



COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

Language Policy Division
Division des Politiques Linguistiques

III. Tests de langue en faveur de la cohésion sociale et de la citoyenneté : descriptif à l'intention des décideurs

Groupe d'auteurs ALTE :

Andrew Balch

Shalini Roppe

Michael Corrigan

Heinrich Rübeling

Sara Gysen

Steffi Steiner

Henk Kuijper

Piet Van Avermaet

Michaela Perlmann-Balme

Beate Zeidler

Avec la contribution significative du Groupe d'intérêt spécial sur l'évaluation des compétences linguistiques dans le cadre des migrations et de l'intégration.

1.	Introduction	2
2.	Que tester ?	4
2.1	<i>Généralités</i>	4
2.2	<i>Déterminer l'objectif du test et les besoins des candidats dans le monde réel</i>	4
2.3	<i>Déterminer les demandes linguistiques</i>	5
2.4	<i>Déterminer le niveau de difficulté adéquat</i>	5
2.5	<i>Produire des spécifications pour les tests</i>	7
3.	Application pratique des spécifications de test	7
3.1	<i>Généralités</i>	7
3.2	<i>Critères d'évaluation et format des tests</i>	8
3.3	<i>Rédaction des items</i>	8
3.4	<i>Pré-évaluation des items</i>	8
3.5	<i>Passation des tests</i>	9
3.6	<i>Participation des candidats présentant des besoins spéciaux</i>	9
3.7	<i>Notation</i>	10
3.8	<i>Contrôle</i>	10
3.8.1	<i>Contrôle des réponses des candidats</i>	10
3.8.2	<i>Contrôle de l'impartialité</i>	11
4.	Conclusion	11

1. Introduction

De nombreux pays européens soumettent la migration, la résidence et la citoyenneté à certaines conditions linguistiques et, souvent, exigent de recourir à des tests linguistiques ou à d'autres procédures d'évaluation. Dans ce chapitre, notre intention n'est pas de promouvoir l'usage de ces tests linguistiques mais, là où des examens obligatoires sont déjà en place ou en voie de l'être, de proposer des conseils professionnels fondés sur de bonnes pratiques d'évaluation afin d'assurer la prise en compte des besoins des parties prenantes et l'équité des tests. En effet, l'équité des tests est particulièrement importante dès lors qu'ils concernent la migration, la résidence ou la citoyenneté. Ils mettent en jeu les droits civiques ou humains de migrants. Un certain nombre de standards (voir la section *Documents de référence*) fournissent une orientation pour l'élaboration et la passation de tests équitables. En complément du reste de ce chapitre, elles montrent comment divers éléments du processus d'évaluation s'inscrivent dans un cadre éthique.

Pour une évaluation linguistique, les décideurs doivent d'abord examiner certains points en détail :

- Quelle est la solution la plus adéquate : un test ou une autre forme d'évaluation ?
- Est-il souhaitable d'utiliser une combinaison de plusieurs méthodes d'évaluation ?
- Comment les résultats des tests seront-ils utilisés ?
- Quelles seront les conséquences d'un test pour la société ?
- Quelle sera l'incidence pour le migrant ?
- Quelle sera l'incidence pour la société du migrant ?

Mode d'évaluation

S'agissant des premier et deuxième points, les décideurs doivent être conscients qu'il existe d'autres formes d'évaluation qui, elles aussi, peuvent convenir. Tests et autres méthodes d'évaluation offrent chacun leurs propres avantages, notamment quant à l'incidence sur le candidat, aux possibilités d'interprétation et à la fiabilité des résultats, à la standardisation, au coût et aux applications pratiques. Aussi est-il important d'examiner attentivement les conditions précises de la situation afin de déterminer le type d'évaluation le mieux adapté. À noter, également, la possibilité de combiner plusieurs méthodes d'évaluation. Nous mentionnons, ci-après, quelques-uns des avantages des tests et autres formes d'évaluation.

S'ils sont correctement conçus, élaborés et appliqués, les tests présentent les avantages suivants :

- Les résultats offrent cohérence et fiabilité ; autrement dit, différentes administrations peuvent facilement comparer les candidats.
- Les candidats sont évalués en toute indépendance et objectivité.
- Il est possible de faire passer des tests à un grand nombre de candidats en peu de temps.

