den zuständigen Personen, angemessene Ausschreibungen für Projektevaluationen zu formulieren und die dafür notwendigen Zeiträume und finanziellen Mittel einzuplanen. Weiterhin machen umfassende Validierungsmodelle, die von der Vorbereitung der Teilnehmenden bis zu den Auswirkungen auf die Gesellschaft alles miteinbeziehen, deutlich, dass Testvalidierung ein andauernder Prozess ist, der die gesamte Lebensdauer des Tests umfasst. Dies ist vor allem für jene Personen von Bedeutung, welche die Testverfahren einführen und für sie verantwortlich sind: Die Vorstellung einer einmaligen Testentwicklung ist nicht haltbar.

Wie bereits erwähnt, ist die Aussagekraft der Interpretationen, welche aufgrund der Testergebnisse formuliert werden, in den vorliegenden Publikationen ein häufig diskutierter Punkt. Von verschiedenen Seiten wird angemerkt, dass

die getesteten Fähigkeiten und die Fähigkeiten, die für das Funktionieren in der Gesellschaft und für aktive Beiträge zugunsten der Gesellschaft nötig sind, wenig miteinander zu tun haben. Die wichtigste Massnahme, um eine bessere Passung zu erreichen – und auch, um die Erwartungen bezüglich der Sprachkenntnisse von Migrantinnen und Migranten zu überwinden, welche auf Ideologien basieren – ist eine gründliche datenbasierte Bedürfnisanalyse<sup>18</sup>. Eine Bedürfnisanalyse war eine der erfolgreich angewendeten Methoden des schweizerischen *fide* Projekts (Müller & Wertenschlag, 2013). Darin wurden nicht nur Migrantinnen und Migranten aus verschiedenen sozialen und ethnischen Gruppen interviewt, sondern auch Mitglieder der Aufnahmegesellschaft, welche regelmässigen Kontakt mit Personen haben, die nur begrenzte Sprachkompetenzen mitbringen. Ein weiterer kritischer Punkt ist die Gewährleistung einer angemessenen Interpretation der Resultate der Bedürfnisanalyse. Diese könnten z.B. zeigen, dass die Notwendigkeit, die Lokalsprache (auf einem bestimmten Niveau) zu beherrschen, in verschiedenen Gruppen sehr unterschiedlich ist, dass es für bestimmte Personen wenig Sinn macht, die Schriftsprache zu erlernen oder dass einige Migrantinnen und Migranten sogar sehr gute Sprachkompetenzen benötigen würden, um ihren Fähigkeiten entsprechend zur Gesellschaft beizutragen. Ein differenzierendes Prüfungs- und Zertifizierungssystem, welches verschiedenen Bedürfnissen und Fähigkeiten Rechnung zu tragen versucht, könnte einen modularen Ansatz wählen. Im Hinblick auf Integration wäre es vermutlich günstig, zunächst ein optimales Sprachförderungssystem zu entwickeln und die Zertifizierung dann darauf aufzubauen. Das von den schweizerischen Behörden unterstützte *fide*-System (Gerber & Schleiss, 2013) bemüht sich, auf die beobachtete Vielfalt zu reagieren, indem es verschiedene Teilnehmerbedürfnisse und -profile in Betracht zieht und Sprachbeurteilung in ein System der

18 | Die sprachliche Bedürfnisanalyse kann zu einer Serie von Sprachaudits erweitert werden, welche innerhalb relevanter Kontexte nicht nur die Bedürfnisse der verschiedenen Stakeholder in Betracht ziehen, sondern auch deren mehrsprachige Sprachkompetenzen.

Sprachförderung einbettet<sup>19</sup>.

