

Les prérequis linguistiques et l'usage de tests de langue à des fins d'immigration et d'intégration:

—

une synthèse de la littérature académique

—

Evelyne Pochon-Berger
Peter Lenz

2014

—

Rapport du Centre scientifique de compétence sur le plurilinguisme

Impressum

Publié par

Institut de plurilinguisme
www.institut-plurilinguisme.ch

Auteurs

Evelyne Pochon-Berger
Peter Lenz

Traduction

Evelyne Pochon-Berger

Le projet dont il est question a été financé par la Confédération suisse dans le cadre du programme de travail 2012-2014 du Centre scientifique de compétence sur le plurilinguisme. La responsabilité du contenu de la présente publication incombe à ses auteurs.

Fribourg, mars 2014

Layout

Billy Ben, Graphic Design Studio

Les prérequis linguistiques et l'usage de tests de langue à des fins d'immigration et d'intégration:

—

une synthèse de la littérature académique

—

Evelyne Pochon-Berger

Peter Lenz

2014

—

Rapport du Centre scientifique de compétence sur le plurilinguisme

Résumé

Ce document est le résultat d'une étude de la littérature sur les développements récents en matière de prérequis linguistiques et d'usage de tests de langue à des fins d'immigration et d'intégration. L'étude a été menée dans le cadre du programme de recherche 2011-2014 du Centre scientifique de compétence sur le plurilinguisme (CSP¹) en Suisse.

L'introduction de prérequis linguistiques et l'usage de tests de langue formels comme conditions pour l'obtention d'un visa d'entrée, d'un permis de séjour et/ou de la nationalité sont des pratiques plutôt récentes, mais évoluant rapidement dans beaucoup de pays occidentaux et en Europe en particulier. Ces prérequis et formes d'évaluations sont devenus l'objet de débats dans les milieux académiques et de la recherche et le sujet d'un certain nombre de publications scientifiques. Ce rapport vise à offrir une vue d'ensemble structurée des différents aspects de ces prérequis et formes d'évaluations en se basant sur un corpus de publications scientifiques.

Le résumé de la littérature présenté dans ce document se divise en deux chapitres principaux. Le premier offre une vue d'ensemble des développements récents, qui se sont produits surtout au cours de la dernière décennie, en matière de prérequis linguistiques pour l'immigration, le séjour permanent et la naturalisation; il décrit également l'introduction récente d'instruments d'évaluation formelle aux différentes étapes du processus d'immigration. Dans la mesure où l'information présentée est souvent basée sur des prises de positions d'experts critiques quant aux nouvelles tendances, ce chapitre aborde également différentes problématiques dans une perspective critique. Le deuxième chapitre traite de manière systématique les thèmes émergents du débat sur l'usage de tests de langue à des fins d'immigration et d'intégration par le biais des principales catégories utilisées dans la recherche en matière de validation de l'évaluation. Le modèle de

Bachman & Palmer (2010) dénommé *Assessment Use Argument* (AUA) constitue l'arrière-fond conceptuel pour notre analyse.

Notre étude conclut avec une série de recommandations pour la recherche future dans le domaine de l'évaluation, les publications et les pratiques en matière de validation des tests.

En général, nos résultats confirment la perception subjective d'une augmentation des prérequis linguistiques et de l'usage de tests de langue formels (parfois, ceux-ci étant complétés par un test de connaissances culturelles) à des fins d'immigration et d'intégration dans différents pays 'occidentaux', en particulier en Europe, au cours de la dernière décennie. Bien qu'une tendance générale commune se dessine très clairement, nous constatons également une grande diversité dans la définition concrète des prérequis et des instruments d'évaluations y afférents, par exemple, en ce qui concerne les niveaux de langue requis.

L'analyse de la littérature à travers le modèle de validation de Bachman et Palmer révèle que les différents thèmes sont traités de manière très inégale au sein de la littérature publiée. En particulier, seuls quelques publications visent réellement à démontrer systématiquement (ou à remettre en question) la validité de tests spécifiques et leur usage. La plupart des publications traitent soit de la pertinence (*meaningfulness*) des contenus des tests et des niveaux en ce qui concerne le savoir et les capacités choisies pour l'évaluation (typiquement 'une connaissance suffisante pour s'intégrer dans la société') et les retombées positives (*beneficence*) d'un test et de la phase préparatoire pour les différents acteurs, soit elles interrogent l'usage en soi de tests pour l'immigration qui est une question si fondamentale qu'elle n'est traitée que de manière marginale par les modèles actuels de validation.

Dans notre conclusion, nous formulons des recommandations pour la recherche à venir en rapport avec les points suivants: l'inclusion de disciplines spécialisées dans les problématiques sociales, éthiques et légales en lien avec l'usage de tests de langue à des fins d'immigration et d'intégration, une obligation de rendre des comptes et de soumettre les instruments

1 | Les auteurs voudraient remercier Séverine Beaud Duarte Rodrigues pour son importante contribution lors du repérage et de l'examen des publications scientifiques dans la phase initiale du projet, ainsi que son relecteur interne, Thomas Studer, pour ses judicieux commentaires à différents moments du projet.

d'évaluation à une validation systématique au moyen d'un modèle de validation actualisé, le recours à l'analyse des besoins (*needs analysis*) comme base pour la conception des tests et leur usage, la nécessité pour plus d'études d'impact (*impact studies*) sur les régimes de test, le souhait pour plus de transparence en ce qui concerne la validation des tests et de leurs résultats.

Table des matières

1	Introduction	5
1.1	Le contexte suisse	5
1.2	Objectif et structure de ce rapport	5
1.3	Sources	6
2	L'évaluation linguistique à des fins d'immigration et d'intégration dans un contexte en mutation	8
2.1	Politiques linguistiques et pratiques d'évaluation émergentes	8
2.2	L'implémentation de politiques d'immigration plus strictes en matière de langues	9
2.2.1	Prolifération des prérequis linguistiques	10
2.2.2	Augmentation de l'usage de tests de langue standardisés	11
2.2.3	Niveaux de langue requis	12
2.2.4	Les tests de langue et de connaissances culturelles et civiques	12
3	Des questions de validité dans les tests de langue à des fins d'immigration et d'intégration	13
3.1	L'approche argumentative de Bachman et Palmer pour la validation de l'évaluation	13
3.2	Justice et évaluation linguistique à des fins d'immigration et d'intégration	16
3.3	Les retombées positives de l'usage des instruments d'évaluation	20
3.4	Décisions sensibles aux valeurs	24
3.5	Décisions équitables	24
3.6	Pertinence des conclusions relatives au construct	26
3.7	Impartialité des conclusions pour tous les groupes de candidats	31
3.8	Cohérence des résultats du test	32
4	Résumé et conclusions	34
5	Bibliographie	38

1

Introduction

1.1

Le contexte suisse

Dans de nombreux pays occidentaux, particulièrement en Europe, les autorités ont de plus en plus souvent recours à des tests de langues pour prendre des décisions en matière de permis de séjour. La Suisse n'est pas au premier plan de ce développement mais des efforts récents des organes législatifs et administratifs à différents niveaux du système politique indiquent une tendance allant dans cette direction. Au niveau fédéral, les problématiques linguistiques sont traitées comme un aspect de l'intégration. En suivant la formule 'fördern und fordern' ('encourager et exiger') qui guide la politique d'intégration, les autorités ont lancé un plan directeur pour l'intégration en 2010 qui stipule un encouragement linguistique accru pour les locuteurs de langues qui ne sont pas l'une des langues nationales suisses (Gerber & Schleiss, 2013). Dans un premier temps, un certain nombre d'instruments ont été développés pour encourager et soutenir l'enseignement et l'apprentissage des langues, tels qu'un curriculum-cadre pour les cours de langue de 'niveau-seuil faible' (Lenz et al., 2009), une spécification des objectifs orientés vers les besoins des apprenants et basés sur des scénarios d'utilisation de la langue (Müller & Wertenschlag, 2013) ou encore du matériel pour le placement des apprenants ayant des besoins spécifiques, pour la planification des cours, pour l'enseignement et l'(auto-)évaluation. Par ailleurs, un programme de workshops a été mis sur pied pour faciliter la dissémination des nouveaux instruments.

Depuis 2008, la Loi fédérale sur les étrangers a permis aux autorités de définir des prérequis linguistiques (en général, il s'agit de la fréquentation d'un cours de langue) dans le cadre d'une 'convention d'intégration' pour les candidats à une autorisation de séjour (Skenđerovic, 2013). Des instruments spécifiques

d'évaluation linguistique (y inclus des tests²) pour les autorisations de séjour³ n'existent pas encore. Au moment de la rédaction du présent rapport (2013), alors que la révision sur la Loi sur les étrangers est dans le processus législatif, la partie 'exiger' de la formule se trouve de plus en plus au centre de l'attention. Il est fort probable que dès 2015 certains groupes de migrants devront fournir des preuves de leur maîtrise d'une langue nationale, sans doute en passant soit un test officiellement reconnu ou une évaluation liée à un cours contrôlé par les autorités.

1.2

Objectif et structure de ce rapport

L'objectif principal de la présente étude est de présenter une vue d'ensemble structurée de différents aspects relatifs à l'évaluation linguistique à des fins d'immigration et d'intégration sur la base de publications scientifiques. Ces aspects sont :

- les développements récents en matière de prérequis linguistiques et de tests de langue formels à des fins d'immigration et d'intégration ;
- les problématiques en jeu et les arguments avancés dans les discussions essentiellement critiques qui portent sur ces développements ;
- la validité de l'évaluation linguistique et des tests de langue à des fins d'immigration et d'intégration.

En établissant cette vue d'ensemble, nous espérons identifier non seulement des problématiques actuelles et les principales tendances, mais aussi mettre en évidence des lacunes

2 | 'Evaluation' (*assessment*) est un terme plus général que 'test'. Il couvre les tests ainsi que d'autres activités permettant d'établir le degré de connaissances ou de compétences. Lorsque l'évaluation prend la forme d'un test, les deux termes sont utilisés de manière interchangeable dans le présent texte.

3 | L'accès à la nationalité est une problématique différente, dans la mesure où les communes ont (encore) le droit de l'octroyer ou non et de définir les prérequis linguistiques et les procédures d'évaluation. Les différences d'une commune à une autre sont considérables.

dans la recherche et les débats actuels dans les milieux scientifiques. Les informations ainsi recueillies constituent la base sur laquelle sont formulées des recommandations en vue de poursuivre le travail conceptuel, la recherche et le développement dans ce domaine relativement délicat.

La présente étude a une structure simple : le chapitre 1 pose l'arrière-fond de l'étude, à savoir le contexte local, l'objectif et la structure de l'étude, ainsi que les sources publiées disponibles. Le chapitre 2 informe sur les développements de cette dernière décennie en ce qui concerne les prérequis linguistiques pour les migrants ; il décrit également l'introduction récente d'instruments d'évaluation formels aux différentes étapes du processus d'immigration (immigration, séjour permanent, nationalité). Dans la mesure où ces informations se basent sur des comptes-rendus d'experts s'opposant pour la plupart à l'introduction de conditions d'immigration plus strictes, le chapitre 2 offre également une vue d'ensemble structurée des divers aspects qui font l'objet de critiques. Le chapitre 3 commence par esquisser un cadre largement accepté pour la validation des tests. Les catégories principales de ce modèle sont ensuite utilisées pour présenter la discussion actuelle sur l'usage de tests de langue à des fins d'immigration et d'intégration. Enfin, le chapitre 4 offre une synthèse des résultats des chapitres précédents, formule des conclusions et esquisse des pistes pour les travaux futurs.

1.3 Sources

L'usage de prérequis linguistiques et de tests de langue formels comme conditions pour l'octroi d'un visa d'entrée, d'une autorisation de séjour ou de la nationalité est un développement assez récent, mais dont l'évolution est rapide. En conséquence, l'évaluation linguistique à des fins d'immigration et d'intégration est à son tour un champ de recherche scientifique et un objet de débat récent, mais croissant, qui a attiré l'attention de nombreux experts. Les études qui ont paru jusqu'à ce jour sont principalement des descriptions critiques

et des analyses des politiques d'immigration en lien avec les questions linguistiques et les régimes de test qui ont été introduits dans divers pays. Les questions d'équité des tests (*test fairness*) et de justice sociale en rapport avec ces nouvelles politiques et régimes ont été largement abordées. Au cours des dix dernières années, de nombreux volumes collectifs portant sur ces questions ont été publiés, tels que Adami & Leclercq (2012), Hogan-Brun et al. (2009a), Mar-Molinero & Stevenson (2006), Oers et al. (2010), Slade & Mollering (2010). Par ailleurs, quelques numéros spéciaux dans des revues scientifiques ont été dédiés à ces problématiques, p. ex. *International Journal on Multicultural Studies*, vol. 10 (1) 2008 ; *Language Assessment Quarterly*, vol. 6 (1), 2009. Un article de Kunnan (2012) dans le cadre d'un *handbook* offre une synthèse récente sur ce sujet. Bien que la plupart des publications accessibles sont jusqu'ici encore essentiellement programmatiques, elles ont néanmoins contribué à ouvrir ce champ à la recherche et ont permis une prise de conscience chez les linguistes et les concepteurs de tests de ces développements récents et des problématiques qu'ils soulèvent. Cependant, des études empiriques publiées sont à ce jour encore rares et très peu d'études portent sur la conception et la validation de tests de langue déjà existants. Les articles de De Jong et al. (2009), Perlmann-Balme (2011) et Plassmann (2011) sont des exceptions notables. L'article de De Jong et al. a une orientation de recherche différente, alors que les deux autres articles sont accessibles à un public plus large. Les trois articles ont en commun le fait qu'ils visent plus à *confirmer* – en représentant ainsi la perspective des concepteurs de tests et leurs intérêts – qu'à prendre un point de vue critique.

Du fait que l'utilisation de tests de langue à des fins d'immigration et d'intégration constitue une préoccupation pour les gouvernements et les institutions intergouvernementales (telles que le Conseil de l'Europe) et qu'il s'agit également d'un sujet présent dans le discours public et politique, il n'est pas surprenant de trouver sur cette problématique des rapports émanant de mandats officiels (p. ex. Schneider et al., 2006 : des recommandations concernant

l'utilisation de tests de langue dans le cadre de demandes de naturalisation, sur un mandat de la Commission fédérale pour les questions de migration ; Balch et al., 2008 ; Extramania & Van Avermaet, 2010 : les résultats de deux enquêtes mandatées par le Conseil de l'Europe ; Little, 2010 : un rapport portant sur l'intégration linguistique de migrants adultes mandaté par le Conseil de l'Europe). Notre étude se concentre sur les articles publiés dans des revues ou des ouvrages scientifiques ; seule une petite sélection de rapports mandatés, tels que les rapports mentionnés précédemment, sont consultés de manière ponctuelle.

La littérature à disposition est plutôt limitée dans sa portée géographique. Elle couvre surtout les politiques et les pratiques des pays européens (membres de l'Union Européenne et non-membres) et quelques rares autres pays, dont notamment les Etats-Unis, le Canada et l'Australie. Les publications se sont concentrées sur ces pays en partie parce que ceux-ci ont été confrontés à d'importantes vagues d'immigration. Par ailleurs, bien que certains pays aient une longue tradition d'immigration (p. ex. les Etats-Unis, le Canada, l'Australie et certains pays d'Europe occidentale), pour d'autres, le phénomène est plus récent (p. ex. les pays du Sud de l'Europe, tels que l'Espagne, ou d'autres pays d'Europe occidentale) (Extra et al., 2009). D'après Van Avermaet (2009), les changements drastiques dans les politiques d'immigration se sont produits en réponse à l'immigration croissante. Les pays européens sont particulièrement touchés par ce développement récent, car, dans ce cas précis, l'immigration vers l'Europe vient s'ajouter à la mobilité à l'intérieur de l'Union européenne, qui, elle, est déjà importante. Le contexte européen est par conséquent un objet propice pour la recherche. La grande diversité dans les pratiques et les politiques – malgré les tendances communes – que l'on trouve à travers les pays européens a contribué à en faire un objet de recherche. Des études sont disponibles sur des pays pionniers en matière d'utilisation de tests de langue pour la naturalisation – les Pays-Bas, l'Allemagne, le Danemark et le Royaume-Uni (voir p. ex. Blackledge, 2009a et b ; Extra & Spotti, 2009a et b ; Laversuch, 2008 ; Michalowski, 2010 ; Möllering,

2010 ; Slade, 2010a). Les pratiques dans les pays d'outre-mer à tradition d'immigration, tels que les Etats-Unis, le Canada et l'Australie, sont bien décrites (p. ex. Cox, 2010 ; Farrell, 2010 ; Hargreaves, 2010 ; Holland, 2010 ; McNamara, 2009a). Des informations sur des pays européens tels que l'Autriche, la Belgique, le Luxembourg, l'Espagne et les pays baltes sont plutôt rares, alors que les pratiques dans d'autres pays sont même largement sous-explorées (p. ex. Grèce, Italie, Pologne, Portugal). Enfin, nous ne savons presque rien sur les pratiques dans des pays non-occidentaux (p. ex. les pays asiatiques ou africains, mais Lee, 2010, sur le contexte coréen constitue une exception notable).

Dans le présent rapport, nous nous consacrerons essentiellement aux publications portant sur les pays européens.