Les tests peuvent aussi prendre la forme d'une évaluation continue tout au long d'un cours, ou de l'évaluation d'une série de justificatifs présentés par le candidat¹. Si l'évaluation doit avoir une réelle valeur formative, elle peut être effectuée dans le cadre du cours ; auquel cas l'apprenant interviendra plus directement dans l'orientation de l'apprentissage. L'évaluation peut aussi comprendre une autoévaluation et/ou une évaluation mutuelle entre apprenants.

¹ Voir le chapitre I: « Le Cadre européen commun de référence pour les langues et l'élaboration de politiques en faveur de l'intégration des migrants adultes », David Little

Outre le potentiel formatif de ce type d'évaluation (comme indiqué plus haut), cette approche présente plusieurs avantages :

- Les documents servant à l'évaluation peuvent être recueillis dans des circonstances neutres (en classe, par exemple), ce qui peut renforcer leur validité comme preuve des capacités du candidat.
- Les tâches à accomplir sont très proches des activités rencontrées dans le monde réel – il peut s'agir de tâches jugées particulièrement importantes pour des migrants et des citoyens.
- Les performances du candidat sont évaluées de manière plus globale et, par conséquent, visent davantage ses capacités sous-jacentes à réussir telle ou telle tâche – au lieu d'évaluer des éléments linguistiques isolés.

Impact

En ce qui concerne les conséquences d'un test, l'on part du principe que tous les résultats contiennent une certaine marge d'erreur, aucune évaluation ne pouvant prétendre à la perfection. Le but de ce chapitre est, par conséquent, d'aider à *minimiser* la possibilité de conséquences négatives dues à l'utilisation des tests, notamment en réduisant la possibilité d'erreur sans viser à l'éliminer totalement. Par ailleurs, les tests à grande échelle ne peuvent pas facilement prendre en compte la personnalité, le parcours d'apprentissage ni l'histoire personnelle des candidats en évaluant les compétences. Si, contre toute attente, le candidat a plus ou moins échoué à un test, ses capacités réelles seront sous-estimées. Aussi la valeur générale d'un test doit-elle être relativisée et envisagée par rapport aux conséquences d'un échec, car certaines des décisions fondées sur les résultats du test peuvent procéder d'informations inexacts.

Une fois le mode d'évaluation choisi, il est très important de réfléchir à l'utilisation et aux conséquences, car l'utilisation peut avoir des conséquences graves et inattendues. Lors de la planification de l'évaluation, il est recommandé d'examiner attentivement les conséquences possibles et, durant le processus d'évaluation, d'explorer quelles sont les conséquences réelles. Parmi les conséquences possibles, citons un changement des pratiques d'enseignement et d'apprentissage à la suite du test, ou un changement du système éducatif dans le pays du migrant.

Après considération des points ci-dessus mentionnés, si un test linguistique est jugé nécessaire, les parties prenantes – notamment les décideurs – doivent s'assurer (i) que le test répond bien au(x) besoins identifiés et (ii) qu'il permet d'appliquer correctement et en toute équité la politique associée. Si l'équité des tests concerne tous les types de tests linguistiques et tous les candidats, reste que, comme nous l'avons dit plus haut, elle revêt une particulière importance dans les cas de migration, de résidence et de citoyenneté du fait des graves implications pour les candidats en termes de droits civiques et humains. Les efforts déployés pour veiller à l'équité d'un test doivent commencer dès la phase de planification et se poursuivre tout au long de l'application du test. Ces efforts permettront aux utilisateurs du test de bien interpréter et exploiter les résultats. Pour aider les décideurs dans cette tâche, le reste du chapitre fournit des informations sur toutes les étapes du test, de l'élaboration à l'application ; ces informations serviront soit à sélectionner/superviser un fournisseur de tests et à interpréter des résultats, soit à orienter la production interne de tests. Comme avec tous les tests, le résultat de l'application de la bonne pratique d'évaluation décrite dans ce chapitre permettra d'assurer que non seulement les compétences et connaissances adéquates sont testées (ce qui fait la validité du test), mais aussi que cette évaluation est la même pour tous les candidats et pour toutes les versions du test (ce qui fait la fiabilité du test). Dans ce

chapitre, on trouvera également des références à d'autres ressources pouvant être utiles à ce travail.