Eine weitere Erkenntnis aus unserer Studie ist das Bedürfnis nach mehr Wirkungsstudien, die eine grössere Palette von Aspekten in mehr verschiedenen Kontexten untersuchen. In vielen Artikeln wurden Unsicherheiten bezüglich der tatsächlichen Wirkung von sprachlichen Anforderungen und Sprachtests deutlich. Zum Teil traten gar Spekulationen und anekdotenhafte Belege an die Stelle von fundierten Beweisen. Die INTEC-Studie (Strik et al., 2010) hat mit der Zusammenführung von quantitativen und qualitativen Daten über die Auswirkungen von Testverfahren in neun europäischen Ländern bereits viel erreicht und kleinere Studien ergänzen unser Wissen über andere Kontexte. Diese Bemühungen müssen jedoch fortgesetzt und auf andere Kontexte und Forschungsfragen ausgedehnt werden. Aktuellen Validierungsmodellen wie dem AUA zufolge, welche die Verwendung von Tests als gesellschaftliches Phänomen verstehen und von ihr positive Auswirkungen für alle Beteiligten und die Gesellschaft als Ganzes erwarten, stellen Wirkungsstudien ein wesentliches Element einer umfassenden Testvalidierung dar, das zu einer fundierten Rechtfertigung des Testgebrauchs erheblich beiträgt.

In unserer letzten Empfehlung möchten wir betonen, dass es äusserst wünschenswert ist, dass Tests im Migrations- und Integrationsbereich nicht nur systematischer, sondern auch transparenter validiert werden. Aus unserer Sammlung von Artikeln zielt nur ein einziger (De Jong et al., 2009) tatsächlich darauf ab, eine umfassende Argumentation für die Validität eines Tests darzustellen, indem relevante Punkte angesprochen und dann Argumente und Belege dazu geliefert werden. Es muss allerdings festgehalten werden, dass der Artikel einen konfirmatorischen Charakter hat<sup>20</sup>. Verständlicherweise wurde er erst zu einem Zeitpunkt ver-

19 | Schneider et al. (2006) empfehlen in der gleichen Absicht im Kontext der Einbürgerung die Erstellung zweier Sprachkompetenzprofile: eines Sprachförderungsprofils und eines Sprachbeurteilungsprofils.

20 | Chapelle, Enright, & Jamieson (2008) ist ein weiterer lohnender Bericht, der von (einigen) Testautoren aus einem anderen Bereich, nämlich TOEFL iBT, verfasst wurde. Dabei handelt es sich um einen relativ neuen Test für angehende Studierende an englischsprachigen Universitäten. Die Autoren verwenden ein Validierungsmodell, das auf Kane und Bachman & Palmer basiert.

öffentlich, als der Test bereits einsatzbereit war. Wie Briggs in seinem Kommentar zu Kanes (2004) Aufruf zu einer *expliziten* interpretierenden Argumentation (ähnlich dem AUA von Bachman und Palmer) schreibt, gibt es nämlich in der Testpraxis ein „paradoxes“ Problem, welches nicht leicht zu überwinden ist: Wurde ein Test nicht ausreichend validiert, bevor er eingesetzt wird, dann ist sein Einsatz für Entscheidungen mit grosser Tragweite fragwürdig. Wurde er aber nicht einem Teil der Zielgruppe für den vorgesehenen Zweck vorgelegt, wird es unmöglich, ihn gänzlich zu validieren (Briggs, 2004, S. 172). Aus unserer Sicht befreit dieses Paradoxon die Entwicklerinnen und Entwickler von Tests und jene, die einen Test verwenden und/oder dafür verantwortlich sind, auf keinen Fall davon, Validierungsstudien durchzuführen und sie einem unabhängigen Expertenpublikum vorzulegen. Im besten Fall rechtfertigt es eine leicht verzögerte Veröffentlichung. Ein wünschenswerter nächster Schritt, nun da Validierungsmodelle von guter Qualität und praktischer Handhabung vorliegen, ist die Schaffung einer „Testvalidierungskultur“, in der die Publikation von Validierungsstudien zu einsatzbereiten Tests die Norm, und nicht die Ausnahme sind. Ausserdem sollten, um die Befangenheit der Autorinnen und Autoren auszuschliessen, welche gleichzeitig den Test auch entwickelt haben, externe Testexpertinnen und -experten beauftragt werden, kritische Punkte eines Testsystems (einschliesslich seiner Auswirkungen) zu untersuchen. Zu diesem Zweck ist es notwendig, dass Einsicht in die internen Prozesse und in vertrauliche Daten gewährt wird, wenn dies nötig sein sollte. Das Streben nach Testvalidität, ein hoher Grad an Transparenz und natürlich positive Ergebnisse von Validierungsstudien würden die Glaubwürdigkeit und Legitimität einer Institution, welche Sprachkompetenzen in einem so sensiblen Bereich wie der Immigration testen möchte, erheblich erhöhen. Dies trifft auch auf Behörden zu, die einen Test einsetzen (lassen).