2

L'évaluation linguistique à des fins d'immigration et d'intégration dans un contexte en mutation

2.1 Politiques linguistiques et pratiques d'évaluation émergentes

Des débats politiques animés sur le sujet de l'évaluation linguistique et des tests de langue à des fins d'immigration et d'intégration ont vu le jour dans de nombreux pays ; ces débats ont donné lieu à des changements dans les législations nationales. Les publications suivantes offrent une vue d'ensemble : Blackledge (2009a & b) pour le contexte britannique, Slade (2010a) pour le contexte néerlandais, Möllering (2010) pour le contexte allemand, Farrell (2010) pour l'Australie. A un niveau supranational, Balch et al. (2008) formulent des recommandations pour l'usage des tests de langue pour la naturalisation et Extramania & Van Avermaet (2010) font une synthèse des politiques linguistiques en matière d'immigration dans un certain nombre de pays européens. Un troisième rapport mandaté par le Conseil de l'Europe, et qui devrait paraître en 2013, se base sur une enquête à grande échelle des politiques d'intégration linguistique des migrants dans les pays membres qui a été menée en 2013.

Certaines études font apparaître un changement majeur dans les politiques nationales d'immigration dans de nombreux pays qui s'est produit au cours des dix dernières années (principalement en Europe, mais aussi en Australie, au Canada et aux USA). Ce changement se manifeste dans 'l'augmentation dans le nombre et

l'étendue des mécanismes de régulation'⁴ (Saville, 2009). Autrement dit, il y a une tendance générale à implémenter des mesures d'immigration et d'intégration de plus en plus strictes (Hogan-Brun et al., 2009b; Slade, 2010a; Van Avermaet, 2009), ce qui traduit une approche 'assimilationniste' de l'immigration (Möllering, 2010; Slade, 2010a). Certains pays ont même mis en place des programmes d'intégration obligatoires fonctionnant sur un système de sanctions (Kostakopoulo, 2010). D'après Böcker & Strik (2011), cette tendance vers des politiques d'immigration de plus en plus strictes dans de nombreux pays européens contraste fortement avec les politiques antérieures que ces pays avaient adoptées dans les années 1980-1990. Les politiques précédentes visaient l'encouragement de l'immigration et de l'intégration au moyen de mesures facilitées pour l'obtention du permis de séjour et de la nationalité. La législation néerlandaise, par exemple, était autrefois libérale à cet égard, mais les Pays-Bas sont devenus le premier pays européen à adopter des conditions d'immigration et d'intégration strictes (Slade, 2010a). Depuis ce revirement législatif, l'octroi d'un permis de séjour ou de la nationalité est considéré comme une 'récompense' pour ceux qui ont réussi à surmonter un certain nombre d'obstacles et qui, par conséquent, méritent d'appartenir à la communauté (Adamo, 2008; Böcker & Strik, 2011; Kostakopoulo, 2010). Auparavant, ces démarches étaient considérées avant tout comme un moyen encourageant l'intégration et moins comme une 'récompense'.

Hogan-Brun et al. (2009b) situent ce changement dans les politiques linguistiques pour l'immigration en Europe aux alentours de 2004, en le reliant à l'extension de l'UE aux anciens pays soviétiques, ce qui a engendré un flux migratoire massif des pays de l'UE de l'Est vers les pays de l'UE de l'Ouest. Certains auteurs, cependant, expliquent cette croissance par d'autres facteurs économiques et politiques (p. ex. Hargreaves, 2010; Saville, 2009; Slade, 2010a; Van Avermaet, 2009): globalisation, immigration de main-d'œuvre, éducation, tour-

4 | Toutes les citations dans ce rapport ont été traduites en français, sauf si indication contraire.

isme et persécution politique. L'augmentation de la multiculturalité et du multilinguisme dans les pays d'accueil qui en découle a souvent été perçue comme une menace pour l'identité nationale, et ceci, même dans des pays tels que l'Australie qui se sont construits grâce à l'immigration (Slade, 2010a; Wright, 2008). D'après Hogan-Brun et al. (2009b), cet effet s'observe particulièrement dans les pays membres de l'UE où les processus d'unification politique et économique en cours ont engendré en soi une crise des identités nationales. Il en résulte que de nombreux états ont réagi par l'adoption de politiques d'immigration plus strictes dans le but de préserver l'identité nationale et la cohésion sociale. La xénophobie montante dans le courant des années 2000, alimentée par les attaques terroristes, est considérée comme un autre facteur ayant mené à des politiques d'immigration et d'intégration plus strictes pour renforcer 'la sécurité nationale' (Blackledge, 2009c; Laversuch, 2008; Van Avermaet, 2009; Wright, 2008).

2.2 L'implémentation de politiques d'immigration plus strictes en matière de langues

Les changements dans les politiques linguistiques ont donné lieu à une remarquable diversité en termes de prérequis linguistiques dans les différents pays. Cette diversité s'observe dans les méthodes, la conception et l'administration des tests, les compétences testées, le niveau de langue requis, l'âge minimal du candidat, les frais, l'exemption de certains groupes, ainsi que le statut des cours préparatoires de langue et d'intégration (obligatoire vs. volontaire). Cette diversité est soulignée par des études récentes sur les régimes de tests dans les pays européens (voir Böcker &

Strik, 2011⁵; Extramania & Van Avermaet, 2010; Van Avermaet, 2009). Sur la base de données recueillies auprès de 17 pays, Van Avermaet (2009) observe que le niveau de langue requis peut varier de A1 à B1 pour ce qui est de l'admission dans le pays d'accueil, de A1 à B2 pour les demandes de permis de séjour et de A2 à B2 pour la naturalisation. Ces données révèlent des différences flagrantes en matière de prérequis linguistiques d'un pays à un autre. De manière semblable, se basant sur la comparaison de pays anglophones (les Etats-Unis, le Canada et le Royaume-Uni), Hargreaves (2010) identifie des différences significatives entre les trois pays tant en ce qui concerne les formats de tests utilisés pour la naturalisation que leur contenu. Le Royaume-Uni et le Canada, par exemple, ont recours à un test à choix multiples (administré par ordinateur dans le premier cas et sur papier dans le deuxième cas), alors qu'aux Etats-Unis, un entretien informel est mené avec le candidat. Certains auteurs (Hargreaves, 2010; Hogan-Brun et al., 2009b; McNamara, 2009) stipulent qu'une telle variation dans les pratiques et les régimes de test est due en partie au fait que ces régimes de test sont mis en place par les gouvernements (au lieu d'un groupe d'experts) en réponse au contexte sociopolitique propre à chaque état et en fonction de la manière dont ils perçoivent l'attitude de la population. Ces études témoignent de pratiques qui évoluent rapidement, mais qui sont très peu coordonnées, en particulier en Europe.

Les sections suivantes mettent en évidence les mesures prises dans le but d'implémenter des conditions d'immigration et d'intégration plus strictes en ce qui concerne la ou les langue(s) du pays d'accueil : en premier lieu, l'introduction récente ou le renouvellement de

5 | L'article de Böcker & Strik (2011) expose uniquement les résultats concernant le cas spécifique de demandes de séjour permanent; l'étude plus large (INTEC, voir Strik et al., 2010) porte sur les différentes étapes du processus d'immigration : entrée dans le pays d'accueil, séjour permanent et naturalisation. Cette étude n'offre pas seulement une vue d'ensemble détaillée des régimes de test dans chaque pays analysé, mais présente également des données recueillies en ce qui concerne l'impact des tests de langue par le biais d'entretiens informels avec différents acteurs dans le but d'identifier leur point de vue sur les pratiques et leurs conséquences.

prérequis linguistiques pour l'immigration dans différents pays, ainsi que l'extension de ces critères aux différentes étapes du processus d'immigration; en deuxième lieu, le recours croissant à des tests de langue standardisés dans le but d'appliquer ces prérequis; en troisième lieu, le niveau de langue requis; en quatrième lieu, les tests de connaissances civiques et de la société et les prérequis implicites en termes de connaissances linguistiques.

2.2.1 Prolifération des prérequis linguistiques

Les prérequis linguistiques sont devenus un aspect central des politiques d'immigration et d'intégration dans de nombreux pays considérés dans ce rapport. Non seulement un nombre plus élevé de pays a introduit de tels prérequis, mais les prérequis linguistiques ont également été étendus aux différentes étapes du 'parcours du migrant' (Saville, 2009).

Plusieurs études signalent une forte augmentation des prérequis linguistiques qui s'accompagne parfois par des taxes plus élevées dans le processus d'immigration et par des sanctions pour ceux qui ne remplissent pas les conditions (Böcker & Strik, 2011; Hogan-Brun et al., 2009b; Van Avermaet, 2009). Sur la base des informations données dans différents rapports officiels, Extramania (2012) identifie un pic dans l'introduction de prérequis linguistiques dans les pays d'Europe occidentale (Danemark, Belgique, Allemagne, Grèce, Norvège, Autriche, Pays-Bas, France et Liechtenstein) entre les années 2003 et 2008⁶. Pour ce qui est des pays d'Europe de l'Est (Slovaquie, Arménie, Slo-

vénie, Pologne et Turquie), un tel pic s'observe entre les années 2007 et 2009. Extramania (2012) indique par ailleurs que la majorité des pays ayant introduit des prérequis linguistiques à des fins d'immigration et d'intégration ont également mis en place des cours de langues pour les candidats.

De manière intéressante, lorsque certains pays mettent en place des politiques linguistiques et les régimes de tests plus stricts, ceux-ci tendent à être suivis par les autres pays. Par exemple, lorsque le gouvernement néerlandais a introduit des prérequis linguistiques pour l'octroi du permis de séjour et l'admission dans le pays en 2006, cinq autres états membres de l'UE ont suivi le modèle néerlandais en adoptant des conditions semblables la même année (Böcker & Strik, 2011; Van Avermaet, 2009).

De plus, les études mettent en évidence une tendance des gouvernements à étendre progressivement les prérequis linguistiques aux différentes étapes du processus d'immigration. Initialement, les prérequis linguistiques ont été introduits par la plupart des pays dans le cadre des demandes de naturalisation (d'après Böcker & Strik, 2011, cette pratique est devenue courante dans les années 1990). Par la suite, elles ont été imposées aux individus souhaitant obtenir un permis de séjour permanent. Se basant sur les chiffres émanant de deux enquêtes mandatées par le Conseil de l'Europe et portant sur les politiques linguistiques à des fins d'immigration dans un certain nombre de pays européens, Extramania (2012) observe une augmentation considérable entre 2007 et 2009 dans la proportion des pays imposant des prérequis linguistiques comme condition pour le séjour permanent: tandis qu'en 2007, 57% des pays ayant participé à l'enquête ont des prérequis linguistiques, ce chiffre s'est élevé à 69% en 2009.

Plus récemment, et dans des cas plus rares, des prérequis linguistiques ont été établis comme condition pour l'admission dans le pays (c.-à-d. pour l'obtention d'un visa d'entrée), à l'instar des Pays-Bas (Extra & Spotti, 2009b) ou de l'Allemagne (Michalowski, 2010). Extramania (2012) signale une augmentation allant de 19% des pays ayant participé à l'enquête qui appliquent une telle condition en 2007 à 26% en

6 | Mar-Molinero (2006) mentionne l'Espagne comme une exception remarquable à cette tendance générale. La nationalité est en effet difficilement accessible en Espagne: le taux de naturalisation y est très faible, les prérequis linguistiques pour le séjour permanent et la nationalité sont inexistantes. L'Espagne n'impose pas de tests de langue explicites, mais un entretien avec le juge permet d'évaluer indirectement l'adaptation du candidat à la culture locale. L'absence de prérequis linguistiques (connaissances d'une langue 'nationale') pour la naturalisation peut s'expliquer par l'histoire du pays (qui implique d'éviter d'éventuelles réactions des minorités linguistiques dans les Communidades autónomas).

2009. Selon Michalowski (2010), cette mesure restrictive a été introduite aux Pays-Bas et en Allemagne en réponse à l'augmentation de l'immigration familiale. Ainsi, l'obligation de passer un test de langue avant l'arrivée dans le pays devient un moyen indirect pour limiter l'immigration familiale.

Les prérequis linguistiques comme condition préalable à l'entrée dans le pays font l'objet de commentaires critiques (Goodman, 2011 ; Kostakopoulo, 2010) qui se basent sur l'argument selon lequel le processus d'intégration ne peut commencer que lorsqu'un individu se trouve physiquement dans un pays et qu'il a la possibilité d'interagir avec la communauté d'accueil. Ainsi, Goodman (2011) souligne :

Les conditions d'intégration préalables à l'entrée dans le pays imposent un degré d'intégration au sein de l'état tandis que le candidat est physiquement et conceptuellement – par rapport à son statut légal – *en-dehors* de l'état (p. 237).

De plus, ainsi que Extra & Spotti (2009a) et Hogan-Brun et al. (2009b) le soulignent, l'accès à la langue cible tout en vivant à l'étranger peut être difficile (p. ex. manque de cours de langue). Dans un tel cas de figure, il est attendu des candidats de connaître une langue qu'ils n'ont pas eu l'opportunité d'apprendre et de pratiquer – ou du moins pas dans un environnement propice à cela.

2.2.2 Augmentation de l'usage de tests de langue standardisés

Se corrélant avec la prolifération des prérequis linguistiques dans les politiques d'immigration européennes, le recours à des tests de langue formels est devenu de plus en plus courant (Van Avermaet, 2009). Chacune des trois étapes de l'immigration, à savoir, admission, séjour et nationalité, est concernée. Selon Van Avermaet (2009), qui rend compte d'une étude menée en 2007 par ALTE (Association des concepteurs de tests en Europe) auprès de 19 pays européens, neuf pays sur onze ayant des prérequis linguistiques pour la naturalisation, ont implémenté un test de langue ; de manière semblable, six pays sur huit imposant des prérequis linguistiques pour le séjour permanent ont également institué un test de langue pour ceux qui souhaitent obtenir ce statut. Les résultats pour ce qui est de l'admission dans le pays sont similaires : sept pays sur neuf ayant des prérequis linguistiques pour cette étape ont également recours à un test de langue formel. En d'autres termes, presque tous les pays ayant introduit des prérequis linguistiques dans leur politique d'immigration ont recours à un test formel lorsqu'il s'agit de déterminer si un candidat a acquis le niveau de langue visé.

Dans certains cas, des tests standardisés viennent remplacer d'autres formes d'évaluation pratiquées jusqu'alors (Hogan-Brun et al., 2009b). Böcker & Strik (2011) citent des cas de procédures de naturalisation plus anciennes aux Pays-Bas et au Danemark où les compétences linguistiques étaient d'abord évaluées de manière informelle au cours d'un entretien avec un fonctionnaire d'état. Par la suite, au début des années 2000, ces pays ont introduit des tests standardisés pour évaluer les connaissances linguistiques des candidats à la naturalisation.

2.2.3 Niveaux de langue requis

Les niveaux de langue requis pour l'admission dans le pays, pour le séjour permanent et pour la nationalité peuvent varier énormément d'un pays à l'autre. La vue d'ensemble de Van Avermaet (2009) sur les régimes de test dans 17 pays européens montre que les niveaux de langue requis pour l'admission varient entre A1.1 et B1 sur l'échelle du Cadre européen commun de référence pour les langues (CECR – Conseil de l'Europe, 2001); pour le *séjour permanent*, les niveaux vont de A1 à B2, et pour la *nationalité*, de A2 à B2. Se référant à l'Allemagne où le niveau B1 est requis pour la nationalité, Möllering (2010) note que ce niveau de langue relativement élevé peut décourager les individus à faire une demande de naturalisation et suggère que le niveau A2, tel qu'il est requis aux Pays-Bas, serait plus adapté.

Les niveaux de langue requis ne sont pas toujours décrits en rapport avec une échelle de niveaux telle que celle du CECR (Conseil de l'Europe, 2001). La législation canadienne, par exemple, exige 'une capacité à parler l'anglais ou le français suffisamment bien pour communiquer avec la population' (Hargreaves, 2010); le test de naturalisation australien (*Australian Citizenship Test*) exige 'des connaissances élémentaires d'anglais' (Farrell, 2010; Möllering & Silaghi, 2010). A propos de ce test, le *Woolcott Report* (voir McNamara & Ryan, 2011; Möllering & Silaghi, 2010) critique la nature très vague de la définition qui est donnée du niveau de langue requis – dans la mesure où il n'est pas très clair ce qui doit être maîtrisé – et en dénoncent le manque de transparence; cela rend plus difficile pour les candidats de savoir comment remplir les conditions pour obtenir la nationalité.

Böcker & Strik (2011) attirent l'attention sur le fait que les niveaux de langue requis pour le séjour permanent et la nationalité tendent à s'élever lorsque des prérequis linguistiques sont également posés comme condition préalable à l'entrée dans le pays. En d'autres termes, l'introduction d'un prérequis linguistique à un stade initial du processus d'immigration et d'intégration a apparemment pour effet d'élever la barre pour les stades ultérieurs.

2.2.4 Les tests de langue et de connaissances culturelles et civiques

Dans certains pays, comme le Royaume-Uni, le Danemark ou les Pays-Bas, les candidats pour le séjour permanent et la nationalité doivent passer un examen de connaissances culturelles et civiques en plus d'un test de langue (Böcker & Strik, 2011; Van Avermaet, 2009). Dans les tests de connaissances culturelles et civiques, les candidats sont évalués sur des thèmes divers, tels que la religion, la culture, le droit et l'histoire du pays d'accueil. Dans des tâches telles que celles visant la connaissance et la compréhension du Serment d'allégeance australien (*Pledge of Allegiance*) (Slade, 2010a), Möllering (2010) identifie une approche assimilationniste de la naturalisation. Selon Slade (2010a), le choix d'utiliser de tels tests se base typiquement sur la présupposition que les candidats vont adhérer aux valeurs de la société d'accueil par l'acquisition d'un savoir sur les faits socioculturels, les normes et les valeurs de cette société – une croyance qui peut être difficilement soutenue selon l'auteure.