Documents de référence :

[Code de bonne pratique ALTE](http://www.alte.org/quality_assurance/code/code_practice_fr.pdf) – http://www.alte.org/quality_assurance/code/code_practice_fr.pdf

Multilingual Glossary of Language Testing Terms, Studies in Language Testing, volume 6, Cambridge University Press (ISBN: 0-521-65877-2)

ILTA, *Code of Ethics* – <http://www.iltaonline.com/code.pdf>

JCTP, *Code of Fair Testing Practice in Education* – <http://www.apa.org/science/FinalCode.pdf>

2. Que tester ?

2.1 Généralités

Cette section indique les étapes à suivre pour assurer l'adéquation entre la conception du test et l'objectif visé par le test. La première étape du processus consiste à identifier avec précision et sans ambiguïté l'objectif du test. Ensuite, il s'agit de déterminer le contenu et les difficultés en se basant sur des principes. Enfin, reste à rédiger les spécifications du test, document essentiel pour les étapes finales d'élaboration et d'examen du test.

2.2 Déterminer l'objectif du test et les besoins des candidats dans le monde réel

Avant d'élaborer un test linguistique, il est nécessaire d'en déterminer l'objectif précis. En effet, il ne suffit pas de déclarer qu'un test est pratiqué à des fins de migration et de citoyenneté, car même dans ce domaine, les raisons d'évaluer les migrants sont multiples — inciter les apprenants (les aider à utiliser et à améliorer leur compétence dans la langue cible), s'assurer que leurs compétences sont suffisantes pour participer à des situations sociales bien définies (études ou travail, par exemple), prendre des décisions qui concernent leurs droits juridiques (par exemple, droit de rester dans un pays ou d'en acquérir la citoyenneté).

Une fois l'objectif clairement défini, il est alors possible d'identifier les demandes auxquelles les candidats se trouveront confrontés dans le monde réel (par exemple, nécessité de participer à des processus sociétaux et d'exercer les droits et les responsabilités d'un citoyen) ; le test doit rendre compte de ces besoins. Un objectif clair et explicite ne servira pas seulement à étayer l'élaboration des tests et à clarifier les attentes des candidats — contribuant ainsi à l'équité des tests —, mais permettra aussi aux autres membres de la société de correctement interpréter et exploiter les résultats. Ce processus est appelé « *analyse des besoins* ».

Lors d'une analyse des besoins, il est également nécessaire de tenir compte du fait qu'il existe différents sous-groupes de migrants avec leurs propres besoins spécifiques. Ainsi, les besoins diffèrent souvent entre ceux qui veulent entrer sur le marché du travail le plus rapidement possible et ceux qui envisagent d'élever de jeunes enfants au foyer. Dans une analyse des besoins, les concepteurs de tests linguistiques doivent se faire une habitude de définir les contextes et les situations intéressant le public ciblé. Lors de la planification de ce type d'analyse, les décideurs veilleront à prévoir les ressources suffisantes et des représentants de différents secteurs de la société interviendront dans la définition des besoins.

2.3 Déterminer les demandes linguistiques

Une fois les demandes du monde réel identifiées, il convient de les traduire en exigences linguistiques en indiquant non seulement connaissances et compétences, mais aussi le niveau de capacité requis pour chacune de ces deux catégories. Par exemple, si un test linguistique vise à vérifier que les candidats possèdent la compétence langagière nécessaire pour suivre une formation professionnelle, il est probable qu'il évaluera la capacité à suivre des cours et des ateliers, à communiquer avec des enseignants et d'autres apprenants, à lire la documentation voulue, à rédiger des devoirs, etc. Cette analyse pourrait alors aider à déterminer le niveau adéquat de compétence langagière requis – ou les niveaux requis dans chacune des compétences testées (compréhension écrite et production écrite, par exemple). En revanche, si un test linguistique est conçu pour la certification de traducteurs ou d'interprètes, le même type d'analyse des besoins appliqué à la profession montrerait que le niveau de compétence requis est très supérieur et que, pour le traducteur, les exigences en lecture et écriture sont plus fortes, alors que pour les interprètes, c'est la compétence orale qui prime.