## 5

## Bibliographie

## A

Adami, H., & Leclercq, V. (Hrsg.). (2012). *Les migrants face aux langues des pays d'accueil*. Villeneuve d'Ascq: Presses Universitaires du Septentrion.

Adamo, S. (2008). Northern exposure: the new Danish model of citizenship test. *International Journal on Multicultural Societies*, 10(1), 10–28.

Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal*, 91(4), 659–663. doi:10.2307/4626093

ALTE (Hrsg.). (1994). The ALTE code of practice. Heruntergeladen von [http://www.alte.org/attachments/files/code\\_practice\\_eng.pdf](http://www.alte.org/attachments/files/code_practice_eng.pdf)

ALTE. (2001). Principles of good practice for ALTE examinations. Revised draft. Heruntergeladen von [http://www.alte.org/attachments/files/good\\_practice.pdf](http://www.alte.org/attachments/files/good_practice.pdf)

ALTE (Hrsg.). (2007). 17 Mindeststandards zur Sicherstellung von Qualität in den Prüfungen der ALTE. Heruntergeladen von [http://www.alte.org/attachments/files/minimum\\_standards\\_de.pdf](http://www.alte.org/attachments/files/minimum_standards_de.pdf)

American Educational Research Association, American Psychological Association, National Council on Measurement in Education & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

## B

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2(1), 1–34.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.

Balch, A., Corrigan, M., Gysen, S., Kuijper, H., Perlmann-Balme, M., Roppe, S., ... Zeidler, B. (2008). *Sprachtests für gesellschaftlichen Zusammenhalt und Staatsbürgerschaft – ein Leitfaden für Entscheidungsträger*. Strasbourg: Europarat.

Blackledge, A. (2009a). "As a country we do expect": the further extension of language testing regimes in the United Kingdom. *Language Assessment Quarterly*, 6(1), 6–16. doi:10.1080/15434300802606465

Blackledge, A. (2009b). Being English, speaking English: extension to English language testing legislation and the future of multicultural Britain. In G. Hogan-Brun, C. Mar-Molinero, & P. Stevenson (Hrsg.), *Discourses on language and integration: critical perspectives on language testing regimes in Europe* (S. 83–108). Amsterdam; Philadelphia: John Benjamins Publishing Company.

Blackledge, A. (2009c). Inventing English as convenient fiction: language testing regimes in the United Kingdom. In G. Extra, M. Spotti, & P. Van Avermaet (Hrsg.), *Language testing, migration and citizenship: cross-national perspectives on integration regimes* (S. 66–86). London; New York: Continuum International Publishing Group.

Böcker, A., & Strik, T. (2011). Language and knowledge tests for permanent residence rights: help or hindrance for integration? *European Journal of Migration and Law*, 13(2), 157–184. doi:10.1163/157181611X571268

Brennan, R. L., National Council on Measurement in Education, & American Council on Education (Hrsg.). (2006). *Educational measurement* (4. Auflage). Phoenix, AZ: Greenwood.

Briggs, D. C. (2004). Comment: making an argument for design validity before interpretive validity. *Measurement: Interdisciplinary Research & Perspective*, 2(3), 171–174. doi:10.1207/s15366359mea0203\_2

Buhlmann, R., Ende, K., Kaufmann, S., Kilimann, A., & Schmitz, H. (2007). *Rahmencurriculum für Integrationskurse Deutsch als Zweitsprache*. München: Goethe-Institut. Heruntergeladen von <http://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Kurstraeger/KonzepteLeitfaeden/rahmencurriculum-integrationskurs.pdf>

## C

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Hrsg.). (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York; London: Routledge.

Cooke, M. (2009). Barrier or entitlement? The Language and Citizenship Agenda in the United Kingdom. *Language Assessment Quarterly*, 6(1), 71–77. doi:10.1080/15434300802606580

Cox, L. (2010). The value of values? Debating identity, citizenship and multiculturalism in contemporary Australia. In C. Slade & M. Möllering (Hrsg.), *From migrant to citizen: testing language, testing culture* (S. 77–97). Basingstoke: Palgrave Macmillan.