Deux autres problèmes relatifs aux tests de connaissances de la société sont soulevés dans la littérature analysée. Premièrement, bien que les questions semblent porter sur des connaissances purement factuelles, certaines d'entre elles portent en réalité sur les attitudes et les croyances personnelles des candidats. Ce fait est jugé inéquitable (Joppke, 2010a; Michalowski, 2011) et peut possiblement constituer une source de discrimination sociale. Deuxièmement, dans la mesure où le test de connaissances de la société est souvent effectué dans la (ou l'une des) langue(s) officielle(s) de la société d'accueil, la réussite dépend par conséquent des compétences linguistiques du candidat. McNamara (2009b) considère de tels tests de connaissances de la société comme étant *de facto* des tests de langue. Extra & Spotti (2009a) discutent de cette problématique à l'exemple du test néerlandais de langue et de culture. McNamara & Ryan (2011) décrivent la controverse qui a accompagné l'introduction du nouveau test australien de naturalisation; des objections ont été émises, car les connais-

3

Des questions de validité dans les tests de langue à des fins d'immigration et d'intégration

sances linguistiques nécessaires pour réussir le test dépassent (même après révision, elles dépassent toujours) le niveau requis officiellement pour la naturalisation.

En somme, les travaux évoqués dans le chapitre 2 prennent une position critique sur les développements récents concernant les prérequis linguistiques pour l'immigration. Ils attestent d'une tendance dans les pays occidentaux, en particulier en Europe, au cours de ces dix dernières années, à l'introduction de prérequis linguistiques plus stricts pour les différents étapes de l'immigration. De plus, une forte tendance allant d'une forme d'évaluation linguistique informelle vers des instruments d'évaluation formels est constatée. Alors que ces tendances générales se dessinent assez clairement, les études témoignent d'une forte diversité d'un pays à un autre en ce qui concerne les niveaux de langues requis. Les tests portant sur les connaissances culturelles et civiques sont généralement critiqués dans ces études, non seulement en raison du type de connaissances qui est évalué mais également parce que le niveau de langue implicite pour le réussir dépasse les niveaux de langue requis.

Ce chapitre présente comment et dans quelle mesure les aspects de validité de l'évaluation sont discutés dans la littérature sur l'évaluation linguistique à des fins d'immigration et d'intégration. La qualité et l'adéquation de l'instrument d'évaluation (formel) sont généralement établies au moyen d'un modèle de validité ou de validation qui guide les discussions et les analyses à visée confirmative ou critique. La présentation des questions de validité est organisée en fonction des catégories prises d'un modèle de validation qui est à la fois exhaustif et largement reconnu par les professionnels des tests de langue: le modèle de Bachman et Palmer pour la validité et la validation de l'évaluation, appelé *Assessment Use Argument* (AUA). Dans un premier temps, les éléments essentiels de ce cadre sont brièvement présentés. Par la suite, des problématiques provenant de la littérature portant sur la validité sont mis en évidence en utilisant les catégories du cadre AUA ainsi que quelques catégories additionnelles qui s'avèrent nécessaires.

3.1

L'approche argumentative de Bachman et Palmer pour la validation de l'évaluation

La validité de l'évaluation a été un sujet récurrent d'une importance centrale en matière de tests dans le domaine de la psychologie, de l'éducation et des langues depuis plusieurs décennies. Le modèle *Assessment Use Argument* (AUA) de Bachman et Palmer a été récemment décrit et opérationnalisé en détail dans Bach-

man & Palmer (2010). Il se fonde sur deux types de sources : a) des travaux théoriques et pratiques qui ont été menés par différents auteurs⁷, y inclut des associations reconnues dans le domaine⁸; b) les travaux antérieurs de Bachman et Palmer sur les tests de langue se basant sur des tâches communicatives (Bachman, 1990; Bachman, 2005; Bachman & Palmer, 1996). Le chapitre 5 de Bachman & Palmer (2010) offre une vue d'ensemble concise du modèle AUA.

La raison principale de l'usage d'un modèle de validation tel que le AUA réside dans le fait que les évaluations linguistiques sont typiquement utilisées pour des décisions qui entraînent des conséquences pour les différents acteurs (p. ex. les candidats, les enseignants, les employeurs ou la société plus généralement). Dans la mesure où les individus sont directement concernés par les décisions se basant sur des évaluations, Bachman & Palmer posent une obligation⁹ pour les concepteurs de tests et leurs utilisateurs (p. ex. les autorités en charge des questions d'immigration) de rendre des comptes aux acteurs concernés. 'Rendre des comptes' signifie qu'ils 'doivent démontrer, au moyen d'un argumentaire et de données à l'appui, que le recours à un instrument d'évaluation particulier est justifié' (B&P¹⁰: 85-86). Une fois établi, l'argumentaire consiste en une série d'affirmations de validité bien-fondées qui sont interconnectées (par des

inférences) dans le but de constituer un argumentaire de validité exhaustif pour des usages spécifiques d'un instrument d'évaluation. Les critères habituels de qualité des instruments d'évaluation (formels), tels que la fiabilité des tests et la fiabilité inter-individuelle, absence de biais de groupe ou validité des contenus font partie intégrante de ces affirmations et de l'argumentaire de validité dans son ensemble (voir les indications entre crochets dans la liste ci-dessous). La tâche principale qui incombe aux concepteurs de tests et/ou aux utilisateurs est de démontrer que les différentes affirmations sont respectées à un degré suffisant, en fonction du but de l'instrument d'évaluation et de ses conséquences. Un AUA valide justifie à la fois une méthode spécifique de test et l'usage du test pour un objectif particulier dans la société.

Dans ce qui suit, les différentes affirmations qui devraient être démontrées au moyen du AUA sont présentées très succinctement.

Affirmation 1 – les conséquences visées

Les *conséquences* de l'usage d'un instrument d'évaluation et les décisions qui sont prises sont **bénéfiques** pour les acteurs (candidats, enseignants, société dans son ensemble) (B&P: 105).

Affirmation 2 – les décisions

Les décisions prises sur la base de conclusions à partir des résultats du test...

- prennent en compte les **valeurs** de la communauté et les exigences légales pertinentes et
- sont **équitable**s pour les acteurs qui sont concernés par la décision [absence de biais] (B&P: 111).

Affirmation 3 – les conclusions

Les conclusions en ce qui concerne la compétence évaluée sont...

- **pertinentes** au regard d'un syllabus d'apprentissage spécifique, d'une analyse des capacités requises pour réaliser des tâches

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

8 | Les 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; les codes de pratiques : ILTA (International Language Testing Association), 2007; ALTE [Association of Language Testers in Europe] (1994); ALTE (Association of Language Testers in Europe), 2001; les manuels : 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 introduisent la nécessité de rendre des comptes aux individus concernés et l'obligation de démontrer que l'usage de l'évaluation est justifié comme 'deux axiomes fondamentaux' de leur approche (Bachman & Palmer, 2010: 85).

10 | Abréviation pour Bachman & Palmer (2010).

dans le domaine du TLU¹¹, d'une théorie générale de la compétence langagière ou d'une combinaison de ces aspects [validité de construct] ;

- **impartiales** pour l'ensemble des groupes de candidats [absence de biais] ;
- **généralisables** au domaine du TLU pour lequel la décision est prise [validité de construct] ;
- **importantes** pour les décisions qui vont être prises [validité de construct] et
- **suffisantes** pour les décisions qui vont être prises [validité de construct] (B&P : 111).

Affirmation 4 – les résultats du test

Les *résultats du test* (notes, descriptions des performances des candidats) sont **cohérents** à travers différentes tâches d'évaluation, différents aspects de la procédure d'évaluation (p. ex. formats du test, passations, examinateurs) et à travers les différents groupes de candidats [fiabilité, absence de biais] (B&P : 124).

Ainsi qu'il l'a été mentionné plus haut, une validation de test exhaustive et la mise à disposition de preuves de validité convaincantes sont hautement souhaitables dans tout contexte dont les enjeux sociaux sont élevés, tel que l'octroi d'un visa d'entrée, d'un permis de séjour permanent ou de la nationalité. En dépit de cela, la vue d'ensemble de la littérature sur l'évaluation linguistique à des fins d'immigration et d'intégration présentée dans cette publication fait apparaître d'emblée qu'un argumentaire de validité exhaustif est très rarement appliqué. Bon nombre de travaux adressent seulement quelques aspects sélectifs de la validité, à savoir, des aspects relatifs aux affirmations 1 (les retombées positives - *beneficence*), 2 (les valeurs et l'équité - *values and equitability*) et 3 (la pertinence - *meaningfulness*) ; ces aspects ont fait l'objet de discussion plus souvent que les autres. Cependant, ces discussions sont souvent d'ordre général et ne se rapportent pas à un instrument d'évaluation spécifique et à son usage.

11 | TLU : usage de la langue-cible (*Target-Language Use*), c'est-à-dire l'usage dans des situations de communication authentique.

Seules quelques rares publications visent à produire un argumentaire de validité exhaustif ou une analyse d'une série de caractéristiques cruciales d'un test particulier et de ses usages spécifiques. Cela n'est pas étonnant, dans la mesure où il existe relativement peu de publications se concentrant sur un test particulier dans le domaine des tests de langues à des fins d'immigration et d'intégration.

L'objectif des sections suivantes est de rendre compte des débats dans la littérature sur les tests de langue à des fins d'immigration et d'intégration en rapport avec chacune des catégories principales (à savoir, les affirmations de validité) utilisées dans le modèle AUA. Cela signifie que la présente publication ne traite pas des argumentaires de validité pour des instruments d'évaluation spécifiques mais traite plutôt des problématiques soulevées de manière générale à travers les contextes, dans leur rapport à chacun des aspects de la validité du modèle AUA.

Par ailleurs, le système des catégories du modèle AUA doit être modifié de façon à intégrer les études qui traitent des tests de langue à des fins d'immigration et d'intégration depuis un angle de vue plus radical et qui soulèvent la question de savoir si les tests de langue, pour diverses raisons, sont *justes*, c'est-à-dire s'ils constituent *en soi* un moyen légitime dans le contexte de l'immigration.

McNamara et ses collègues (p. ex. McNamara, 2006; McNamara & Roever, 2006; McNamara & Ryan, 2011) ont contribué de manière considérable à la discussion concernant le caractère *juste* de l'usage d'un instrument d'évaluation. Ils se positionnent comme appartenant au courant du *Critical Language Testing* (p. ex. Shohamy, 2001; Shohamy, 2006), lequel se concentre sur les relations implicites et explicites entre langage et pouvoir. En parallèle, ils s'inspirent des travaux fondateurs de Messick sur la validité (Messick, 1989; Messick, 1996; McNamara, 2006), lesquels mettent l'accent sur l'usage (de l'instrument d'évaluation) en tant que dimension constitutive de la validité.

Dans une publication récente sur la problématique de la *justice* de l'évaluation linguistique, McNamara & Ryan (2011) présentent les concepts d'équité (*fairness*) et de justice

comme étant complémentaires et tous les deux nécessaires à la validation des tests (de langues). La notion d'*équité*, telle qu'ils la définissent, n'implique pas seulement un traitement égalitaire des groupes et la prévention de biais psychométriques, mais également tous les aspects de la validation empirique des conclusions à partir des résultats dans l'intérêt d'une évaluation raisonnable et défendable des candidats au test. La problématique de la *justice* des tests complète cette approche en y ajoutant une prise en compte des conséquences de l'interprétation des résultats et leur usage (ou abus) dans un contexte social spécifique¹², ainsi que les valeurs sociales et politiques implicites dans les constructs sur lesquels se fondent les tests.

Sans doute, la notion d'*équité* dans son ensemble et certains aspects de la notion de *justice* sont également couverts par le modèle du *Assessment Use Argument* de Bachman et Palmer. Bachman (2005) et Bachman & Palmer (2010) illustrent brièvement comment l'abus dans l'usage de tests mis en évidence par le *Critical Language Testing* (voir notamment Shohamy, 2001) pourrait être intégré dans le cadre du modèle AUA. Ils argumentent que de nombreux abus potentiels mentionnés pourraient être contrecarrés par l'établissement de critères de qualité traitant de 'l'apport bénéfique des conséquences' pour les différents acteurs.

Cependant, dans leur discussion au sujet de Messick (1989), McNamara et Ryan (2011) ne sont pas d'accord avec cet aspect et soulignent que l'évaluation linguistique ne peut pas être auto-suffisante pour le traitement des 'questions de valeur sociale'. Selon eux, l'évaluation linguistique doit aussi

se fonder sur d'autres types d'analyse, plus habituelles dans le domaine de la *Cultural Analysis* et *Critical Policy Analysis*, lesquelles prennent généralement en compte les aspects historiques, idéologiques et le contexte discursif et lesquelles ont recours principalement à des outils analytiques qualitatifs (McNamara & Ryan, 2011: 164).

12 | McNamara & Ryan (2011) évoquent les 'agendas cachés' selon les termes de Shohamy, c'est-à-dire des tests de langue utilisés à des fins politiques – comme étant un aspect de la question de la *justice* des tests.

3.2 Justice et évaluation linguistique à des fins d'immigration et d'intégration

Cette section offre un compte rendu de la discussion sur la 'question des valeurs sociales' (McNamara & Ryan, 2011) dans la littérature sur l'évaluation linguistique à des fins d'immigration et d'intégration. Elle se concentre sur deux domaines critiqués par le *Critical Language Testing* et lesquels s'articulent partiellement :

1. les valeurs sociales et politiques, ainsi que les idéologies implicites dans les constructs sur lesquels se fondent les tests ;
2. la régulation sociale (*gate-keeping*) au moyen de l'évaluation linguistique se basant très souvent sur un 'agenda caché'.

Une grande partie des études consultées identifie et critiquent les idéologies et les valeurs incontestées sur la base desquelles est justifiée l'introduction de prérequis linguistiques (et de connaissances culturelles) et leur évaluation formelle.

L'une de ces idéologies est le mythe du langage comme symbole de l'identité nationale et d'appartenance à une communauté (Extra et al., 2009 ; Piller, 2001 ; Shohamy, 2009 ; Van Avermaet, 2012). Se basant sur une analyse de textes et de discours sur le sujet de l'immigration dans la sphère publique suédoise, Milani (2008) identifie la croyance récurrente qu'un manque de maîtrise du suédois est la cause principale d'un engagement faible des migrants dans la vie politique, sociale et économique. Ainsi, la maîtrise de la langue nationale est perçue comme une condition indispensable au 'bon fonctionnement' dans la société d'accueil et une condition pour la compréhension des dimensions culturelles de la communauté. Selon Milani, il existe la présupposition

[...] que la connaissance d'une langue commune (le suédois) ne donne pas simplement accès au domaine civique des droits, des obligations de la participation politique et de la sphère économique du marché du travail, mais est en réalité le SEUL

moyen pour les migrants de réellement comprendre une société donnée, y inclus ses lois, sa vie quotidienne et ses normes culturelles (Milani, 2008 : 44).

La maîtrise de la langue nationale est donc souvent prise comme indicateur d'une intégration réussie dans la communauté d'accueil et le manque de maîtrise de cette langue est souvent compris comme un manque de volonté du migrant de s'intégrer (Shohamy, 2009). Certaines interprétations vont parfois même jusqu'à conclure que le manque de maîtrise de la langue nationale (par incapacité ou par refus) est une menace pour la cohésion sociale et l'identité nationale (Blackledge, 2009c).

Certains auteurs tels que Oers (2010) and Mackenzie (2010) soulignent que, en parallèle à la hausse des prérequis pour le séjour permanent et/ou la nationalité, le concept même de citoyenneté a changé¹³. Mackenzie (2010) constate une évolution allant d'une conception de la citoyenneté comme phénomène passif, se basant sur un ensemble de droits et d'obligations, à une conception de la citoyenneté en tant que phénomène actif qui implique un engagement envers les institutions politiques. Dans cette perspective, la citoyenneté est prise comme indicateur de l'appartenance à une communauté donnée :

La citoyenneté est envisagée comme un signe qui SYMBOLISE la reconnaissance institutionnelle de la réalisation d'un ensemble de prérequis qui sont prétendument indispensables si un individu veut recevoir en guise de récompense l'identité de citoyen suédois, et par là même entrer pleinement dans la société suédoise (Milani, 2008 : 43).

Dans cette conception, il existe un lien direct entre intégration réussie et obtention de la nationalité. Ceci constitue cependant une simplification évidente, car il a été démontré que des

individus peuvent être intégrés dans une société sans avoir été naturalisés auparavant (Farrell, 2010) alors que d'autres sont moins bien intégrés, bien qu'ils soient citoyens depuis toujours. La nationalité est aujourd'hui considérée comme une récompense, quelque chose de précieux, dans la mesure où elle implique non seulement des droits et des obligations, mais également des privilèges (Cox, 2010).

Une conception de la nationalité en tant que point culminant du processus d'intégration grâce auquel un individu prouve son engagement envers la société d'accueil s'articule étroitement à une conception de la nationalité en tant qu'assimilation : il est attendu du futur citoyen d'abandonner son identité et ses valeurs antérieures et de pleinement adhérer aux normes culturelles et sociales de la société d'accueil (Kostakopoulou, 2010 ; Slade, 2010a).