À l'opposé, établir des exigences linguistiques à partir de tâches à accomplir dans le monde réel est beaucoup moins simple dans le cas de migrants et de candidats à la citoyenneté. En effet, entre compétence linguistique dans la langue officielle et capacité à s'intégrer dans la société et/ou à exercer les droits et responsabilités de la citoyenneté, la relation est plus vague et beaucoup plus difficile à cerner. De fait, si la compétence linguistique était le seul et unique facteur en jeu, tous les habitants et locuteurs natifs d'un pays seraient des citoyens pleinement intégrés ; ce qui n'est pas le cas. L'on peut donc en conclure que d'autres facteurs entrent aussi en jeu. En tout état de cause, la tâche du concepteur de tests de langue consiste à déterminer les besoins et les attentes linguistiques applicables. Après une analyse approfondie des besoins (ainsi que décrit plus haut), il importe également de s'assurer que les tests ne sont pas orientés par des hypothèses erronées sur le bagage éducatif et culturel des candidats. En effet, dans la majorité des cas, les tests linguistiques conçus en vue d'études et du travail sont destinés à des candidats présentant une certaine homogénéité en termes de bagage éducatif et de compétences cognitives, alors que les tests pour l'intégration et la citoyenneté (pour acquérir des droits civiques) doivent s'adresser à un vaste éventail de candidats possibles et, par conséquent, être accessibles tout aussi bien à des personnes maîtrisant peu ou pas les mécanismes de la lecture et de l'écriture qu'à celles bénéficiant d'un niveau élevé d'enseignement et de culture générale.

2.4 Déterminer le niveau de difficulté adéquat

Une fois les demandes linguistiques déterminées, les concepteurs doivent essayer de les faire correspondre aux « capacités de faire » proposées par le *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer* (CECR)². Ce Cadre...

[...] offre une base commune pour l'élaboration de programmes de langues vivantes, de référentiels, d'examens, de manuels, etc. en Europe. Il décrit aussi complètement que possible ce que les apprenants d'une langue doivent apprendre afin de l'utiliser dans le but de communiquer ; il énumère également les connaissances et les habiletés qu'ils doivent acquérir afin d'avoir un comportement langagier efficace.

CECR, p. 1

Le CECR contient un certain nombre d'échelles illustratives – une par compétence : parler, écrire, lire, écouter et interagir – qui définissent des niveaux de compétence, indiquent ce qu'un apprenant *est capable de faire* à chaque niveau et permettent de mesurer les progrès sur une échelle à six niveaux – de A1 (niveau introductif ou découverte) à C2 (maîtrise).

² Conseil de l'Europe / Editions Didier, 2001. Disponible en ligne : www.coe.int/lang/fr

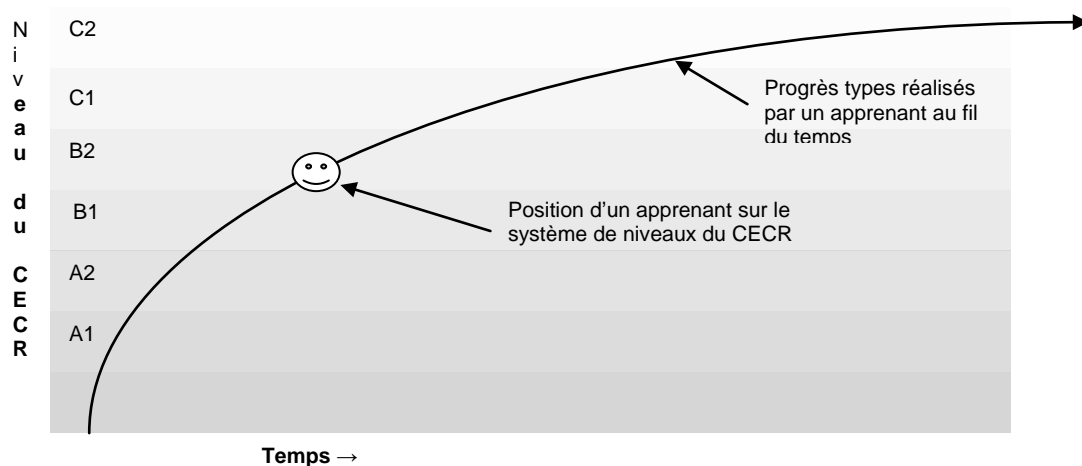


Schéma 1 : Courbe d'apprentissage et niveaux du CECR

Le schéma 1 montre les progrès généralement réalisés par un apprenant au fil du temps par rapport aux niveaux du CECR (représentés par des bandes horizontales). À noter que, sur ce schéma, la largeur de chacun des niveaux ne doit pas faire croire qu'un même laps de temps est nécessaire pour atteindre chaque niveau supplémentaire. Vous remarquerez que la courbe représentant les progrès types d'un apprenant monte rapidement au début et s'aplatit vers la fin ; cela s'explique par le fait que la gamme des compétences linguistiques ajoutées à chaque niveau et, par conséquent, le temps nécessaire pour passer d'un niveau au suivant, augmente à mesure des progrès réalisés. Aussi l'analyse des besoins ne doit-elle pas se baser sur le nombre d'heures d'enseignement, mais sur la correspondance rigoureuse entre besoins de l'apprenant et « capacités de faire ».