## D

De Jong, J. H. A. L., Lennig, M., Kerkhoff, A., & Poelmans, P. (2009). Development of a Test of Spoken Dutch for prospective immigrants. *Language Assessment Quarterly*, 6(1), 41–60. doi:10.1080/15434300802606564

## E

Educational Testing Service (ETS). (2002). ETS standards for quality and fairness. Educational Testing Service. Heruntergeladen von <http://www.ets.org/s/about/pdf/standards.pdf>

Europarat (Hrsg.). (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin: Langenscheidt.

Extra, G., & Spotti, M. (2009a). Language, migration and citizenship: a case study on testing regimes in the Netherlands. In G. Hogan-Brun, C. Mar-Molinero, & P. Stevenson (Hrsg.), *Discourses on language and integration: critical perspectives on language testing regimes in Europe* (S. 61–81). Amsterdam; Philadelphia: John Benjamins Publishing.

Extra, G., & Spotti, M. (2009b). Testing regimes for newcomers to the Netherlands. In G. Extra, M. Spotti, & P. Van Avermaet (Hrsg.), *Language testing, migration and citizenship: cross-national perspectives on integration regimes* (S. 125–147). London; New York: Continuum International Publishing Group.

Extra, G., Spotti, M. A. & Van Avermaet, P. (Hrsg.). (2009a). *Language testing, migration and citizenship: cross-national perspectives on integration regimes*. London; New York: Continuum International Publishing Group.

Extra, G., Spotti, M. A. & Van Avermaet, P. (2009b). Testing regimes for newcomers. In G. Extra, M. Spotti, & P. Van Avermaet (Hrsg.), *Language testing, migration and citizenship: Cross-national perspectives on integration regimes* (S. 3–33). London; New York: Continuum International Publishing Group.

Extramania, C. (2012). Les politiques linguistiques concernant les adultes migrants: une perspective européenne. In H. Adami & V. Leclercq (Hrsg.), *Les migrants face aux langues des pays d'accueil: acquisition en milieu naturel et formation* (S. 135–152). Villeneuve d'Ascq: Presses Universitaires du Septentrion.

Extramania, C., & Van Avermaet, P. (2010). *Language requirements for adult migrants in Council of Europe member states: report on a survey*. Strasbourg: Europarat. Heruntergeladen von [http://www.coe.int/t/dg4/linguistic/Source/Mig-ReportSurvey2011\\_EN.doc](http://www.coe.int/t/dg4/linguistic/Source/Mig-ReportSurvey2011_EN.doc)

## F

Farrell, E. (2010). "Do I feel Australian? No you tell me": debating the introduction of the Australian formal citizenship test. In C. Slade & M. Möllering (Hrsg.), *From migrant to citizen: testing language, testing culture* (S. 164–187). Basingstoke: Palgrave Macmillan.

Fulcher, G. (2004). Deluded by Artifices? The Common European Framework and harmonization. *Language Assessment Quarterly*, 1(4), 253–266. doi:10.1207/s15434311laq0104\_4

## G

Gerber, A., & Schleiss, M. (2013). Langue et intégration: une responsabilité partagée / Sprache und Integration: eine gemeinsame Verantwortung. *Babylonia*, (1), 9–12.

Goodman, S. W. (2011). Controlling immigration through language and country knowledge requirements. *West European Politics*, 34(2), 235–255. doi:10.1080/01402382.2011.546569

## H

Hargreaves, M. (2010). Citizenship testing in the anglophone countries: the UK, Canada and the USA. In C. Slade & M. Möllering (Hrsg.), *From migrant to citizen: testing language, testing culture* (S. 101–124). Basingstoke: Palgrave Macmillan.