Selon certains auteurs, l'idéologie nationaliste d'«une nation – une langue» est sous-jacente dans de nombreux pays européens qui exigent un degré de maîtrise de la langue 'nationale' – au détriment d'autres variétés régionales (Blackledge, 2009a ; Extra & Spotti, 2009a ; Shohamy, 2009) contribuant par là même à consolider l'hégémonie linguistique et culturelle déjà existante (Hogan-Brun et al., 2009b ; Shohamy, 2009). Il est attendu des migrants qu'ils démontrent leur intégration dans la société d'accueil en apprenant la langue standard et en adhérant aux valeurs culturelles du pays. Blackledge (2009c) va même jusqu'à dire que les régimes de tests, dans la mesure où ils imposent des niveaux d'anglais standardisés, reflètent une conception de l'*anglais* comme un ensemble de pratiques linguistiques homogènes, alors qu'en réalité il n'y a rien de tel que l'anglais standard, puisque les pratiques varient à travers les contextes. Van Avermaet (2009) attire l'attention sur le fait que l'évaluation de connaissances linguistiques et culturelles est généralement basée sur la présupposition que les sociétés sont homogènes en termes de normes culturelles et linguistiques, car des constructs simplifiés se prêtent mieux à l'évaluation.

Pour McNamara, la fonction symbolique de la maîtrise d'une langue nationale se trouve au

13 | Oers (2010) identifie dans la littérature trois concepts de citoyenneté : a) la citoyenneté en tant que statut légal (un rapport privilégié entre une personne et l'état) ; b) la citoyenneté en tant qu'activité (participation dans la vie sociale de la communauté) ; c) la citoyenneté comme marque identitaire (appartenance à une communauté, identification, loyauté, adhésion à la culture de la société).

premier plan :

La raison de l'inclusion d'un prérequis linguistique n'est pas en premier lieu dû à la fonction communicative du langage, mais à sa fonction symbolique. La fonction primaire du test n'est pas d'encourager le bien-être des migrants, mais d'exprimer une idéologie qui associe l'usage du langage à des valeurs culturelles (McNamara, 2009: 158).

Par conséquent, McNamara soutient que le construct sur lequel se base un test devrait être reformulé, étant donné que ces tests ne mesurent pas réellement une compétence fonctionnelle en langue (p. ex. des habiletés communicatives pratiques qui seraient pertinentes pour les candidats dans leur vie quotidienne en tant que résidents ou citoyens), mais évalue implicitement la 'conformité à un ensemble de valeurs sociales et culturelles' ou une 'idéologie nationale' (McNamara, 2009b). En fait, les compétences sont évaluées sur la base d'une dichotomie 'acceptable' vs. 'inacceptable' – établie par les décideurs politiques plutôt que par les concepteurs de tests – et interprétées en termes d'éligibilité ou non-éligibilité pour le séjour permanent ou la nationalité (McNamara, 2010).

Pour Kostakopoulo (2010), le lien entre test de connaissances et adhésion à la société d'accueil n'est pas évidente :

En effet, il est très rare de trouver une réponse à la question de savoir pourquoi l'on présuppose qu'une 'appartenance partagée' est quelque chose qui peut être obtenue en testant la fluidité de quelqu'un dans la langue d'accueil et l'accumulation d'informations factuelles à propos du domaine civique, de l'histoire ou de la vie dans le pays, qui risquent d'être oubliées quelques mois après le test, plutôt que sur la base d'expériences communes, de travail et de contribution aux biens communs et à l'amélioration du bien-être de la société (Kostakopoulo, 2010: 9).

Dans cette perspective, les pratiques habituelles en matière de tests ne semblent pas très pertinentes.

D'après Milani (2008), les tests de langue dans le contexte de l'immigration n'encouragent

que très peu l'intégration ; bien au contraire, ils encouragent l'exclusion sociale. Pour cet auteur, une contradiction émerge entre l'objectif déclaré des tests à des fins d'immigration et d'intégration tel qu'il est mis en avant dans le discours public et les conséquences effectives de ces tests : alors que les tests de langue pour la naturalisation sont censés encourager l'*inclusion* des migrants dans la société d'accueil – dans la mesure où l'on présuppose que la cohésion sociale et l'identité nationale peuvent être réalisées au moyen d'une langue commune – les effets de ces tests tendent en réalité plutôt à l'*exclusion*, puisque certains individus vont inévitablement échouer au test. Milani (2008) montre, sur la base d'entretiens menés avec différents acteurs, que les migrants eux-mêmes perçoivent ces politiques linguistiques comme ayant été établies dans le but d'exclure certains groupes sociaux. Milani (2008) souligne que les tests de langue contribuent à l'exclusion sociale, dans la mesure où une nouvelle délimitation est tracée entre des catégories sociales : les tests mettent en opposition le groupe des candidats ayant réussi (et ayant par là même obtenu la nationalité) et le groupe des candidats ayant échoué (lesquels sont confinés à leur statut de non-citoyen). Ils contribuent également à la distinction entre le groupe des migrants qui souhaitent devenir des citoyens et qui sont obligés de passer un test pour ce faire et le groupe des autochtones qui n'ont pas besoin de le passer. Par conséquent, au lieu de réduire la discrimination sociale, comme cela est clamé officiellement, les tests de langue pour la naturalisation contribuent à la reproduire en excluant certains groupes sociaux de certains domaines civiques et culturels (Milani, 2008).

Les auteurs prenant position dans le cadre du *Critical Language Testing* (Shohamy, 1998, 2001) dénoncent l'usage (ou l'abus) par les gouvernements de l'évaluation linguistique à des fins d'immigration et d'intégration en tant qu'instrument de régulation sociale (*gate-keeping*) (Blackledge, 2009c; Hogan-Brun et al., 2009b; Van Avermaet, 2009; Slade, 2010a; Shohamy, 2009; Wright, 2008) et pour préserver les privilèges des citoyens (Piller, 2001). Le recours à l'évaluation linguistique à cette fin

permet aux gouvernements de non seulement contrôler le flux migratoire, mais aussi de démontrer auprès de la population que tout est sous contrôle (Böcker & Strik, 2011; Slade, 2010a). Ainsi, les acteurs politiques peuvent affirmer qu'ils ont pris les mesures nécessaires pour assurer la cohésion sociale (McNamara, 2009b) en préservant la nation d'une trop grande diversité linguistique et culturelle (Kostakopoulo, 2010; Milani, 2008).

Dans le mécanisme de régulation sociale, la langue est prise comme indicateur d'une appartenance à un groupe social et la non-conformité avec les prérequis linguistiques entraîne par conséquent le refus de la nationalité et par là même l'exclusion sociale. Milani (2008: 53) souligne la tendance de certains états à 'élever la BARRE D'APPARTENANCE qui régle l'accès au groupe' par l'introduction d'une évaluation linguistique pour faire face aux flux migratoires. Le test australien nommé *Australian Dictation Test* (McNamara, 2005) est évoqué comme un cas extrême d'abus d'un test de langue par un gouvernement. Semblable au test Shibboleth dans les temps bibliques, le test australien visait délibérément à discriminer certains groupes sociaux qui n'étaient pas les bienvenus en Australie, en imposant aux candidats à l'immigration de passer un test dans une langue qu'ils ne connaissaient pas, ce qui ne leur laissait aucune chance de réussir le test (McNamara, 2005).

Siiner (2006) décrit un cas récent de régulation sociale en Estonie qui a eu un impact sérieux, voire même nuisible sur un grand nombre d'individus. Dans le but de reconstruire la nation après la fin de l'occupation soviétique et d'imposer l'estonien comme seule langue nationale, le gouvernement a imposé la maîtrise de l'estonien comme condition pour l'acquisition de la nationalité et a instauré une évaluation des compétences conversationnelles et écrites dans cette langue. Beaucoup de résidents russophones n'étaient pas en mesure de remplir cette condition, car ils ne savaient pas suffisamment la langue nationale. La nationalité leur était ainsi refusée, quand bien même ils étaient nés en Estonie et que, dans de nombreux endroits, il était tout à fait possible de fonctionner dans la vie quotidienne avec le

russe comme seule langue de communication. De par leur statut de non-citoyens, ils ont été privés de droits fondamentaux; par exemple, ils n'étaient pas autorisés à postuler pour certains emplois dans le secteur public. Par conséquent, une grande part de la minorité russophone est restée apatride, puisque ces individus n'étaient ni citoyens d'Estonie ni de Russie, et ont fini par être marginalisés de la société. L'exclusion de groupes minoritaires dans les pays baltes sur la base d'une compétence insuffisante dans la langue nationale, et rendue légitime par les législations nationales, a été largement critiquée par les institutions européennes (Ozolins, 2003).

Shohamy traite de la question de savoir si le recours à l'évaluation linguistique pour le séjour permanent et la nationalité est *juste* dans une perspective plus radicale, en prenant en compte des arguments éthiques et politiques fondamentaux.

Dans Shohamy (2006), l'auteure exprime son opposition aux prérequis linguistiques et à l'évaluation linguistique à des fins d'immigration et d'intégration, car elle ne considère pas éthique que des individus ne maîtrisant pas la langue dominante dans un contexte donné soient privés de l'opportunité de participer pleinement dans la société en tant que citoyens. Shohamy ajoute que, puisque cette privation n'est pas éthique, le refus d'octroyer la nationalité à quelqu'un sur la base d'une compétence linguistique insuffisante est une violation de ses droits individuels – une pratique discriminatoire. Les droits civiques et les obligations sont attribués à celui qui maîtrise la langue dominante, alors que ces mêmes droits et obligations sont refusés à celui qui ne la maîtrise pas. Selon Shohamy, le refus d'octroyer ces droits sur la base d'une compétence linguistique n'est pas une raison équitable, dans la mesure où 'il n'y pas d'indication que la maîtrise de la langue nationale crée automatiquement de meilleurs citoyens' (Shohamy, 2006: 148). De plus, Shohamy (2009; voir aussi Blackledge, 2009a) interroge la légitimité d'un état à imposer à un individu l'usage et l'apprentissage d'une langue donnée dans le cadre d'une politique linguistique et de tests de langue. Elle défend ce point de vue en déclarant :

L'on peut se demander si l'acquisition d'une langue 'nationale' ne devrait pas être un choix des individus qui sont en mesure de prendre les meilleures décisions de manière rationnelle quant à la langue qu'ils ont besoin de savoir et d'utiliser dans un monde multilingue, transnational (Shohamy, 2009 : 49).

En d'autres termes, obliger les individus à apprendre la langue nationale constitue une violation de leur droit à la liberté de parler la langue qui convient le mieux à leurs besoins quotidiens et transforme la connaissance d'une langue nationale en un 'devoir civique' (Shohamy, 2009).

A la lueur de ce qui a été présenté précédemment, Shohamy et d'autres experts critiques (p. ex. Blackledge, 2009c; Cox, 2010; McNamara & Shohamy, 2008; Van Avermaet, 2012) dénoncent les conséquences sociales d'un prérequis linguistique pour accéder au séjour permanent et à la nationalité, ainsi que l'instrumentalisation politique des tests de langue à cet égard. À travers leurs écrits, ils essaient de soulever la prise de conscience des utilisateurs de tests de l'impact que ces tests peuvent avoir sur la vie des candidats et appellent à leur responsabilité lors d'abus dans l'utilisation de tels tests.

En général, la notion de justice est utilisée comme moyen – souvent radical – pour critiquer les prérequis linguistiques et l'usage de tests de langue dans le contexte de l'immigration. Des raisonnements idéologiques, tels qu'une nation – une langue (standard) se révèlent à travers les prérequis linguistiques qui semblent avoir été établis pour le bien-être des candidats, c'est-à-dire leur intégration. Par ailleurs, selon ces auteurs, les tests de connaissances linguistiques et culturelles évaluent principalement la conformité à un ensemble de valeurs socio-culturelles, car il n'y a pas de preuves pour soutenir l'affirmation selon laquelle les constructs évalués seraient essentiels pour la participation sociale. De plus, il est également souligné qu'en réalité, les prérequis linguistiques et l'usage de tests débouchent sur une plus grande exclusion sociale (l'usage de tests pour la régulation sociale) que l'intégration visée officiellement par les autorités. Enfin, il y a

un courant critique qui voit dans les prérequis linguistiques et leurs conséquences une infraction aux droits fondamentaux des individus et relèvent par conséquent de pratiques contraires à l'éthique et qui sont discriminatoires.

3.3

Les retombées positives de l'usage des instruments d'évaluation

Dans le modèle conceptuel du AUA, lorsque les conséquences de l'usage d'un instrument d'évaluation et les décisions qui en découlent sont bénéfiques pour tous les acteurs, et non pas nuisibles, cela constitue un signe de validité de l'instrument en question. Cet aspect concerne aussi bien les individus que la société dans son ensemble.

Certains arguments relatifs aux retombées positives des instruments d'évaluation ont été discutés dans la section ci-dessus sur la notion de *justice*, mais dans une perspective générale (et foncièrement critique) seulement. Dans cette section, nous nous concentrerons davantage sur les aspects concrets signalés dans les études s'intéressant à l'impact des prérequis linguistiques sur les individus (p. ex. accès à l'emploi ou à l'éducation) et sur la société (p. ex. effets sur la cohésion sociale). Une première sous-section résume les éléments soulevés dans une multitude de publications portant sur les effets présumés ou réels des prérequis linguistiques et des régimes de test. Une deuxième sous-section se concentre sur les résultats provenant d'études se consacrant spécifiquement à l'impact des régimes de tests.

Typiquement, les mesures restrictives pour l'immigration et l'intégration sont présentées par les décideurs politiques, et parfois aussi par les concepteurs de test, comme étant motivées par des considérations pratiques et les effets potentiellement bénéfiques pour les migrants sont soulignés (Blackledge, 2009c). De Jong et al. (2009) font un effort évident pour justifier l'usage de leur nouveau Test pour le néerlandais parlé (TGN) en mettant en évidence les retombées positives d'une meilleure compétence linguistique pour les candidats et la société. Selon ces auteurs, un manque d'intégra-

tion (dû à des déficits linguistiques) entraîne ‘un cercle vicieux, dans lequel les parents ne peuvent pas aider leurs enfants, les enfants abandonnent l’école, ce qui mène à un sentiment de désespoir et d’angoisse’, alors que les tests aident à combattre les ‘conséquences négatives de la ségrégation sociale d’un grand nombre de migrants dans les Pays-Bas’. Des effets positifs sont aussi mis en avant en ce qui concerne les tests de langue pour l’entrée dans le pays d’accueil, dans la mesure où cela aide à combattre les mariages forcés et la traite des femmes, ‘une nouvelle forme de traite d’esclaves’. En guise d’argument additionnel, les auteurs citent les résultats d’une enquête qui a été menée avant le début de la conception du test. Les auteurs affirment que cette enquête a montré que

[...] malgré le potentiel pour une controverse politique, l’objectif et l’intention de la loi sur l’intégration des nouveaux résidents ont été soutenus et considérés justifiés par une large majorité à l’extérieur et à l’intérieur du parlement [et même par les migrants interrogés] (De Jong et al., 2009 : 43).

L’esprit des arguments amenés par ces auteurs se retrouve également ailleurs. Ce qui est habituellement mis en avant pour justifier l’introduction de prérequis linguistiques dans les politiques d’immigration est que ces mesures facilitent l’intégration des migrants dans la communauté d’accueil et encouragent leur participation active dans des activités quotidiennes et civiques (Blackledge, 2009c; Holland, 2010; Oers et al., 2010), et leur permettent de devenir économiquement indépendants (Böcker & Strik, 2011). Dans la perspective de ces retombées positives, les migrants sont supposés être équipés avec un degré suffisant d’une compétence linguistique ‘fonctionnelle’ (Blackledge, 2009c). L’obligation d’un niveau B1 selon le Cadre européen commun de référence pour les langues (Conseil de l’Europe, 2001) pour les migrants en Allemagne souhaitant obtenir le permis de séjour permanent ou la nationalité était motivée par l’objectif de rendre les migrants indépendants à la fois au niveau professionnel et privé (Oers et al., 2010). L’introduction d’un test de langue

exigeant est présentée comme moyen pour garantir de nouveaux citoyens qui soient indépendants sur le plan économique et actifs sur le plan social. Les cours et tests de connaissances sur l’histoire, la culture et le droit de la société d’accueil sont supposés compléter le répertoire nécessaire pour pouvoir participer pleinement à la vie sociale (Cooke, 2009; Milani, 2008).

Nous avons montré dans la section sur la justice ci-dessus que l’argument selon lequel les mesures prescrites seraient bénéfiques est perçu de manière critique dans bon nombre d’études. Il y a cependant d’autres études qui présentent les effets positifs des prérequis linguistiques (plus stricts) comme un fait. Kiwan (2008), par exemple, argumente que la préparation pour le test de naturalisation britannique peut être envisagée comme une opportunité pour les migrants d’acquérir des connaissances utiles au sujet de leurs droits et obligations en tant que futurs citoyens. Dans cette optique, le test peut être conçu comme un instrument encourageant une meilleure compréhension d’une participation active sur le plan civique. Il est à noter que ce n’est pas le test de langue en soi qui est considéré comme encourageant l’intégration mais plutôt la préparation à ce test dans le cadre d’un cours d’intégration. De manière semblable, Yoffe (2010) voit des effets positifs du Contrat d’accueil et d’intégration qui a été introduit en France en 2007. Ce contrat oblige les migrants nouvellement arrivés à suivre une formation linguistique si leur niveau de français est plus bas qu’A1.1 de l’échelle du CECR. Pour Yoffe, l’obligation de suivre des cours de langue offre l’opportunité – notamment aux femmes musulmanes non-instruites – de sortir de leur environnement social circonscrit et de communiquer avec des locuteurs natifs, alors que ces femmes ne seraient pas autorisées à socialiser avec des autochtones si ces cours n’étaient pas obligatoires. Un autre effet positif que Yoffe observe est le fait que les migrants ayant complété avec succès leur formation linguistique gagnent plus de ‘confiance dans leur capacité à fonctionner dans la société et la motivation à poursuivre la formation linguistique’ (Yoffe, 2010 : 77).