Les concepteurs de tests doivent comprendre que les compétences de la majorité des candidats ne se répartissent pas de manière égale entre les quatre aptitudes : lire, écouter, écrire et parler (comme le suggère la figure 1). Au contraire, les compétences de production et de compréhension orales sont généralement supérieures à celles de production et de compréhension écrites. Ainsi, l'apprenant présenté sur le schéma pourrait avoir un profil « *en dents de scie* » par rapport aux quatre aptitudes linguistiques (comme l'illustre le schéma 2). En conséquence, les concepteurs de tests auront peut-être intérêt à définir explicitement les compétences par le biais de l'évaluation.

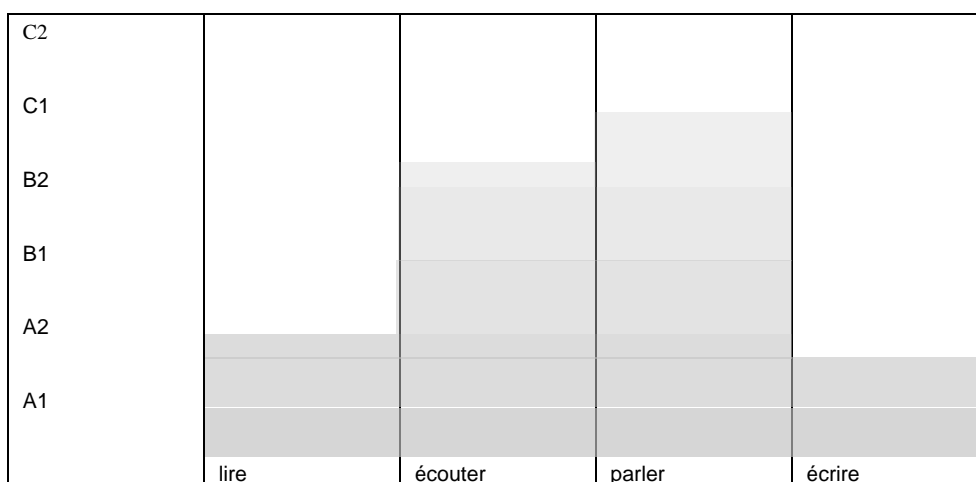


Schéma 2 : Profil « en dents de scie » des capacités linguistiques d'un apprenant

Outre qu'elle donne une image plus fidèle des capacités du candidat, une approche modulaire de l'évaluation présente plusieurs avantages. S'il y a possibilité de passer un test

séparément pour chaque aptitude, les candidats peuvent se présenter à celles qu'ils maîtrisent le mieux, ce qui peut être plus encourageant que d'avoir à se confronter aux quatre aptitudes à la fois. Cette possibilité conviendrait particulièrement bien au groupe des migrants analphabètes ou aux participants possédant très peu d'expérience de l'écrit. Sans compter que les cours de langue peuvent alors cibler les aptitudes nécessitant davantage d'attention. Pour communiquer au public les résultats d'un test de compétences partielles, les institutions d'évaluation doivent indiquer clairement les compétences testées, afin d'éviter toute confusion entre tests de compétences partielles et tests des quatre aptitudes. En conséquence, au lieu d'indiquer un seul et unique niveau de capacité global pour le candidat, les rapports ou certificats de tests doivent préciser le niveau atteint dans chaque aptitude.

Documents de référence :

Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer – Conseil de l'Europe : www.coe.int/lang/fr [http://www.coe.int/t/dg4/linguistic/cadre_FR.asp]

2.5 Produire des spécifications pour les tests

Une fois que la candidature cible, la finalité de l'évaluation et l'orientation de l'évaluation sont établies, elles doivent faire l'objet de spécifications détaillées. Les spécifications décrivent aussi les types d'items ou de tâches à utiliser, le format du test et d'autres questions pratiques — qui, elles aussi, seront décidées sur la base des préceptes élémentaires (candidature cible, par exemple). Ces spécifications servent alors de document de référence, notamment pour toutes les décisions prises aux étapes ultérieures du test.