Hogan-Brun, G., Mar-Molinero, C., & Stevenson, P. (Hrsg.). (2009a). *Discourses on language and integration: critical perspectives on language testing regimes in Europe*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Hogan-Brun, G., Mar-Molinero, C., & Stevenson, P. (2009b). Testing regimes: introducing cross-national perspectives on language, migration and citizenship. In G. Hogan-Brun, C. Mar-Molinero, & P. Stevenson (Hrsg.), *Discourses on language and integration*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Holland, A. (2010). Australian citizenship in the twenty-first century: historical perspectives. In C. Slade & M. Möllering (Hrsg.), *From migrant to citizen: testing language, testing culture* (S. 39–59). Basingstoke: Palgrave Macmillan.

## I

ILTA (International Language Testing Association). (2007). Guidelines for practice. Heruntergeladen von [http://www.iltaonline.com/images/pdfs/ILTA\\_Guidelines.pdf](http://www.iltaonline.com/images/pdfs/ILTA_Guidelines.pdf)

## J

Joppke, C. (2010). How liberal are citizenship tests? *EUI Working Papers*, 41, 1–4.

## K

Kane, M. (2004). Certification testing as an illustration of argument-based validation. *Measurement: Interdisciplinary Research & Perspective*, 2(3), 135–170. doi:10.1207/s15366359mea0203\_1

- Kane, M. (2006). Validation. In R. L. Brennan, National Council on Measurement in Education, & American Council on Education (Hrsg.), *Educational measurement* (4. Auflage, S. 17–64). Phoenix, AZ: Greenwood.
- Kane, M., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, 18(2), 5–17. doi:10.1111/j.1745-3992.1999.tb00010.x
- Kiwan, D. (2008). A journey to citizenship in the United Kingdom. *IJMS: International Journal on Multicultural Societies*, 10(1), 60–74.
- Klein, G. (2013). Do gender, age and first language predict the results in the Deutsch-Test für Zuwanderer (DTZ)? In E. Dimitrova-Galaczi & C. J. Weir (Hrsg.), *Exploring language frameworks: proceedings of the ALTE Kraków Conference, July 2011* (S. 389–404). Cambridge: Cambridge University Press.
- Kostakopoulo, D. (2010). Introduction. In R. van Oers, E. Ersbøll, & D. Kostakopoulo (Hrsg.), *A re-definition of belonging? Language and integration tests in Europe* (S. 1–23). Leiden; Boston: Martinus Nijhoff Publishers.
- Krumm, H.-J. (2007). Profiles instead of levels: the CEFR and its (ab)uses in the context of migration. *The Modern Language Journal*, 91(4), 667–669. doi: 10.1111/j.1540-4781.2007.00627\_6.x
- Kunnan, A. J. (2009). Testing for citizenship: The U.S. Naturalization Test. *Language Assessment Quarterly*, 6(1), 89–97. doi:10.1080/15434300802606630
- Kunnan, A. J. (2012). Language assessment for immigration and citizenship. In G. Fulcher & F. Davidson (Hrsg.), *The Routledge handbook of language testing* (S. 162–177). New York: Routledge.
- L
- Laversuch, I. M. (2008). Putting Germany's language tests to the test: an examination of the development, implementation and efficacy of using language proficiency tests to mediate German citizenship. *Current Issues in Language Planning*, 9(3), 282–298. doi:10.1080/14664200802139554
- Lee, J. (2010). Amendment to the Naturalization Examination and its social impact on international marriage immigrants in South Korea. *TESOL Quarterly*, 44(3), 575–585. doi:10.5054/tq.2010.232866
- Lenz, P., Andrey, S., & Lindt-Bangerter, B. (2009). Rahmencurriculum für die sprachliche Förderung von Migrantinnen und Migranten. Bundesamt für Migration BFM. Heruntergeladen von <http://www.bfm.admin.ch/content/dam/data/migration/integration/berichte/rahmencurriculum-d.pdf>
- Linn, R. L., National Council on Measurement in Education, & American Council on Education (Hrsg.). (1989). *Educational measurement*. New York; London: Macmillan Publishers.
- Little, D. (2010). The linguistic integration of adult migrants: evaluating policy and practice. Council of Europe, Language Policy Division. Heruntergeladen von [http://www.coe.int/t/dg4/linguistic/liam/Source Events/2010/2010evaluatingpolicy\\_EN.pdf](http://www.coe.int/t/dg4/linguistic/liam/Source%20Events/2010/2010evaluatingpolicy_EN.pdf)
- M
- Mackenzie, C. (2010). Citizenship, identity, and immigration: Contemporary philosophical perspectives. In C. Slade & M. Möllering (Hrsg.), *From migrant to citizen: testing language, testing culture* (S. 191–216). Basingstoke: Palgrave Macmillan.