Les études empiriques s’intéressant aux

conséquences de l'usage des instruments d'évaluation sont traditionnellement connues dans la recherche en matière de validation comme *études d'impact* (ou *études de was-hback* pour le contexte éducatif). Ces études identifient les conséquences des tests de langue sur la société dans son ensemble (p. ex. effets sur la cohésion sociale) et sur la vie des candidats (p. ex. meilleur accès à l'emploi ou à la formation).

Les études d'impact en matière d'évaluation linguistique à des fins d'immigration et d'intégration sont plutôt rares, ce qui est en partie dû au fait que l'introduction de prérequis linguistiques et de tests est un phénomène récent (Böcker & Strik, 2011). Van Avermaet & Rocca (2013) observent que les études d'impact tendent à se limiter à des décomptes statistiques, tels que les taux de réussite aux tests ou le nombre de participants dans les cours préparatoires. Ils ne donnent donc que très peu d'informations concernant l'impact d'un régime de test sur le processus d'intégration ou sur la participation sociale. Les auteurs soulignent ainsi le besoin de mener plus de recherches sur l'impact social des politiques d'intégration sur le processus même d'intégration des immigrants.

Une étude importante récente qui relève précisément de ce type de recherche est l'étude INTEC (Strik et al., 2010). Elle examine les pratiques effectives résultant des politiques linguistiques et l'usage de tests à différentes étapes du processus d'immigration (l'entrée sur le territoire, la demande de permis de séjour permanent et la naturalisation), ainsi que les effets des régimes de tests mis en place dans neuf états membres de l'UE (Allemagne, Autriche, Belgique, Danemark, France, Hongrie, Lettonie, Pays-Bas, Royaume-Uni). L'étude présente deux types de données : a) des données statistiques sur le nombre de permis octroyés ou refusés ; b) des données qualitatives provenant d'un total de 329 entretiens menés avec différents acteurs (principalement des immigrants ayant l'obligation de remplir les conditions d'intégration, mais aussi des enseignants de langue, des représentants des instances administratives et politiques et des collaborateurs d'organisation non-gouvernementales).

Pour certains pays, les *données statistiques* sur les procédures de naturalisation montrent un déclin dans le nombre de migrants ayant obtenu la nationalité depuis que les conditions sont devenues plus restrictives. Au Danemark par exemple, le nombre de refus de la nationalité a augmenté en 2002-2003 lorsque des tests standardisés ont été introduits, et à nouveau en 2007-2009 lorsque le niveau de langue requis a été rehaussé de B1 à B2. De manière semblable, Oers (2008) observe une diminution significative des demandes de naturalisation aux Pays-Bas en 2004 (70% de demandes en moins en comparaison à 2002) immédiatement après l'introduction de tests de naturalisation. Dans le cas des Pays-Bas et de l'Allemagne, qui ont introduit des tests pour l'entrée dans le pays, l'étude INTEC constate un déclin dans le nombre de demandes – qui semble d'ailleurs concerner principalement le regroupement familial. Cela signifie que, probablement, pour la plupart des femmes concernées qui souhaitaient rejoindre leur partenaire (que ce soit leur mari ou futur mari), les nouveaux régimes de test ont eu un effet non-bénéfique. L'échec au test a eu pour conséquence d'empêcher la réunification familiale ou la conclusion d'un mariage.

Selon l'étude INTEC, les *données qualitatives* indiquent également que les migrants eux-mêmes perçoivent les effets des nouvelles mesures comme n'étant pas bénéfiques. En ce qui concerne les procédures de naturalisation, les migrants en Autriche, au Danemark, en Allemagne, au Royaume-Uni et en Lettonie étaient sceptiques quant à l'impact positif de ces tests sur leur processus d'intégration – dans les domaines de l'emploi et de la vie sociale par exemple. Au lieu de cela, les informateurs semblent attribuer la réussite d'une intégration à d'autres facteurs. Cela n'est pas très surprenant, dans la mesure où la plupart d'entre eux se sentaient déjà intégrés au moment de la demande de naturalisation. Certains migrants ont explicitement souligné comme effets négatifs le stress et l'anxiété qu'ils ont ressentis (ainsi que leurs familles) lorsqu'ils devaient passer un test dont le résultat est suffisamment important pour avoir un impact considérable sur leur vie. Les cours préparatoires ont également été critiqués pour l'investissement en temps qu'ils

impliquent et leur incompatibilité avec des obligations professionnelles.

L'étude INTEC (Strik et al., 2010) identifie également des effets positifs (ou : bénéfiques). Par exemple, les enseignants d'allemand ont observé des cas où des candidates avaient échoué délibérément au test d'entrée afin d'éviter le mariage forcé. Certains informateurs qui avaient fait une demande de permis de séjour permanent s'accordaient à dire que les cours de préparation obligatoires avaient eu un 'effet émancipatoire'. Les cours de langue obligatoires (en Allemagne ou en Autriche par exemple) semblent constituer une opportunité pour certaines femmes de socialiser, car elles ne seraient pas autorisées à suivre un programme d'intégration sur une base volontaire. La possibilité d'avoir une vie sociale a des effets positifs visibles sur l'estime de soi auprès de femmes peu instruites.

Les résultats d'une étude d'impact plus petite menée en Flandres et présentée dans Van Avermaet & Rocca (2013) et Van Avermaet (2012) montre un impact positif limité. L'étude se fonde sur 40 entretiens semi-directifs avec divers acteurs (des migrants, des enseignants, ainsi que des représentants du groupe 'majoritaire' tels que les employeurs). En Flandres, les nouveaux arrivants ainsi que les migrants établis ont l'obligation de suivre des cours d'intégration dans le but d'atteindre le niveau A1 en flamand et d'acquérir des connaissances de base sur la société. Au terme du cours d'intégration, les participants ayant terminé le cours avec succès reçoivent un certificat d'intégration, dont la décision de réussite ou d'échec incombe à l'enseignant.

Les migrants interrogés qui ont suivi le cours plus qu'une année auparavant et qui n'ont pas trouvé d'emploi après le cours étaient d'avis, sans surprise, que le cours d'intégration et le certificat n'étaient pas utiles pour l'accès au marché du travail. Les entretiens avec les employeurs et les agences de placement pour l'emploi ont quant à eux démontré que le certificat n'était même pas pris en considération dans les postulations et que les compétences linguistiques des migrants étaient évaluées de manière informelle au cours des entretiens d'embauche. De manière générale, la compétence linguistique

n'apparaît pas comme un facteur important pour l'accès au marché du travail en Flandres. L'étude conclut que les mesures d'intégration ont dans leur ensemble un impact plutôt limité sur le réel processus d'intégration des immigrants.

En somme, les études disponibles montrent que les retombées positives visées par les prérequis linguistiques et les tests de langue à des fins d'immigration et d'intégration restent une question controversée. Les décideurs politiques et les concepteurs de test ont tendance à légitimer les nouvelles mesures en soulignant les retombées positives potentielles pour les migrants et la société : une meilleure intégration et plus d'indépendance économique au moyen d'une connaissance approfondie de la langue et de la société. Mais, les quelques études existantes ne soutiennent cette affirmation que de manière limitée : certains experts ont observés que des migrants ont profité des nouveaux régimes de tests dans le cadre des cours préparatoires, notamment les femmes musulmanes qui ont ainsi eu la possibilité de socialiser en-dehors de leur famille. Les études d'impact se basant sur des données statistiques et des entretiens avec un nombre raisonnable d'acteurs divers ont montré plutôt des effets négatifs ou l'absence même de retombées : là où un test de langue est obligatoire pour obtenir un visa d'entrée, le nombre d'individus (souvent des femmes qui souhaitent se marier) pouvant immigrer est réduit ; là où un certificat de langue est une condition pour les demandes de naturalisation, le nombre de résidents obtenant la nationalité est réduit ; pour les candidats qui sont déjà bien intégrés, les cours d'intégration et les examens entraînent avant tout un horaire chargé et plus de stress. Une étude montre que les certificats délivrés lorsque les conditions sont remplies ont peu ou pas du tout de valeur sur le marché de l'emploi. Ainsi, il y a très peu d'éléments dans la littérature pouvant soutenir les déclarations optimistes concernant les retombées positives des mesures plus strictes pour ce qui est de la compétence linguistique et des connaissances culturelles. Il apparaît également que plus d'études d'impact sont nécessaires.

3.4 Décisions sensibles aux valeurs

Bachman & Palmer (2010) suggèrent que pour qu'un instrument d'évaluation soit sensible aux valeurs, les concepteurs de test doivent s'intéresser aux valeurs des différents acteurs et les prendre en compte d'une manière ou d'une autre. Les auteurs vont encore plus loin en encourageant les concepteurs de test à remettre en question les valeurs existantes de la communauté si celles-ci mettent en danger le traitement équitable des candidats.

Sur la base de la littérature incluse ici, nous pouvons dire de manière générale que les concepteurs de test n'accordent que peu d'importance aux valeurs, alors que ces dernières semblent être abordées de manière prioritaire par les experts en critique sociale qui traitent des prérequis linguistiques et des tests de langue.

La manière dont Plassmann, en tant que représentante du *telc* ayant contribué à développer le Test allemand pour les immigrants (DTZ), traite le fait que les candidats doivent témoigner d'un niveau B1 (généralement considéré comme exigeant) à la fin du cours d'intégration, peut donner des pistes pour expliquer cet apparent désintérêt. Elle argumente simplement que le niveau B1 a été imposé par les législateurs et que, par conséquent, une discussion quant à savoir quel niveau est adéquat pour quel objectif n'est pas nécessaire (Plassmann, 2011 : 25). Les développeurs du Test de néerlandais parlé (TGN) se trouvent dans une situation semblable au cas allemand – le TGN est également un test qui a été mandaté par les autorités – mais abordent cependant, bien que brièvement, la question des valeurs avant de poursuivre avec la présentation du travail de validation technique (De Jong et al., 2009). Ils mentionnent ainsi différents faits et résultats afin de démontrer que leur test se fonde sur une base solide en termes de valeurs et d'acceptabilité. Ils soulignent que l'introduction du nouveau test se base sur un vote parlementaire ayant reçu un large soutien, qu'il a été mandaté par le Ministère de la Justice, et que ces mêmes autorités ont également choisi le CECR comme échelle de référence. Ils rapportent par ail-

leurs que l'opposition politique n'a pas trouvé de soutien, pas même auprès de la population migrante. Ils poursuivent avec une explication des retombées sociales attendues de leur test: le TGN encourage l'intégration, tandis que le manque d'intégration peut créer un 'cercle vicieux' entraînant le désespoir et l'anxiété; le test d'entrée sur le territoire permet de contrôler la traite des femmes; ceux qui souhaitent fournir un effort minimal ont quand même la possibilité d'immigrer puisque le test n'est pas trop exigeant (De Jong et al., 2009 : 41-43). Les concepteurs du test néerlandais font clairement preuve d'une attitude positive envers leur test et ne prennent pas en considération les points de vue opposés¹⁴.

3.5 Décisions équitables

Selon Bachman & Palmer (2010), la nature équitable d'une décision basée sur une évaluation dépend principalement de deux aspects: une opportunité égale pour tous les candidats d'un niveau de compétence semblable d'être classés dans le même groupe (p. ex. 'réussi' ou 'échoué') et une opportunité égale pour les différents groupes de candidats d'acquiescer la ou les compétence(s) visée(s). L'appartenance à un groupe social ne devrait pas avoir d'influence sur les décisions prises à partir des résultats des tests.

L'argument selon lequel les tests formels seraient (au moins potentiellement) plus équitables que, par exemple, un entretien informel avec un fonctionnaire est mentionné dans plusieurs publications. Joppke (2010; aussi Wright, 2008), par exemple, considère les tests standardisés introduits récemment pour la naturalisation plus équitables et libéraux que l'évaluation informelle antérieure, dans la mesure où tous les candidats sont traités de la même manière. Plusieurs études citent des exemples de régimes de tests informels qui mènent à une inégalité de traitement évidente: par exemple, Schneider et al. (2006) dé-

14 | Il y a même une pointe de triomphe et de ridicule dans certains passages.

crivent des pratiques locales de naturalisation en Suisse et Piller (2001) décrit des pratiques similaires en Allemagne, qui sont déterminées en grande partie par les préférences des fonctionnaires. Laversuch (2008) montre comment cette situation a été utilisée à mauvais escient à des fins de régulation sociale et en particulier à l'encontre des musulmans. Le groupe d'experts ALTER LAMI, dont la position générale est plutôt sceptique envers les tests, mentionne les avantages que peuvent avoir des tests qui sont 'conçus, élaborés et administrés de manière adéquate' par exemple, des 'résultats qui sont hautement standardisés et fiables' et des 'candidats qui sont évalués avec un degré élevé d'indépendance et d'objectivité'. Ils défendent par conséquent le développement et l'usage de tests selon des standards de qualité – en particulier dans le cas où les tests sont utilisés pour des prises de décision en matière d'immigration (Balch et al., 2008 : 2).

Alors que le potentiel des tests formels d'assurer plus d'équité est reconnu par différents auteurs, les régimes de tests hautement standardisés sont également critiqués comme n'étant pas équitables, dans la mesure où ils impliquent de nouveaux obstacles à franchir qui désavantagent certains groupes de migrants. Shohamy souligne que beaucoup de migrants peu ou pas instruits ne sont pas habitués aux procédures de test et qu'ils ne les réussissent que rarement. Böcker & Strik (2011) mettent en avant que les gens plus pauvres ne sont pas en mesure de payer les taxes demandées pour passer le test. Pour Michalowski (2010) et Shohamy (2009), les tests sont des instruments de sélection parce que seuls des groupes hautement qualifiés et instruits sont acceptés, puisqu'ils sont les seuls à pouvoir surmonter tous les obstacles.

D'après plusieurs auteurs, cette inégalité commence bien avant le test en tant que tel, car les candidats n'ont pas le même accès à l'information, à la langue d'instruction et à la préparation du test. Ceci concerne en particulier les contextes dans lesquels un test est demandé pour l'entrée sur le territoire (Allemagne, Pays-Bas). Les informateurs de l'enquête INTEC (Strik et al., 2010) au Maroc attirent l'attention sur le fait que l'accès au matériel pour préparer

le test implique de se déplacer à l'ambassade et qu'il n'y a pas toujours de cours de langue là où vivent les candidats.

Shohamy (2009) observe que les migrants peuvent même rencontrer des difficultés à entrer en contact avec la langue dans le pays d'accueil et qu'ils manquent ainsi d'opportunités pour suivre une formation dans la langue cible jusqu'au niveau requis pour le test de naturalisation.

Pour Krumm (2007), il est peu probable que l'application d'un même test à tous les migrants puisse mener à plus d'équité. Il observe qu'il y a une grande hétérogénéité dans la population migrante en termes d'arrière-fond linguistique, éducatif et culturel, ainsi que des niveaux de compétences linguistiques nécessaires pour fonctionner dans la société et le marché du travail. D'après Krumm, cette hétérogénéité devient problématique, dès le moment où l'on veut tester ces personnes :

Le fait de tester des gens qui ne sont pas égaux dans tous les aspects de leurs compétences linguistiques et culturelles au moyen du même test et attendre un niveau de compétence identique de leur part dans tous les domaines ne peut possiblement pas être un moyen pour démontrer l'égalité dans la société (Krumm, 2007 : 668).

Se basant sur des données statistiques concernant le Test allemand pour les immigrants (DTZ), Klein (2013) observe que certains groupes de candidats rencontraient plus de problèmes pour passer le DTZ pour le niveau requis que d'autres. Elle suggère par conséquent de donner plus de soutien aux groupes désavantagés lors de la préparation de l'examen: les femmes ayant le turc comme L1, les hommes ayant le russe ou le polonais comme L1, ainsi que les personnes âgées ayant l'italien, le chinois ou le turc comme L1. Oers (2008) observe également que certains groupes sont particulièrement défavorisés par ces formats de test. Il mentionne les personnes âgées ainsi que les personnes peu ou pas instruites. Oers présente l'usage de tests avec des ressortissants du Surinam aux Pays-Bas comme un autre exemple de traitement inéquitable: bien que leur langue maternelle soit le néerlandais, ils

sont obligés de passer le Test de néerlandais afin d'obtenir la nationalité alors qu'il n'y a pas de test pour les citoyens néerlandais. Par ailleurs, Stevenson & Schanze (2009) évoquent un manque d'équité qui vient du fait que beaucoup de migrants hautement qualifiés, tels que les ressortissants de l'UE ou d'autres pays comme les Etats-Unis ou l'Australie ne sont pas soumis à des prérequis linguistiques et les tests y afférents, indépendamment de leur maîtrise de la langue locale.

Dans l'ensemble, ces études font apparaître une image mitigée en ce qui concerne l'équité des tests. Alors qu'il est généralement admis que les tests remplacent souvent des pratiques informelles biaisées, certains auteurs soulignent avec prudence les aspects discriminatoires que ces nouveaux régimes de tests impliquent.

3.6 Pertinence des conclusions relatives au construct

Afin que les conclusions à partir des données recueillies (c.-à-d. l'interprétation des résultats enregistrés) soient pertinentes en regard de l'usage de la langue dans un contexte d'usage authentique (TLU), le construct de compétence qui sous-tend l'évaluation doit être articulé à cet usage de la langue (p. ex. basé sur une analyse des besoins linguistiques), selon Bachman & Palmer (2010). Les tâches utilisées lors de l'évaluation doivent avoir des propriétés pertinentes semblables aux tâches existant dans le domaine d'usage authentique de la langue cible, ainsi que la notation des performances doivent pouvoir saisir les aspects mis en avant dans la définition du construct. Par ailleurs, le construct et les prérequis devraient être communiqués aux différents acteurs (y inclut les candidats au test) de manière accessible. L'affirmation de pertinence concerne des aspects qui ont souvent été discutés dans le cadre de la *validité de construct*.