3. Application pratique des spécifications de test

3.1 Généralités

Une fois les spécifications de test finalisées, plusieurs étapes restent à effectuer pour assurer le bon fonctionnement du test : définir des critères d'évaluation et le format du test, rédiger des items (« questions »), puis les combiner conformément aux spécifications. Une fois produit, le test doit être appliqué aux candidats de manière cohérente et équitable. Enfin, après passation du test, les données obtenues doivent être analysées afin de vérifier le bon déroulement du test – et, au besoin, ajustées. Tout au long du processus, des contrôles d'assurance qualité sont nécessaires et décrits dans le texte. Pour mettre en œuvre les codes de bonnes pratiques et d'éthique, les fournisseurs de tests pourront se référer utilement aux divers documents ici mentionnés (voir la section *Documents de référence* du chapitre d'introduction). Ces documents, qui proposent une forme plus détaillée et concrète pour les concepts examinés dans les documents cités en introduction, sont répertoriés dans la section suivante.

Documents de référence :

AERA/APA/NCME, *Standards for Educational and Psychological Testing* – <http://www.apa.org/science/standards.html>

ALTE COP QMS Checklists (listes de contrôle) – http://www.alte.org/quality_assurance/code/checklist.php

ALTE *Minimum standards for establishing quality profiles in ALTE examinations* – http://www.alte.org/quality_assurance/index.php

ALTE *Principles of Good Practice* – http://www.alte.org/quality_assurance/code/good_practice.pdf

EALTA *Guidelines for Good Practice* – <http://www.ealta.eu.org/guidelines.htm>

ILTA *Draft Code of Practice* – <http://www.iltaonline.com/ILTA-COP-ver3-21Jun2006.pdf>

3.2 Critères d'évaluation et format des tests

Pour être utile au concepteur de tests, l'orientation de l'évaluation telle que décrite dans les spécifications des tests (voir la section 2.5) devra être divisée en points spécifiques. C'est seulement alors, lorsque sera clairement tracé un tableau de l'évaluation des performances des candidats, que pourra commencer l'élaboration d'une combinaison adéquate des tâches et des types de tâches à réaliser pour le test. Avant tout, le concepteur doit avoir à l'esprit qu'il faut donner aux candidats des possibilités de faire la preuve qu'ils répondent aux critères d'évaluation.

3.3 Rédaction des items

Pour produire des items de test adéquats, les rédacteurs doivent disposer de directives claires. Le plus souvent, ces directives fourniront une vue d'ensemble de la candidature ciblée et de l'objectif du test ainsi que des conseils d'ordre général – par exemple, pertinence ou non de certains sujets, longueur des éléments de départ (nombre de mots dans un texte à lire, par exemple) et de la production attendue (nombre de mots à écrire par le candidat), degré d'authenticité des textes, etc. – avant de porter individuellement sur chaque type d'item. Une fois les items rédigés, d'autres experts vérifieront alors si la lettre et l'esprit des directives sont respectés.

Documents de référence :

Item Writer Guidelines - http://www.alte.org/projects/item_writer.php

3.4 Pré-évaluation des items

Le recours au jugement d'experts (voir la section 3.3 ci-dessus) doit permettre d'obtenir des items *apparemment* adaptés aux objectifs d'évaluation. Toutefois, pour être sûr que les items produisent les résultats escomptés (évaluation de la langue cible, différenciation effective entre candidats plus forts et plus faibles, impartialité du test, etc.), il est nécessaire de les pré-évaluer en conditions d'examen sur des candidats présentant un profil démographique aussi proche que possible de celui des candidats réels. L'évaluation préalable de tests objectifs (questions à choix multiples, par exemple) débouche sur une analyse statistique détaillée, alors que celle de tests subjectifs (production orale, par exemple) conduit plus certainement à l'analyse qualitative des productions des candidats afin de déterminer dans quelle mesure elles correspondent aux attentes du fournisseur.

À partir de ces analyses, les items et les tâches peuvent soit être acceptés, modifiés et à nouveau pré-évalués, soit être rejetés. En outre, si une *banque d'items* est créée, en dotant chacun des items de caractéristiques légèrement différentes, il est possible de construire des tests spécifiquement adaptés (par exemple, couvrant une gamme précise de niveaux de capacité).

À diverses étapes du processus de rédaction et de pré-évaluation des items, il est important que des experts révisent les tâches et les items. Cette révision doit être répétée car les items peuvent subir des modifications ou de nouvelles informations peuvent intervenir (résultats de la pré-évaluation, par exemple). Des outils facilitant ce processus de révision sont mentionnés à la section *Documents de référence*. Les items et les tâches – et, parfois, les réponses apportées aux tâches – peuvent être analysés selon les catégories proposées dans ces outils. D'autre part, il est possible d'établir en quoi les tâches diffèrent les unes des autres et dans quelle mesure elles respectent les spécifications de test. Les grilles mentionnées ci-après, bien qu'initialement conçues pour aider les fournisseurs de tests à aligner leurs examens sur le CECR, peuvent être utilisées dans cette perspective plus large.