- Mar-Molinero, C. (2006). The European linguistic legacy in a global era: linguistic imperialism, Spanish and the Instituto Cervantes. In C. Mar-Molinero & P. Stevenson (Hrsg.), *Language ideologies, policies and practices: language and the future of Europe* (S. 76–88). Basingstoke: Palgrave Macmillan.
- Mar-Molinero, C., & Stevenson, P. (Hrsg.). (2006). *Language ideologies, policies and practices: language and the future of Europe*. Basingstoke: Palgrave Macmillan.
- McNamara, T. (2005). 21st Century Shibboleth: language tests, identity and intergroup conflict. *Language Policy*, 4(4), 351–370. doi:10.1007/s10993-005-2886-0
- McNamara, T. (2006). Validity in language testing: the challenge of Sam Messick's legacy. *Language Assessment Quarterly: An International Journal*, 3(1), 31–51. doi:10.1207/s15434311laq0301\_3
- McNamara, T. (2009a). Australia: the Dictation Test redux? *Language Assessment Quarterly*, 6(1), 106–111. doi:10.1080/15434300802606663
- McNamara, T. (2009b). Language tests and social policy: a commentary. In G. Hogan-Brun, C. Mar-Molinero, & P. Stevenson (Hrsg.), *Discourse on language and integration* (S. 153–163). Amsterdam; Philadelphia: John Benjamins Publishing Company.
- McNamara, T. (2010). The use of language tests in the service of policy: issues of validity. *Revue Française de Linguistique Appliquée*, 15, 7–23.
- McNamara, T., & Roever, C. (2006). *Language testing: the social dimension*. Malden, MA: Blackwell Publishing.
- McNamara, T., & Ryan, K. (2011). Fairness versus justice in language testing: the place of English literacy in the Australian Citizenship Test. *Language Assessment Quarterly*, 8(2), 161–178. doi:10.1080/15434303.2011.565438
- McNamara, T., & Shohamy, E. (2008). Language tests and human rights. *International Journal of Applied Linguistics*, 18(1), 89–95. doi:10.1111/j.1473-4192.2008.00191.x
- Messick, S. (1989). Validity. In R. L. Linn, National Council on Measurement in Education, & American Council on Education (Hrsg.), *Educational measurement* (3. Auflage, S. 13–103). New York; London: American Council on Education; Macmillan Pub. Co.; Collier Macmillan Publishers.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23. doi:10.2307/1176219
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. doi:10.1177/026553229601300302
- Michalowski, I. (2010). Integration tests in Germany. A communitarian approach? In R. van Oers, E. Esrboll, & D. Kostakopoulo (Hrsg.), *A re-definition of belonging? Language and integration tests in Europe* (S. 185–210). Leiden; Boston: Martinus Nijhoff Publishers.
- Michalowski, I. (2011). Required to assimilate? The content of citizenship tests in five countries. *Citizenship Studies*, 15(6-7), 749–768. doi:10.1080/13621025.2011.600116
- Milani, T. M. (2008). Language testing and citizenship: a language ideological debate in Sweden. *Language in Society*, 37(01), 27–59. doi: 10.1017/S0047404508080020
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to Evidence-Centered Design. CSE Report 632*. Los Angeles, CA: CRESST, CSE; UCLA. Heruntergeladen von <http://www.education.umd.edu/EDMS/mislevy/papers/BriefIntroECD.pdf>
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing*, 19(4), 477–496.

Möllering, C. (2010). The changing scope of German citizenship: from “guest worker” to citizen? In C. Slade & M. Möllering (Hrsg.), *From migrant to citizen: testing language, testing culture* (S. 145–163). Basingstoke: Palgrave Macmillan.

Möllering, M., & Silaghi, L. (2010). From earning the privilege of citizenship to understanding its responsibilities: An update on Australian citizenship testing. In C. Slade & M. Möllering (Hrsg.), *From migrant to citizen: testing language, testing culture* (S. 236–255). Basingstoke: Palgrave Macmillan.