La validité de construct est aussi une question qui est abordée par la littérature analysée pour le présent rapport. Dans leur dis-

cussion sur l'usage de tests de langue à des fins d'immigration et d'intégration, Saville & Van Avermaet (2008) présentent des questions clés que les concepteurs de test doivent traiter au moment de l'élaboration de tests de langue appropriés; l'une de ces questions concerne le construct sur lequel se fonde le test. Saville & Avermaet défendent que, dans la mesure où le test est officiellement déclaré comme encourageant l'intégration du candidat, le construct devrait être défini en conséquence. Cela signifie que le test devrait fournir des informations au sujet de la capacité du candidat à fonctionner dans la société d'accueil (plutôt que d'évaluer simplement des connaissances linguistiques décontextualisées). Le niveau de compétence linguistique visé devrait aussi correspondre à l'objectif du test. Autrement dit, si l'objectif du test est d'évaluer l'intégration sociale du candidat, alors le niveau de compétence attendu devrait correspondre aux besoins posés par un usage de la langue dans la vie quotidienne. Pour le dire plus concrètement, les directives développées par le groupe d'experts ALTE LAMI (Balch et al., 2008) postulent que les besoins de la vie quotidienne des migrants devraient être identifiés au moyen d'une analyse des besoins (*needs analysis*) qui étudie les situations et les tâches pratiques auxquelles les individus (ici: les futurs résidents ou citoyens) seront confrontés au cours de leur vie. L'identification des besoins de la vie quotidienne permet aux concepteurs de test de déterminer dans un prochain pas quelles compétences linguistiques devraient être évaluées, ainsi que le niveau de compétence en langue cible visé. Le groupe d'experts LAMI suggère que les prérequis communicatifs ne devraient pas juste être établis pour les migrants en général, mais devraient être définis de manière spécifique pour les différents sous-groupes (p. ex. les travailleurs, les conjoints, etc.), car leurs besoins de communication dans la vie quotidienne peuvent être très différents. Shohamy (2009) souligne que les tests en tant que tels mettent souvent l'accent sur des standards linguistiques peu authentiques (p. ex. l'exactitude grammaticale), qui n'incluent pas les particularités de l'usage d'une langue seconde et des pratiques multilingues quotidiennes. Par conséquent, les tests de

langue ne reflètent pas les besoins réels de la vie quotidienne que les migrants expérimentent dans le pays d'accueil ; ils peuvent difficilement être pris comme points de repères sur lesquels baser les décisions ultérieures.

Fulcher (2004) observe que de nombreux pays ayant introduit des prérequis linguistiques comme partie intégrante des procédures d'immigration et d'intégration ont défini le niveau de compétence visé par rapport à l'échelle des niveaux du CECR (North, 2009 ; Conseil de l'Europe, 2001). Cette pratique est assez controversée dans la littérature recensée ici.

La controverse relative à l'usage du CECR concerne principalement, mais pas exclusivement¹⁵, la définition du construct. Krumm (2007) attire l'attention sur le fait que l'échelle des niveaux du CECR n'a pas été élaborée initialement dans le but d'évaluer les compétences linguistiques des migrants, mais plutôt celles d'apprenants de langue étrangère au sens classique. Krumm (2007) donne l'exemple d'un descripteur au niveau A1 qui contient l'élément 'proposer un toast', qui n'est évidemment pas une capacité pertinente pour la plupart des migrants pour qui les prérequis linguistiques ont été introduits. Dans la lignée des recommandations du groupe d'expert ALTE LAMI, il suggère que le construct sur lequel se fonde le test et les descripteurs y afférents devraient être adaptés aux contextes d'usage de la langue que les migrants rencontrent effectivement. D'autre part, l'on devrait reconnaître le fait que les migrants, dans leur vie quotidienne multilingue, connaissent souvent et utilisent une panoplie d'autres langues à côté de leur langue maternelle ; des descripteurs adéquats devraient intégrer les dimensions interculturelles et plurilingues. Krumm indique une autre facette du plurilinguisme qui devrait être prise en compte, dans la mesure où il s'agit d'un phé-

nomène fréquent au sein de la population migrante, notamment : un profil de compétences linguistiques différenciées pour la même langue. Les migrants développent souvent des compétences différenciées dans une langue, par exemple une compétence rédactionnelle faible mais de bonnes compétences de lecture et de compréhension orale. Krumm (2007) conclut que le CECR devrait être adapté au contexte spécifique de la migration, notamment en élaborant plus la dimension des compétences partielles et des répertoires plurilingues. En vue de l'évaluation de telles compétences, il suggère que le Portfolio européen pour les langues pourrait être un moyen plus approprié que les tests pour répondre à la diversité de ces répertoires (voir aussi Papp, 2010).

Shohamy (2007) est également critique en ce qui concerne l'adéquation du CECR comme base pour la définition du construct dans un contexte d'immigration. L'un des points qu'elle critique concerne la manière dont les niveaux de compétence plus élevés sont décrits par le CECR. Ces descripteurs impliquent des présuppositions sur le savoir, les compétences et le fonctionnement communicatif qui ne peuvent pas être appliqués à la population migrante. Shohamy critique aussi le fait que les descripteurs ne reflètent pas suffisamment le contexte pour ce groupe cible – l'objectif de l'évaluation, l'âge des apprenants, les différents contextes d'apprentissage ou la distribution fonctionnelle des différentes langues que les individus maîtrisent.

Alderson (2007) attire l'attention sur une problématique un peu différente qui concerne l'usage du CECR. Il observe une tendance des politiciens et des fonctionnaires à définir les standards de compétences que les migrants doivent atteindre en fonction du CECR, sans consulter des experts en langues. Cooke (2009) donne l'exemple du test de naturalisation britannique pour lequel le niveau de compétence ('Entry 3', dans les niveaux du curriculum de base des adultes ESOL, équivalent au niveau B1 du CECR) a été décidé par le gouvernement. A la suite de cette décision, le niveau établi a été fortement critiqué par les enseignants d'anglais langue étrangère comme étant trop élevé. La situation est similaire en Allemagne où

15 | Extra et al. (2009) et aussi Yoffe (2010), par exemple, critiquent le recours à l'échelle de niveaux du CECR comme un *instrument prescriptif* à des fins d'immigration, car, selon eux, le référentiel du CECR était conçu initialement comme *instrument descriptif* qui devrait être utilisé pour rendre compte des compétences langagières des individus. Alors que le CECR a été mis en place pour encourager la mobilité internationale, son instrumentalisation dans le cadre des politiques d'immigration a plutôt l'effet inverse en l'intégrant dans un système d'exclusion et de régulation sociale.

le gouvernement a également fixé les prérequis linguistiques pour le séjour permanent et la nationalité au niveau B1 (Plassmann, 2011) – une décision qui a été critiquée, elle aussi, pour être trop peu réaliste (Laversuch, 2008).

Certaines études remettent sérieusement en question le construct concernant la compétence langagière. La position radicale de Shohamy, qui considère que les tests de langue pour l'immigration touchent aux droits des personnes d'utiliser leur propre langue, a déjà été évoquée plus haut. De manière encore plus radicale, Shohamy (2009) interroge le présupposé selon lequel la maîtrise d'une langue nationale est réellement indispensable pour fonctionner dans la société. Dans le cas où elle *serait* effectivement nécessaire, le degré de maîtrise requis devrait être reconsidéré. Dans une optique semblable à Shohamy, Saville & Van Avermaet (2008) relativisent l'importance de connaître la langue locale, dans la mesure où la difficulté des migrants d'accéder à l'emploi et à la formation ne dépend pas seulement des compétences linguistiques, mais résulte plutôt de leur marginalisation en tant qu'étrangers depuis leur arrivée. Les auteurs rappellent également que ces individus, du fait qu'ils vivent dans des sociétés multilingues, peuvent tout aussi bien fonctionner dans leur vie quotidienne en ayant recours à d'autres langues que la langue locale officielle, bien que ces pratiques linguistiques ne soient pas reconnues comme une manière de participer à la vie sociale. Slade (2010a) souligne que les connaissances linguistiques et culturelles que les migrants parviennent à acquérir à travers le temps et à travers leurs propres expériences peuvent varier considérablement – ce qui rend difficile l'établissement de standards (fondés) que tous les migrants devraient viser.

Papp (2010) rend compte d'une étude de validité à petite échelle qu'elle a menée sur le test britannique pour l'obtention de la nationalité et du permis de séjour permanent, 'Life in the UK' (introduit en 2005). L'auteure y a examiné le test en tant que tel ainsi que le matériel disponible pour les candidats s'y préparant. L'étude se concentre en particulier sur la question de savoir si le matériel du test reflète le domaine d'usage de la langue cible et le niveau de compétence visé ('une compétence fon-

tionnelle requise pour la démonstration réussie de citoyenneté et d'établissement'). L'étude de Papp fait apparaître deux résultats principaux. Premièrement, l'analyse du test et de la documentation montre que les compétences linguistiques évaluées ne donnent pas de preuves concernant la capacité du candidat à s'intégrer dans la société d'accueil. Les contenus du test ne correspondent pas au domaine d'usage de la langue pertinent pour la participation dans la vie sociale quotidienne (p. ex. emploi, administration publique, formation). Deuxièmement, sur la base d'une analyse de corpus, Papp conclut que le niveau de langue utilisé dans le test et la documentation s'avère être plus élevé que le niveau visé. Ce résultat global plutôt négatif met en évidence un problème de pertinence (et de validité de construct), dans la mesure où la conception du test et l'interprétation des résultats ne reflètent pas le construct du test de manière juste.

Le rapport sur le test australien de citoyenneté (*Australian Citizenship Test*) a été établi par un comité présidé par Woolcott, un diplomate, en raison de critiques fréquentes. Le rapport se concentre sur l'adéquation du niveau de langue requis pour le test, en plus des contenus controversés et des biais du test. Dans la loi sur la citoyenneté, le test devrait nécessiter un niveau *basique* d'anglais. Dans les prises de position que les experts en langues ont soumis à Woolcott, les critiques soulignaient le fait que les prérequis linguistiques du matériel de préparation et la formulation des items du test dépassaient largement les 'connaissances basiques'. Le rapport Woolcott tente de définir le prérequis pour des 'connaissances basiques' de manière plus pertinente en tant qu' 'ayant suffisamment de connaissances d'anglais pour vivre dans la communauté australienne au sens large' et de le situer autour des niveaux A1/A2 du CECR (McNamara & Ryan, 2011 : 172). Apparemment, cette clarification a eu peu de conséquences pratiques : McNamara & Ryan (2011) affirment que le niveau d'anglais utilisé dans la version révisée du test et du matériel de préparation n'a pas été modifié.

Kunnan (2009) compare de manière critique l'objectif déclaré (fonctionnant comme construct implicite) du test de naturalisation

américain révisé et introduit en 2008, et son implémentation effective. L'objectif déclaré du test est de soutenir la participation civique et l'intégration sociale. Kunnan soulève deux problématiques. L'une des deux concerne le type de savoir civique qui est demandé dans le test qui tend à se limiter à la mémorisation de faits au lieu de poser des tâches qui encouragent et requièrent une compréhension approfondie de l'histoire américaine et du gouvernement. Un deuxième point concerne le type et le niveau de connaissances d'anglais requis pour passer le test. D'après Kunnan (2009), le niveau n'est probablement pas assez élevé pour démontrer la capacité des candidats à participer à la vie civique et sociale dans la société américaine. À la lueur de ces éléments ainsi que d'autres faiblesses, Kunnan (2009) en conclut que le test de naturalisation américain ne peut pas réaliser son objectif déclaré de promouvoir l'intégration civique et sociale, remettant dès lors en question sa pertinence :

Le test de naturalisation, tel qu'il est conçu, ne peut pas tester le nationalisme civique ou l'intégration sociale au moyen d'une évaluation indirecte des connaissances d'anglais et des connaissances d'histoire américaine et du système politique, étant donné que ceux-ci sont des capacités et des savoirs, mais ne sont pas des preuves d'une participation à la communauté et d'activisme (Kunnan, 2009).

Kunnan mentionne un élément supplémentaire en défaveur du test de naturalisation révisé, à savoir le taux de naturalisation relativement bas qui semble indiquer que le test décourage les candidats potentiels.

Seul un nombre réduit d'articles sont disponibles sur le Test d'allemand pour les immigrants (DTZ) qui ont été écrits par des personnes activement impliquées dans sa conception. Perlmann-Balme (2011) et Plassmann (2011) tentent de démontrer que leur travail respecte les standards de qualité définis par ALTE, une association à laquelle appartiennent la plupart des concepteurs de test en Europe. Les standards 1 et 2 de ALTE sont en rapport avec la pertinence des conclusions se basant sur des tests comme relevant de l'usage

de la langue cible :

- 1) L'examen est fondé sur un concept hypothétique, par exemple sur un modèle de compétences communicationnelles.
- 2) Vous pouvez décrire l'objet et le contexte d'utilisation de l'examen, et le public auquel s'adresse l'examen.¹⁶

Les concepteurs du DTZ tentent d'atteindre ces standards en combinant une approche généralement orientée vers l'action telle que décrite par le CECR (Conseil de l'Europe, 2001) qui divise la compétence en langue dans les quatre compétences 'classiques', avec une orientation vers les besoins spécifiques des groupes cibles. Le curriculum-cadre allemand pour les cours d'intégration (Buhlmann et al., 2007) sert de source d'informations pour connaître les besoins spécifiques. Pour le construct du DTZ, les domaines centraux de l'usage de la langue qui sont communs aux différents groupes de candidats sont tirés d'un ensemble de 12 domaines esquissés dans le document de référence. Le construct et les objectifs du DTZ sont communiqués au grand public au moyen d'un manuel du DTZ (Perlmann-Balme et al., 2009), qui contient également un exemple de test.

Dans leurs travaux, Perlmann-Balme et Plassmann expliquent les étapes de l'élaboration d'un test qui met en pratique un construct pertinent tout en adhérant aux standards selon ALTE. Malheureusement, ces travaux ne fournissent pas assez d'éléments de preuve qui permettraient au lecteur de tirer des conclusions par lui-même.

De Jong et al. (2009) font preuve d'un plus grand effort pour démontrer que le Test de néerlandais parlé (TGN), administré par téléphone et par ordinateur, permet des interprétations pertinentes dans le contexte des tests de langue à des fins d'immigration et d'intégration, bien que le test utilise des types d'items qui ne seraient pas en soi identifiés intuitivement comme relevant de tâches orales communicatives de la vie quotidienne. Trois types d'items

16 | Version française originale (cf. ALTE [Association of Language Testers in Europe], 2007b)

sont utilisés pour le TGN : des répétitions de phrases, des réponses courtes à des questions (ce qui se rapproche le plus d'un type de tâche authentique), et les contraires de mots isolés¹⁷. Il n'y a pas d'interlocuteur humain ; le test entier est administré et noté de manière informatisée. Tous les types d'item fournissent de l'input parlé et un peu d'output parlé. D'après les auteurs,

[...] le test mesure la facilité avec laquelle les candidats sont capables de suivre ce qui est dit, d'extraire du sens en temps réel, et de formuler et produire des réponses pertinentes, intelligibles, à vitesse conversationnelle (De Jong et al., 2009 : 43).

La présentation sous cette forme des besoins posés par les items évoque des parallèles avec des opérations cognitives et linguistiques qui jouent également un rôle important dans les interactions verbales.

Les auteurs présentent plus d'arguments dans le but de démontrer la proximité entre la langue utilisée dans le test et l'usage authentique de la langue cible :

- Les stimuli utilisés pour la répétition de phrases ont été choisis à partir d'un corpus de néerlandais parlé et représentent le langage parlé spontané de la vie quotidienne et provenant de différentes régions.
- Le système automatique de notation a été élaboré sur la base d'un échantillon vaste de locuteurs natifs et non-natifs du néerlandais afin de représenter l'éventail des pratiques de néerlandais dans la vie de tous les jours.
- Le TGN évalue le vocabulaire, la grammaire, la prononciation et la fluidité en tant qu'indicateurs d'une compétence orale en langue cible. Les évaluations réelles des performances portent aussi communément sur ces mêmes aspects du langage du candidat, étant donné qu'ils sont considérés comme fondamentaux.

Une grande différence qui réside entre les deux types d'évaluation, mais aussi entre la langue utilisée dans le TGN et l'usage du langage dans

des situations authentiques de la vie de tous les jours, est le degré d'ouverture des tâches. Alors que les trois types d'items du TGN guident étroitement la (re-)production du langage, l'usage de la langue dans des situations ciblées fait appel souvent à une construction active et à une négociation du discours.

Il apparaît que les concepteurs du TGN ont anticipé cette possible faiblesse dans leur argumentaire de validité et ont fait passer à un grand nombre de candidats des tâches additionnelles telles que celles de répéter une histoire, répondre à des questions ouvertes et participer à un entretien oral. Les résultats ont été évalués par des examinateurs humains, certains en fonction des niveaux du CECR de manière à ce que ces résultats puissent être intégrés dans une analyse commune qui mène à un seul niveau du CECR. Par la suite, cette échelle a pu être utilisée en comparaison avec l'échelle du TGN qui se base sur une évaluation informatisé du vocabulaire, de la grammaire, de la prononciation et de la fluidité avec une précision satisfaisante en rapport avec le CECR, en particulier dans les niveaux les plus faibles où des décisions de réussite ou d'échec sont prises.