Documents de référence :

Content Analysis Checklists – <http://www.alte.org/projects/content.php>

CECR, Conseil de l'Europe : www.coe.int/portfolio/fr.

Grilles pour l'analyse de tâches de test (compréhension orale, lecture, parole et écriture)
[http://www.coe.int/T/DG4/Portfolio/?L=F&M=/documents_intro/Manuel.html]

Voir également: *Illustrations des niveaux européens de compétences en langues*
[http://www.coe.int/T/DG4/Portfolio/?L=F&M=/main_pages/illustrationsf.html]

3.5 Passation des tests

Les fournisseurs de tests doivent veiller à ce que les tests soient passés dans des conditions équitables pour tous les candidats. À cette fin, il est recommandé de mettre en place des procédures visant à minimiser les différences au cours du processus d'examen. Ces procédures doivent assurer les points suivants :

- Le centre d'examen est dûment agréé pour faire passer les tests ;
- Le personnel du centre d'examen est compétent, informé et dûment supervisé ;
- Le centre d'examen assure un niveau élevé de sécurité et de confidentialité durant tout le processus, depuis l'inscription jusqu'à la communication des résultats et des rapports ;
- Les conditions physiques dans la salle d'examen sont correctes (niveau de bruit, température, distance entre les candidats, etc.) ;
- Des dispositions sont prises pour les candidats présentant des besoins spéciaux.

3.6 Participation des candidats présentant des besoins spéciaux

Le système d'évaluation ne doit pas entretenir de discrimination à l'encontre des candidats présentant des besoins spéciaux — lesquels peuvent être le fait d'une incapacité ou d'une infirmité physique, mentale ou émotionnelle temporaire ou de longue durée, de troubles de l'apprentissage, d'une maladie temporaire ou de longue durée, d'un analphabétisme dans la L1 ou dans la langue cible, de règles religieuses, d'une détention pénale ou de toute autre circonstance empêchant un candidat de passer le test dans les mêmes conditions que les autres candidats.

Des dispositions doivent être prévues pour :

- décider si un candidat présentant des besoins spéciaux sera dispensé de passer le test (en partie ou en entier) ;
- prendre des mesures visant à garantir un traitement équitable aux candidats présentant des besoins spéciaux ;
- définir à quelle institution il incombe de décider si le test doit être passé par tel ou tel candidat ;
- déterminer les conditions spéciales à appliquer dans un cas donné (par exemple, épreuves en braille, épreuves en gros caractères, mise à disposition d'une machine à écrire le braille ou d'un ordinateur doté de fonctions spéciales, mise à disposition d'un lecteur/scribe/assistant, temps supplémentaire alloué pour certaines parties du test, pauses supplémentaires, interprète en langue des signes, dates et lieux d'examen spécifiques) ;
- assurer la possibilité d'un recours contre cette décision, informer les candidats sur la manière de déposer un recours et indiquer comment la décision finale doit être prise.

Des informations sur ces règlements et dispenses doivent être publiées et accessibles aux candidats.

Documents de référence :

« *Besoins éducatifs particuliers en Europe : l'enseignement des langues, idées et innovations* »
Commission européenne, DG EAC 23 03 LOT 3, janvier 2005.

3.7 Notation

Les items de test « objectifs » peuvent être notés avec précision par des machines ou par des correcteurs qualifiés. En revanche, la notation des items « subjectifs » nécessite généralement des évaluateurs qualifiés, procédure plus sujette à des erreurs d'écriture ou de jugement. Une erreur de jugement se traduit par des incohérences, de la part de l'examineur, dans l'interprétation des réponses et/ou dans l'application des critères d'évaluation du test. Un examinateur peut se montrer incohérent soit par rapport à d'autres examinateurs, soit dans sa manière d'appliquer les critères dans la durée. Ces incohérences sont largement réductibles par l'application de procédures de formation et de contrôle rigoureuses, ainsi que par une notation croisée de tout ou partie des performances des candidats. Le travail des correcteurs peut aussi être « pondéré » pour compenser une indulgence/sévérité régulière, quoique des différences importantes soient sans doute le signe d'un manque de formation.