Müller, M., & Wertenschlag, L. (2013). „Meine Kinder möchten, dass ich auch zum Elternabend gehe“: Anmerkungen zum Szenarienansatz und zur Entstehungsgeschichte der fide-Szenarien. *Babylonia*, (1), 28–34.

## N

North, B. (2009). The educational and social impact of the CEFR in Europe and beyond: a preliminary overview. In L. Taylor & C. J. Weir (Hrsg.), *Language testing matters: investigating the wider social and educational impact of assessment - proceedings of the ALTE Cambridge Conference April 2008* (S. 357–378). Cambridge: Cambridge University Press.

## O

Oers, R. van. (2008). From liberal to restrictive citizenship policies: the case of the Netherlands. *International Journal on Multicultural Societies*, 10(1), 40–59.

Oers, R. van. (2010). Citizenship tests in the Netherlands, Germany and the UK. In R. van Oers, E. Ersbøll, & D. Kostakopoulo (Hrsg.), *A re-definition of belonging? Language and integration tests in Europe* (S. 49–105). Leiden; Boston: Martinus Nijhoff Publishers.

Oers, R. van, Ersbøll, E., & Kostakopoulo, D. (2010). Mapping the redefinition of belonging in Europe. In R. Van Oers, E. Ersbøll, & D. Kostakopoulo (Hrsg.), *A re-definition of belonging? Language and integration tests in Europe* (S. 307–331). Leiden; Boston: Martinus Nijhoff Publishers.

Oers, R. van, Ersbøll, E., & Kostakopoulou, D. (Hrsg.). (2010). *A re-definition of belonging? Language and integration tests in Europe*. Leiden; Boston: Martinus Nijhoff Publishers.

Ozolins, U. (2003). The impact of European accession upon language policy in the Baltic states. *Language Policy*, 2(3), 217–238.

## P

Papp, S. (2010). The requirements of the UK test for citizenship and settlement: critical issues and possible solutions. In L. Taylor & C. J. Weir (Hrsg.), *Language testing matters: investigating the wider social and educational impact of assessment - proceedings of the ALTE Cambridge Conference April 2008* (S. 118–135). Cambridge: Cambridge University Press.

Perlmann-Balme, M. (2011). Deutsch-Test für Zuwanderer. Internationale Qualitätsstandards bei der Testentwicklung. *Deutsch Als Fremdsprache*, (1), 13–22.

Perlmann-Balme, M., Plassmann, S., & Zeidler, B. (2009). *Deutschtest für Zuwanderer A2-B1 Prüfungsziele, Testbeschreibung*. Berlin: Cornelsen. Heruntergeladen von [http://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Kurstraeger/Sonstiges/dtz-handbuch\\_pdf.pdf](http://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Kurstraeger/Sonstiges/dtz-handbuch_pdf.pdf)

Piller, I. (2001). Naturalization language testing and its basis in ideologies of national identity and citizenship. *International Journal of Bilingualism*, 5(3), 259–277. doi:10.1177/13670069010050030201

Plassmann, S. (2011). *Aktuelle Methoden der Testmethodik und Qualitätssicherung am Beispiel des Deutsch-Tests für Zuwanderer. Deutsch als Fremdsprache*, (1), 23–29.

## S

Saville, N. (2009). Language assessment in the management of international migration: a framework for considering the issues. *Language Assessment Quarterly*, 6(1), 17–29. doi:10.1080/15434300802606499

Saville, N., & Van Avermaet, P. (2008). Language testing for migration and citizenship: contexts and issues. In L. B. Taylor, C. J. Weir, & ALTE Conference (Hrsg.), *Multilingualism and assessment: achieving transparency, assuring quality, sustaining diversity: proceedings of the ALTE Berlin conference, May 2005* (S. 265–275). Cambridge: Cambridge University Press.