Sans surprise, la problématique de la pertinence des conclusions fondées sur des tests – qui couvre en grande partie ce qui a été traité en tant que validité de construct – reçoit une attention considérable dans les études traitées dans cette publication. Certains auteurs traitent la question de ce qu'un construct approprié devrait être, si le construct devrait s'étendre au-delà d'une langue particulière, si l'échelle des niveaux du CECR est pertinente dans ce contexte spécifique ou si les prérequis sont justifiés. D'autres analysent de manière critique les tests existants ; d'autres encore expliquent ou justifient les constructs qu'ils utilisent pour leur propre test. Afin d'illustrer ce dernier cas de figure, l'argument en rapport avec le construct du Test de néerlandais parlé (TGN) qui a été présenté par ses auteurs a été décrit ici dans plus de détails que d'autres, car la manière exhaustive et transparente avec laquelle il a été présenté est plutôt exceptionnelle. Cependant, il est utile de rappeler que malgré ses qualités, la publication susmention-

17 | A la fin de chaque test, deux items de répétition d'une histoire sont donnés. Ceux-ci, cependant, servent à la validation du test et à la recherche uniquement.

née est avant tout un texte à visée affirmative qui a été écrit par les concepteurs du test dans le but de prouver que le test s'ajuste à ses objectifs.

3.7 Impartialité des conclusions pour tous les groupes de candidats

Selon Bachman & Palmer (2010), l'interprétation des données recueillies ne peut être impartiale que si tous les aspects de l'administration du test, les formats et les contenus de l'évaluation ainsi que l'accès aux informations concernant l'évaluation ne désavantagent pas spécifiquement des groupes de candidats.

De Jong et al. (2009) ainsi que Perlmann-Balme (2011) et Plassmann (2011), c'est-à-dire les publications dans notre collection qui fournissent des détails sur la conception du test, abordent la question de l'impartialité du test.

De Jong et al. (2009) considère apparemment l'administration du test dans les ambassades néerlandaises dans le monde comme pouvant poser potentiellement un problème de biais. Dans leur article, les auteurs défendent ou justifient cette pratique avec les arguments suivants : 1) l'explication sur les procédures de test est suffisante pour chaque candidat, parce qu'elle est donnée par un personnel formé dans la langue du candidat. En cas d'absence d'un tel personnel, les candidats au test ont la possibilité d'amener un interprète à l'ambassade. 2) Alors que la fraude est une source bien connue de partialité dans les systèmes de test hautement décentralisés, les auteurs affirment que cela est littéralement impossible avec le TGN, car le test se base sur une grande banque d'items, il est unique pour chaque candidat et la notation se fait automatiquement au moyen d'un système informatisé.

Les concepteurs du TGN font également un effort pour prouver que les formats d'items et les contenus ne constituent pas un risque pour l'impartialité du test. L'un des formats utilisés dans le TGN consiste en des réponses courtes. Celles-ci sont nécessairement des réponses ouvertes à un certain degré et impliquent une

connaissance préalable du contenu. Cela peut générer un biais lorsque le test est administré au niveau international à un public hétérogène. Les concepteurs du TGN contrecarrent ce risque en démontrant les mesures qui ont été prises pour minimiser cette possibilité par un pré-test des items sur trois différents groupes : des candidats potentiels au test inscrits à des cours d'intégration, des migrants ne suivant pas de cours, ainsi que des locuteurs natifs d'arrière-fonds sociaux divers et appartenant à différents groupes d'âge, des femmes et des hommes. De la même manière, les concepteurs du TGN ont testé plusieurs autres hypothèses en rapport avec des variations potentielles en termes d'arrière-fonds qui pourraient excessivement influencer la réussite du TGN, le niveau d'instruction atteint, l'âge, le sexe, le degré d'alphabétisation par exemple, et aussi une possible sanction excessive des accents forts dans le système automatisé de notation.

En lien avec le standard ALTE n°2 – 'Vous pouvez décrire... le public auquel s'adresse l'examen.' – les concepteurs du test allemand DTZ mentionnent certaines mesures qui ont été prises pour minimiser un biais potentiel au sein des groupes visés par le test. Ces groupes incluent, d'un côté, les migrants avec un arrière-fond éducatif modeste qui souhaitent être actifs dans la famille ou trouver un emploi qui exige des qualifications minimales, et d'un autre côté, les migrants avec des expériences d'apprentissage, voire même des diplômes, qui souhaitent poursuivre leur carrière. L'une de ces mesures est en lien avec les contenus du test et le construct : les tâches ont été sélectionnées à partir des domaines centraux décrits dans le curriculum-cadre qui sont censés être pertinents pour *tous* les groupes. L'autre mesure principale concerne les types d'experts et les personnes ressources qui ont été impliquées lors de la conception du test, notamment des enseignants pour le groupe cible, des formateurs d'enseignants et des auteurs de manuels. La responsabilité de ces professionnels était de vérifier qu'aucun des groupes ne soit désavantagé par le choix de sujets et de contenus inappropriés, des formats d'items ou des éléments pour l'administration du test. Les informations obtenues de la part de ces experts

ont été confrontées aux résultats provenant des analyses statistiques lors des séries d'essais sur le DTZ.

3.8 Cohérence des résultats du test

Le degré de cohérence de l'information relative aux capacités qui est obtenue au moyen de l'évaluation détermine à quel point les conclusions peuvent être fiables. Idéalement, les données ne contiennent pas de variation systématique due à des différences dans l'administration du test, dans la notation ou des groupes de candidats. Les problématiques abordées précédemment sous l'égide de la fiabilité (*reliability*) tombent sous ce point. Ainsi, il est important pour les concepteurs de test de démontrer que les résultats du test sont cohérents. Les tests formels sont généralement plus fiables que les évaluations informelles et, ainsi qu'il l'a été mentionné au chapitre 3.5, même des auteurs plutôt sceptiques envers les tests apprécient l'équité accrue de leur usage (p.ex. Balch et al., 2008).

Si l'on considère la longue tradition dont jouissent les analyses de fiabilité, il n'est pas surprenant de voir que des éléments en rapport avec la cohérence sont présentés de manière exhaustive dans De Jong et al. (2009), Perlmann-Balme (2011) et Plassmann (2011).

De Jong et al. (2009) présentent leur système automatisé d'administration du test et de notation, y inclus la banque d'items, comme un facteur important pouvant contribuer à générer des données de test cohérentes; son avantage réside dans le fait que la banque de données réduit les possibilités qu'un individu humain puisse introduire des perturbations ou des biais. Cependant, le grand défi pour les concepteurs du TGN consiste à démontrer que les utilisateurs ont des bonnes raisons de croire que le système informatisé fonctionne de manière fiable. Une grande partie de leur article est consacré à cela précisément. De Jong et al. (2009) fournissent des preuves de corrélations élevées entre la notation humaine et la notation informatisée. Ils montrent également que dans la partie plus basse de l'échelle des niveaux où la limite réussi-échoué se situe, la mesure de la capacité

en question est possible avec une toute petite erreur statistique. L'argumentaire des concepteurs est renforcé au moyen de nombres impressionnants de personnes et d'items à l'essai qui ont généré les données pour les analyses.

Les concepteurs du test allemand DTZ utilisent les standards établis par ALTE comme un élément structurant dans la présentation de leur travail sur 'l'argumentaire de fiabilité' (Perlmann-Balme, 2011).

Le standard ALTE n°3 concerne la bonne pratique dans l'élaboration de tests :

3) Vous fournissez des critères pour la sélection et la formation des concepteurs de tests et vous sollicitez l'avis d'experts pour la conception des tests et la révision des examens.¹⁸

Dans leur rapport sur la conception du DTZ, les auteurs insistent sur la formation des rédacteurs et réviseurs d'items, et sur leurs méthodes de développement interactives et réalisées pas à pas qui combinent des séries d'essais (phases pilotes du test), des commentaires d'enseignants et des analyses statistiques jusqu'à ce que le matériel du test puisse être produit avec fiabilité.

Le standard ALTE n°4 concerne l'équivalence des examens administrés à différentes occasions :

4) Des examens parallèles, bien qu'administrés par différents centres, sont comparables en termes de contenu, de stabilité, de cohérence et de notation.¹⁹

Perlmann-Balme (2011) fait référence à la banque d'items codée comme base technique qui permet de garantir la comparabilité des contenus du test à travers les différentes versions, tandis que des règles d'administration du test permettent d'assurer la cohérence organisationnelle à travers les différentes passations. Le manuel de préparation au test (Perlmann-Balme et al., 2009) qui contient une transcription d'un examen oral joue un rôle important dans la standardisation du déroulement et de

18 | Version française originale (cf. ALTE [Association of Language Testers in Europe], 2007b).

19 | Version française originale (cf. ALTE [Association of Language Testers in Europe], 2007b).

l'évaluation de cet examen. Perlmann-Balme (2011) et Plassmann (2011) soulignent toutes les deux la bonne qualité des procédures qui ont été développées pour évaluer des textes écrits : tous les textes écrits sont notés par la même institution, ce qui, selon Plassmann, a pour avantage que les problèmes puissent être traités directement par les bonnes personnes lorsqu'ils se posent. Afin d'assurer des notations cohérentes, trois mesures concrètes ont été mises en place : la formation de l'examineur, une double notation, des analyses ultérieures de la notation afin de détecter des jugements radicaux.

Le standard ALTE n°14 exige une analyse statistique des items :

14) Les données concernant les niveaux des items (p. ex. pour calculer la difficulté, la discrimination, la fiabilité et les erreurs standards d'évaluation de l'examen) sont collectées à partir d'un échantillon adéquat de candidats et analyses.²⁰

Perlmann-Balme (2011) écrit très brièvement que les analyses psychométriques standards évoquées dans le standard 14 ont été appliquées. D'autre part, des items ont été éliminés sur la base de ces résultats.

Ce bref chapitre sur la cohérence des résultats du test montre de nouveau, ainsi que nous l'avons vu pour les chapitres précédents, que les publications ayant émergé de projets à enjeux élevés aux Pays-Bas et en Allemagne abordent des éléments centraux des procédures habituelles de validation des tests. L'article de De Jong et al. (2009), orienté vers la recherche, procède d'une manière différente des auteurs écrivant sur le DTZ allemand. Tandis que De Jong et al. (2009) font preuve d'une volonté de démontrer les mesures prises pour assurer une validité suffisante, Perlmann-Balme (2011) et Plassmann (2011) tendent à simplement déclarer que des standards de qualité ont été adoptés et que les bonnes procédures ont été suivies. Cette différence peut entièrement être due au style et au public visé par les revues qui ont publié ces articles. Du point de vue d'un lecteur expert, un

retour critique sur les articles déjà existants qui traitent du DTZ serait souhaitable.

Le chapitre 3 en général s'intéresse à la littérature couverte dans la présente étude qui traite des questions de validité et validation de l'évaluation. Les aspects qui ont trait à la validité qui ont été examinés avec plus d'attention font partie du modèle de validation de l'évaluation (AUA) développé par Bachman & Palmer (2010). Nos analyses révèlent que la plupart des auteurs se concentrent sur seuls quelques aspects en rapport avec la validité et la validation. L'attention semble être portée sur des problématiques fondamentales et des considérations socio-politiques en rapport avec l'usage de tests à des fins d'immigration et d'intégration. Ces questions sont abordées seulement partiellement par un modèle de validation actualisé s'orientant vers l'usage des tests et leurs impacts, tel que l'AUA. Les discussions au sujet d'idéologies implicites dans les constructs (p. ex. 'une nation – une langue') ou des réflexions en rapport avec la *justice* ou l'éthique des tests pour la population migrante se situent au-delà de la portée du cadre conceptuel pour la validation de l'évaluation – et aussi au-delà du domaine d'expertise des concepteurs de tests de langue.

Un critère de validité, la 'pertinence des conclusions au regard du construct', souvent traité en tant que *validité de construct*, obtient l'attention des concepteurs de tests, de même que des auteurs sceptiques envers les tests. Alors que les experts critiquant les tests interrogent, par exemple, le lien entre les contenus du test et l'usage de la langue cible en situation authentique, les concepteurs de test font un effort pour démontrer ou expliquer la pertinence de leurs tests.

Nous avons été surpris de constater que peu de publications dans le domaine de l'évaluation linguistique à des fins d'immigration et d'intégration pouvaient effectivement être considérées comme des études de validation. Les véritables études de validation sont soit trop peu nombreuses, soit elles ne sont pas rendues accessibles à un public de spécialistes de l'évaluation indépendants.

20 | Version française originale (cf. ALTE [Association of Language Testers in Europe], 2007b).

4

Résumé et conclusions

Cette étude offre une vue d'ensemble des développements récents dans le domaine de l'évaluation linguistique à des fins d'immigration et d'intégration, des investigations concernant la validité des tests et les discussions ayant lieu dans ce contexte, et implicitement, les problématiques en jeu et les arguments mobilisés. Les commentaires conclusifs et les propositions sont formulés sur la base des informations récoltées.

Nos résultats font apparaître plusieurs changements rapides et plutôt drastiques dans le domaine de l'évaluation linguistique à des fins d'immigration et d'intégration en Europe, ainsi que dans d'autres pays, la plupart des pays 'occidentaux', au cours de la dernière décennie. Nous avons identifié une tendance à introduire, de manière croissante, des prérequis linguistiques aux différentes étapes du processus d'immigration et d'intégration – dans certains pays, ces prérequis linguistiques commencent bien avant l'immigration en tant que telle – et les prérequis sont devenus plus étendus, avec une tendance à utiliser une forme d'évaluation plus formelle, souvent au moyen de tests standardisés. Dans certains pays, l'évaluation ne concerne pas seulement les connaissances linguistiques mais aussi culturelles. Typique pour le contexte européen, les nouvelles conditions ne s'appliquent pas de manière égale à tous les groupes, en raison de régulations spécifiques à l'UE.

Mis à part les tendances générales communes, nous avons pu identifier une diversité considérable dans la conception et la définition des prérequis linguistiques et des évaluations y afférentes. Les éléments de variation sont : le statut des cours de langue et d'intégration préparatoires, l'exemption de certains groupes, l'âge minimal des candidats, le développement du test et son administration, les méthodes de test, les compétences testées et le niveau de compétence requis. Par exemple, le niveau de langue requis pour l'admission dans un pays se situe entre les niveaux A1 et B1 sur l'échelle

des niveaux du CECR; le niveau de langue requis varie entre A1 et B2 pour le séjour permanent et la nationalité. Les arguments avancés pour justifier le choix du niveau, cependant, sont souvent très semblables.

Nous avons utilisé le modèle exhaustif et largement reconnu du *Assessment Use Argument* (AUA) de Bachman et Palmer (Bachman & Palmer, 2010) pour résumer la discussion autour des problématiques se rapportant à la validité dans la littérature sur l'évaluation linguistique à des fins d'immigration et d'intégration. Ce modèle de validation pour l'évaluation linguistique est exhaustif dans la mesure où il prend en compte et articule sur le plan conceptuel tous les aspects d'un test opérationnel qui pourraient être pertinents pour la validité du test – allant des tâches et de l'administration du test à son impact sur les individus concernés ainsi que sur la société dans son ensemble.

Les publications consultées pour la présente étude traitent de manière inégale les différents thèmes relatifs à l'usage valide de tests. Seules quelques publications ont en réalité l'objectif de démontrer de manière exhaustive (ou de remettre en question de manière systématique) la validité de tests spécifiques pour l'immigration. Celles qui se concentrent sur un test particulier ont été rédigées par ses concepteurs afin d'en justifier l'usage. Dans l'une de ces publications (De Jong et al., 2009), les auteurs procèdent de manière très systématique dans leur effort de fournir assez d'éléments de preuves de validité pour justifier l'usage du Test de néerlandais parlé (TGN). Il est évident que les auteurs suivent un modèle de validation basé sur un argumentaire tel que celui qui est présenté par Bachman & Palmer (2010), bien que cela ne soit jamais vraiment exprimé explicitement. La plupart des publications dans notre sélection traitent soit de la pertinence (*meaningfulness*) des contenus des tests et des niveaux, des retombées positives (*beneficence*) d'un test pour les personnes concernées ou soulèvent des questions fondamentales sur l'usage même de tests pour les migrants qui peuvent être difficilement intégrées dans les modèles de validation existants.

Plusieurs auteurs restent sceptiques quant à la pertinence des conclusions basées

sur les résultats de test de langue (et/ou de connaissances culturelles), dans la mesure où ces auteurs ont relevé ou soupçonnent un manque de correspondance entre

- les contenus du test et les besoins et les tâches de la vie quotidienne;
- le construct déclaré du test (p. ex. ‘des connaissances linguistiques suffisantes pour l’intégration sociale’) et les compétences qui sont effectivement évaluées (et qui sont déterminées de manière plus ou moins arbitraire);
- le niveau de compétence requis pour le test et pour les activités quotidiennes authentiques;
- la langue évaluée et les langues nécessaires et utilisées par les migrants dans les contextes sociaux multilingues dans lesquels ils vivent et travaillent;
- ou plus fondamentalement, la réussite des tests de langue et/ou de connaissances culturelles et être un bon citoyen.

Lorsque des régimes de tests (plus stricts) sont instaurés, les autorités affirment souvent que ces tests auront des *retombées bénéfiques*, pour les migrants eux-mêmes notamment. Cet aspect est abordé par un certain nombre d’études, souvent de manière purement anecdotique, mais aussi sur la base d’études d’impact véritables telles que l’étude INTEC (Strik et al., 2010). Des appréciations négatives de l’impact bénéfique du test prédominent dans les publications étudiées. L’obligation de passer un test (plus strict) afin de pouvoir faire un pas dans le processus d’immigration est souvent considérée comme étant inutile, générant du stress pour les candidats, les décourageant ou même excluant certains groupes (p. ex. des personnes peu instruites) d’une intégration réussie. Une appréciation positive peut être relevée en ce qui concerne les cours préparatoires qui accompagnent parfois les tests de langue et de connaissances culturelles; le bénéfice des cours réside dans le fait qu’ils offrent l’opportunité (obligatoire) de sortir des limites de sa propre communauté et d’entrer en contact avec d’autres personnes. L’étude INTEC, qui se base sur des données

statistiques et des entretiens provenant de neuf pays européens, corrobore d’autres observations plus informelles qui mettent en évidence l’effet dissuasif des nouvelles mesures et des effets très peu bénéfiques sur les individus.

Des experts critiques dans le domaine de l’évaluation linguistique tels que Shohamy ou McNamara apportent un éclairage plus radical sur la discussion sur l’évaluation linguistique à des fins d’immigration et d’intégration en posant des questions *fondamentales* (des questions de ‘justice’, dans les termes de McNamara et ses collègues) qui concernent d’autres domaines, en particulier les sciences sociales (critiques) et les sciences politiques, la philosophie (l’éthique) et le droit constitutionnel. En identifiant des idéologies implicites (‘une nation – une langue’), des ‘agendas cachés’ (les tests de langue pour la régulation sociale), des motivations politiques (assurer la cohésion sociale à travers l’exclusion), les conséquences des tests qui sont problématiques sur le plan éthique (l’exclusion des possibilités de participation dans la société d’accueil), ces experts critiques incitent à un débat général qui impliquerait des disciplines qui n’ont que rarement été consultées pour la conception et l’implémentation de tests de langue et de connaissances culturelles. Par conséquent, des experts dans ces disciplines devraient être impliqués bien avant qu’un test soit mandaté, dans le cas où un tel test serait introduit.

Quelles conclusions pouvons-nous tirer de ces observations? Quelles pistes peuvent être dégagées pour le futur?

Il est souhaitable que les questions soulevées par les auteurs critiques quant à l’usage de tests soient prises en compte dans les discussions guidant les politiques linguistiques et les pratiques d’évaluation, dans la mesure où les objections formulées par ces auteurs sont fondamentales. Nous pensons, cependant, qu’aucun changement ne sera initié sans un effort considérable des experts des disciplines qui ne traitent pas habituellement des prérequis linguistiques et des tests de langue. Les décisions concernant les prérequis linguistiques et les tests de langue à des fins d’immigration et d’intégration sont souvent prises

par les pouvoirs législatifs et leur implémentation est mandatée et supervisée par les autorités publiques. Ces organes ne sont pas obligés d'inviter des experts critiques pour débattre et établir des 'questions de valeur sociale' tel que le suggèrent McNamara & Ryan (2011). Les gouvernements se sont intéressés à ces problématiques lorsque des droits exécutoires ont été (possiblement) violés. A la place d'une action gouvernementale, la communauté scientifique devrait entreprendre les démarches nécessaires pour élargir la portée de la discussion de manière autonome et amener des sujets cruciaux au cœur du débat. Afin d'avoir un impact suffisant, il apparaît nécessaire d'impliquer les disciplines qui traitent habituellement des problématiques sociales, juridiques et éthiques qui sont en jeu dans le présent cas de figure.

Un aspect qui se dessine clairement est la potentielle utilité d'un modèle de validation tel que le *Assessment Use Argument* (AUA) de Bachman & Palmer. D'un point de vue professionnel, les concepteurs de test ne peuvent faire autrement qu'accepter la responsabilité et la transparence de leurs travaux comme principe directeur ou 'axiome' (cf. Bachman & Palmer, 2010). Le modèle de validation opérationnalise la démonstration et la transparence afin d'alimenter la recherche sur la conception de test et aussi pour assurer un usage avisé des tests. Il s'agit d'attirer davantage l'attention des autorités, qui mandatent potentiellement des tests, sur l'existence et la portée de modèles actuels tels que le AUA, afin de leur permettre de formuler des appels d'offre et des critères d'évaluation de projets en conséquence et de se préparer à attribuer les ressources temporelles et financières nécessaires. De plus, des modèles de validation exhaustifs qui couvrent l'ensemble du sujet, allant de la préparation d'un candidat pour le test à son impact sur la société font apparaître, en particulier pour ceux qui mettent les régimes de test en place et qui en sont responsables, que la validation est un processus permanent qui s'étend au-delà de la vie d'un test. Ainsi, l'idée d'un processus de développement unique ne peut être soutenue.

Ainsi que nous l'avons mentionné plus haut, la pertinence des conclusions basées sur les résultats des tests que les migrants doivent

passer est souvent contestée dans la littérature étudiée. En particulier, certains auteurs soulignent que les capacités mesurées et les capacités nécessaires pour fonctionner dans la société et y contribuer n'ont que peu de choses à voir ensemble. La meilleure manière d'établir ce lien – et aussi pour surmonter les attentes en ce qui concerne la compétence linguistique qui seraient basées sur des idéologies – est par une analyse approfondie, et empiriquement fondée, des besoins (*needs analysis*²¹). L'analyse des besoins était l'une des méthodes mises en œuvre avec succès dans le projet suisse *fide* (Müller & Wertenschlag, 2013). La démarche n'incluait pas seulement des migrants de différents groupes sociaux et ethniques mais aussi des membres de la société d'accueil qui sont en contact régulier avec des migrants ayant une compétence linguistique limitée. Une problématique qui devrait être prise en compte est l'interprétation adéquate des résultats de l'analyse des besoins. Ceux-ci peuvent montrer par exemple que le besoin de connaître la langue locale (à un niveau spécifique) diverge fortement en fonction de l'appartenance à un groupe; cela ne ferait pas de sens d'exiger de certains individus d'acquérir des compétences écrites; ou que certains migrants auraient en réalité besoin d'avoir une très bonne maîtrise de la langue locale pour pouvoir contribuer à la société en fonction de leurs capacités. Un système d'évaluation et de certification qui prenne en compte des types d'information différents pourrait opter pour une approche modulaire de l'évaluation linguistique, par exemple, ce qui permettrait de répondre aux besoins spécifiques et aux capacités différenciées. En gardant l'intégration à l'esprit, il serait peut-être une bonne idée de commencer par concevoir un encouragement linguistique optimal, pour ensuite aller vers la certification. Le système *fide*, qui est soutenu par les autorités suisses (Gerber & Schleiss, 2013) cherche à répondre à la diversité en proposant des ajustements aux différents besoins et profils des candidats et en intégrant l'évaluation linguistique dans un système d'en-

21 | L'analyse des besoins langagiers peut être étendue à une série d'audits linguistiques qui prennent en compte, au sein des contextes pertinents, non seulement les besoins des différents acteurs, mais aussi leurs compétences langagières plurilingues.

couragement linguistique²².

Notre étude fait également apparaître un besoin pour des études d'impact couvrant un éventail plus large de problématiques dans divers contextes. De nombreux articles laissent entrevoir des incertitudes quant à l'impact effectif des prérequis linguistiques et des tests de langue. Parfois, les spéculations et les anecdotes remplacent des preuves bien fondées. L'étude INTEC (Strik et al., 2010) a fait un travail considérable en articulant des données quantitatives et qualitatives pour identifier l'impact des régimes de test dans neuf pays européens ; des études à plus petite échelle fournissent des connaissances supplémentaires concernant d'autres contextes. Ces efforts doivent être poursuivis et étendus à d'autres contextes et questions de recherche. Dans les modèles de validation actuels tels que le modèle AUA - qui conçoit l'usage des tests comme phénomène social et qui attendent d'un test de profiter aux acteurs concernés et à la société en général - les études d'impact sont un élément constitutif d'une démarche exhaustive de validation de test qui peut contribuer de manière importante à justifier le bien-fondé de son usage.

En guise de recommandation finale, nous tenons à souligner le fait qu'il est hautement souhaitable que les tests à des fins d'immigration et d'intégration soient validés non seulement de manière systématique, mais aussi transparente. Dans notre collection de littérature publiée, un seul article (De Jong et al., 2009) vise réellement à fournir un argumentaire de validation exhaustif en abordant un grand nombre de points essentiels et en offrant des arguments et des éléments de preuve. Il faut noter cependant que cet article a une visée confirmative²³. Comme on peut s'y attendre, l'article a été publié alors que

le test était déjà opérationnel, car, ainsi que Briggs le remarque dans son commentaire sur l'appel de Kane (Kane, 2004) pour une démarche argumentative interprétative *explicite* (semblable au modèle AUA de Bachman et Palmer), il pourrait se poser un problème 'paradoxal' ayant trait à la vie quotidienne qui ne peut pas être surmonté facilement : si un test n'a pas été entièrement validé, alors, son usage dans le but de prendre des décisions dont les enjeux sociaux sont élevés devient problématique. Si le test n'a pas été administré à un échantillon de la population cible pour le but visé, il est alors impossible de le valider entièrement (Briggs, 2004 : 172). Nous pensons que ce 'paradoxe' n'exempte pas les concepteurs de test et ceux qui les utilisent et/ou en sont responsables de mener des études de validation et de les rendre disponible à un public d'experts indépendants ; cela devrait déboucher tout au plus sur une publication légèrement différée. Un prochain pas souhaitable, maintenant que des modèles de validation de bonne qualité et facilement applicables sont disponibles, est de créer une 'culture de validation des tests', dans laquelle la publication de travaux de validation sur des tests opérationnels constitue la norme et non pas l'exception. De même, dans le but de contrecarrer le biais introduit par les auteurs qui sont eux-mêmes concepteurs de test, des spécialistes des tests de langue externes devraient être mandatés pour analyser les éléments sensibles dans un système de test (y inclus son impact) et avoir accès aux processus internes et aux données confidentielles si nécessaire. Un engagement pour la validité des tests, un degré élevé de transparence - et bien évidemment des résultats positifs provenant d'études de validation - pourraient augmenter la crédibilité et la légitimité d'une institution - voire même des autorités publiques - à évaluer les compétences linguistiques dans un domaine aussi délicat que celui de l'immigration.

22 | De manière semblable, Schneider et al. (2006) recommandent de définir deux profils en ce qui concerne l'évaluation linguistique dans le contexte de la naturalisation en Suisse : un profil d'encouragement linguistique et un profil d'évaluation.

23 | Chapelle et al. (2008) est un autre rapport notable rédigé par plusieurs de ses développeurs, dans un contexte différent, notamment celui du TOEFL iBT, un test relativement récent pour les étudiants qui projettent de rejoindre une université anglophone. Les auteurs utilisent un modèle de validation qui s'inspire de Kane et Bachman & Palmer.

5

Bibliographie

A

Adami, H., & Leclercq, V. (Éd.). (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 [Association of Language Testers in Europe] (Éd.). (1994). Le Code de bonne pratique de ALTE. Consulté à l'adresse http://www.alte.org/attachments/files/code_practice_fr.pdf

ALTE [Association of Language Testers in Europe]. (2001). Principles of good practice for ALTE examinations. Revised draft. Consulté à l'adresse http://www.alte.org/attachments/files/good_practice.pdf

ALTE [Association of Language Testers in Europe] (Éd.). (2007). Normes minimales relatives à l'établissement de profils qualité pour les examens de ALTE. Consulté à l'adresse http://www.alte.org/attachments/files/minimum_standards_fr.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 [etc.] : 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 [etc.] : Oxford University Press.

Balch, A., Corrigan, M., Gysen, S., Kuijper, H., Perlmann-Balme, M., Roppe, S., ... Zeidler, B. (2008). *Tests de langue en faveur de la cohésion sociale et de la citoyenneté : descriptif à l'intention des décideurs*. Strasbourg : Conseil de l'Europe. Consulté à l'adresse http://www.coe.int/t/dg4/linguistic/liam/Source/Events/2008/ALTE_migrants08_final_FR.pdf

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 (Éd.), *Discourses on language and integration : critical perspectives on language testing regimes in Europe* (p. 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 (Éd.), *Language testing, migration and citizenship : cross-national perspectives on integration regimes* (p. 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 (Éd.). (2006). *Educational measurement* (4th edition). 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. Consulté à l'adresse <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. (Éd.). (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York ; London : Routledge.

Conseil de l'Europe (Éd.). (2001). *Cadre européen commun de référence pour les langues*. Paris : Editions Didier.

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 (Éd.), *From migrant to citizen : testing language, testing culture* (p. 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. Consulté à l'adresse <http://www.ets.org/s/about/pdf/standards.pdf>

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

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

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

Extra, G., Spotti, M. & Van Avermaet, P. (2009b). Testing regimes for newcomers. In G. Extra, M. Spotti, & P. Van Avermaet (Éd.), *Language testing, migration and citizenship : cross-national perspectives on integration regimes* (p. 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 (Éd.), *Les migrants face aux langues des pays d'accueil : acquisition en milieu naturel et formation* (p. 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 : Council of Europe. Consulté à l'adresse 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 (Éd.), *From migrant to citizen : testing language, testing culture* (p. 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 (Éd.), *From migrant to citizen : testing language, testing culture* (p. 101–124). Basingstoke : Palgrave Macmillan.

Hogan-Brun, G., Mar-Molinero, C. & Stevenson, P. (Éd.). (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 (Éd.), *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 (Éd.), *From migrant to citizen : testing language, testing culture* (p. 39–59). Basingstoke : Palgrave Macmillan.

I

ILTA [International Language Testing Association]. (2007). Guidelines for practice. Consulté à l'adresse 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 (Éd.), *Educational measurement* (4th edition., p. 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. *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 (Éd.), *Exploring language frameworks : proceedings of the ALTE Kraków Conference, July 2011* (p. 389–404). Cambridge : Cambridge University Press.
- Kostakopoulo, D. (2010). Introduction. In R. van Oers, E. Ersbøll & D. Kostakopoulo (Éd.), *A re-definition of belonging? Language and integration tests in Europe* (p. 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:10.2307/4626095
- 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 (Éd.), *The Routledge handbook of language testing* (p. 162–177). London; 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). *Curriculum-cadre pour l'encouragement linguistique des migrants*. Berne : Office fédéral des migrations ODM. Consulté à l'adresse <https://www.bfm.admin.ch/content/dam/data/migration/integration/berichte/rahmencurriculum-f.pdf>
- Linn, R. L., National Council on Measurement in Education & American Council on Education (Éd.). (1989). *Educational measurement* (3rd edition). New York; London : Macmillan Publishers.
- Little, D. (2010). L'intégration linguistique des migrants adultes : évaluation des politiques et des pratiques. Strasbourg : Conseil de l'Europe, Division des Politiques linguistiques. Consulté à l'adresse http://www.coe.int/t/dg4/linguistic/liam/Source/Events/2010/2010_evaluatingpolicy_fr.pdf
- M
- Mackenzie, C. (2010). Citizenship, identity, and immigration : contemporary philosophical perspectives. In C. Slade & M. Möllering (Éd.), *From migrant to citizen : testing language, testing culture* (p. 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 (Éd.), *Language ideologies, policies and practices : language and the future of Europe* (p. 76–88). Basingstoke : Palgrave Macmillan.
- Mar-Molinero, C. & Stevenson, P. (Éd.). (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 (Éd.), *Discourse on language and integration* (p. 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 (Éd.), *Educational measurement* (3rd ed., p. 13–103). New York ; London : 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. Ersbøll & D. Kostakopoulou (Éd.), *A re-definition of belonging? Language and integration tests in Europe* (p. 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. Consulté à l'adresse <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 (Éd.), *From migrant to citizen : testing language, testing culture* (p. 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 (Éd.), *From migrant to citizen : testing language, testing culture* (p. 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 (Éd.), *Language testing matters : investigating the wider social and educational impact of assessment - proceedings of the ALTE Cambridge Conference April 2008* (p. 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 (Éd.), *A re-definition of belonging? Language and integration tests in Europe* (p. 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 (Éd.), *A re-definition of belonging? Language and integration tests in Europe* (p. 307–331). Leiden ; Boston : Martinus Nijhoff Publishers.

Oers, R. van, Ersbøll, E. & Kostakopoulo, D. (Éd.). (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 (Éd.), *Language testing matters : investigating the wider social and educational impact of assessment - proceedings of the ALTE Cambridge Conference April 2008*. 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. Consulté à l'adresse 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 (Éd.), *Multilingualism and assessment : achieving transparency, assuring quality, sustaining diversity : proceedings of the ALTE Berlin conference, May 2005* (p. 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. Consulté à l'adresse 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 ; New York : Routledge.

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

Shohamy, E. (2009). Language tests for immigrants : Why language? Why tests? Why citizenship? In G. Hogan-Brun, C. Mar-Molinero & P. Stevenson (Éd.), *Discourses on language and integration : critical perspectives on language testing regimes in Europe*. (p. 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. *Babylonia*, (1), 14–18.

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

Slade, C. (2010b). Shifting langscapes of citizenship. In C. Slade & M. Möllering (Éd.), *From migrant to citizen : testing language, testing culture* (p. 3–23). Basingstoke : Palgrave Macmillan.

Slade, C. & Möllering, 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 (Éd.), *Language testing, migration and citizenship : cross-national perspectives on integration regimes* (p. 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 (Éd.), *Discourse on language and integration* (p. 15–43). Amsterdam; Philadelphia : John Benjamins Publishing Company.

Van Avermaet, P. (2012). L'intégration linguistique en Europe : analyse critique. In H. Adami & V. Leclercq (Éd.), *Les migrants face aux langues des pays d'accueil : acquisition en milieu naturel et formation* (p. 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 (Éd.), *Exploring language frameworks : proceedings of the ALTE Kraków Conference, July 2011* (p. 11–44). Cambridge : Cambridge University Press.

W

Wright, S. (Éd.). (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.