Schneider, G., Neuner-Anfindsen, S., Sauter, P., Studer, T., Wertenschlag, L., & Widmer, C. (2006). *Rahmenkonzept für den Nachweis der sprachlichen Kommunikationsfähigkeit im Hinblick auf die Einbürgerung, Kurzbericht (erstellt im Auftrag der Eidgenössischen Ausländerkommission (EKA)). Fribourg: Lern- und Forschungszentrum Fremdsprachen, Universität Fribourg. Heruntergeladen von [http://www.ekm.admin.ch/de/dokumentation/doku/kurzbericht\\_rahmenkonzept.pdf](http://www.ekm.admin.ch/de/dokumentation/doku/kurzbericht_rahmenkonzept.pdf)*

Shohamy, E. (1998). Critical language testing and beyond. *Studies In Educational Evaluation*, 24(4), 331–345. doi:10.1016/S0191-491X(98)00020-0

Shohamy, E. (2001). *The power of tests: a critical perspective on the uses of language tests*. Harlow, England; New York: Longman.

Shohamy, E. (2006). *Language policy: hidden agendas and new approaches*. London und New York: Routledge.

Shohamy, E. (2007). Tests as power tools: looking back, looking forward. In J. Fox, M. Wesche, & D. Bayliss (Hrsg.), *Language testing reconsidered* (S. 141–152). Ottawa: University of Ottawa Press.

Shohamy, E. (2009). Language tests for immigrants: Why language? Why tests? Why citizenship? In G. Hogan-Brun, C. Mar-Molinero, & P. Stevenson (Hrsg.), *Discourses on language and integration: critical perspectives on language testing regimes in Europe*. (S. 45–59). Amsterdam: John Benjamins Publishing Company.

Siiner, M. (2006). planning language practice: a sociolinguistic analysis of language policy in post-communist Estonia. *Language Policy*, 5(2), 161–186. doi:10.1007/s10993-006-9004-9

Skenderovic, D. (2013). *Einwanderung und Sprache: Kulturalisierung einer Debatte. Baby-lonia*, (1), 14–18.

Slade, C. (2010a). Civic integration in the Netherlands. In C. Slade & M. Möllering (Hrsg.), *From migrant to citizen: testing language, testing culture* (S. 125–144). Basingstoke: Palgrave Macmillan.

Slade, C. (2010b). Shifting langscapes of citizenship. In C. Slade & M. Möllering (Hrsg.), (S. 3–23). Basingstoke: Palgrave Macmillan.

Slade, C., & Mollering, M. (2010). *From migrant to citizen: testing language, testing culture*. Basingstoke: Palgrave Macmillan.

Stevenson, P., & Schanze, L. (2009). Language, migration and citizenship in Germany: discourses on integration and belonging. In G. Extra, M. Spotti, & P. Van Avermaet (Hrsg.), *Language testing, migration and citizenship: cross-national perspectives on integration regimes* (S. 87–106). London; New York: Continuum International Publishing Group.

Strik, T., Böcker, A., Luiten, M., & Oers, R. van. (2010). *The INTEC Project: synthesis report. Integration and naturalisation tests: the new way to European Citizenship*. Nijmegen: Centre for Migration Law, Radboud University Nijmegen.

## V

Van Avermaet, P. (2009). Fortress Europe? Language policy regimes for immigration and citizenship. In G. Hogan-Brun, C. Mar-Molinero, & P. Stevenson (Hrsg.), *discourse on language and integration* (S. 15–43). Amsterdam; Philadelphia: John Benjamins Publishing Company.

Van Avermaet, P. (2012). L'intégration linguistique en Europe: analyse critique. In H. Adami & V. Leclercq (Hrsg.), *Les migrants face aux langues des pays d'accueil: acquisition en milieu naturel et formation* (S. 153–171). Ville-neuve d'Ascq: Presses Universitaires du Septentrion.

Van Avermaet, P., & Rocca, L. (2013). Language testing and access. In E. Galaczi & C. Weir (Hrsg.), *Exploring language frameworks: proceedings of the ALTE Kraków Conference, July 2011* (S. 11–44). Cambridge: Cambridge University Press.

## W

Wright, S. (Hrsg.). (2008). Citizenship tests in Europe - editorial introduction. *International Journal on Multicultural Societies*, 10(1), 1–9.

## Y

Yoffe, L. (2010). Linguistic integration of adult migrants in France: policy and practice following the introduction of the reception and integration contract (CAI). *The Cultural Review* - 早稲田商学同友会, 36(3), 171–188.