3.8 Contrôle

Outre le contrôle du comportement des examinateurs (voir la section 3.7 ci-dessus), il est important que le concepteur de test de langue recueille et analyse des informations sur les réponses du candidat et sur le candidat lui-même (âge, sexe et nationalité, par exemple). Ce contrôle est nécessaire pour assurer les points suivants :

- Chaque test mesure les capacités pour lesquelles il est prévu ;
- Les capacités sont mesurées de manière cohérente par toutes les versions du même test, antérieures et ultérieures ;
- Chaque test fonctionne de manière équitable pour tous les candidats, indifféremment de leurs antécédents.

Les résultats de ce type de contrôle peuvent alors servir à assurer que les résultats du test reflètent exactement la capacité du candidat. En outre, les conclusions de l'analyse peuvent alimenter les processus d'élaboration, de passation et de notation du test et, par là-même, constamment les améliorer.

3.8.1 Contrôle des réponses des candidats

Lors de tests réels, les réponses apportées par les candidats aux items et aux tâches ont une double utilité : d'une part, fournir des notes brutes et, d'autre part, informer les fournisseurs de test sur l'efficacité des items et des tâches pour mesurer la capacité des candidats (voir la section 3.4 ci-dessus). Les items ne doivent pas être jugés uniquement sur leur niveau de difficulté, mais aussi sur leur capacité à opérer une distinction entre candidats « forts » et candidats « faibles » (le but premier du test étant, précisément, de différencier ces deux groupes). Ces items et autres statistiques sont à conserver et à comparer pour les versions antérieures et ultérieures du test. Ainsi les fournisseurs de test pourront-ils garantir la cohérence des résultats d'une version (ou session) du test et à une autre. Si le test a fait l'objet d'une pré-évaluation, il est probable que les items et les tâches fonctionneront comme prévu lors du test réel. Toutefois, il est important de confirmer ce point par une analyse et, en cas de dysfonctionnement, de procéder à un examen approfondi suivi d'une action corrective. Il est également important, bien entendu, que telle ou telle version d'un test

indique la capacité des candidats de manière cohérente avec des versions antérieures et ultérieures. C'est pourquoi les données obtenues par des réponses réelles servent souvent à statuer sur les limites entre niveaux ou sur les notes d'admission. Si un candidat doit passer deux versions différentes du même test, il est peu probable qu'il obtienne exactement la même note brute. Pourtant, le fournisseur de test peut raisonnablement assurer que le candidat obtiendra le même classement — ou se situera du même côté de la note d'admission — dans les deux tests. Pour ce faire, il faut mettre les tests *en équation*, ou évaluer quelle note de la version B est égale à la note requise pour réussir dans la version A. S'il y a eu pré-évaluation, cette estimation en sera sans doute d'autant plus facile et précise.

3.8.2 Contrôle de l'impartialité

Des techniques semblables à celles décrites à la section 3.8.1 (voir ci-dessus) doivent être appliquées aux groupes de candidats présentant un même niveau de capacité. Si, par exemple, comparativement à d'autres candidats de même capacité, des candidats d'une même nationalité réussissent mieux ou moins bien pour un item (ou groupes d'items), cela peut signifier que cet item est favorable ou défavorable aux personnes de cette nationalité ; ce qui serait injuste. Reste que la cause peut aussi être linguistique — par exemple, différences/similitudes entre langue maternelle et langue cible —, ce qui ne peut être considéré comme injuste. Quoiqu'il en soit, dès lors qu'un examen quantitatif a fait apparaître un possible déséquilibre, un examen qualitatif s'impose. Si un item est jugé favorable/défavorable, cela peut nécessiter son retrait et/ou une modification des procédures employées pour le produire.

4. Conclusion

L'utilisation de tests à des fins de migration et de citoyenneté est beaucoup plus complexe qu'il n'y paraît de prime abord. Dans ce chapitre, nous avons tenté de dégager les éléments à prendre en compte et, par voie de conséquence, ceux dont la responsabilité incombe aux décideurs. Deux questions sont à examiner en premier lieu : quel est le type d'évaluation nécessaire pour l'objectif visé et que peut mesurer cette évaluation ? Si un test est envisagé, il doit absolument satisfaire aux conditions soulignées dans ce chapitre. Par ailleurs, pour en garantir le fonctionnement et la qualité, le test doit être constamment contrôlé. D'autre part, il ne faut pas oublier que les résultats d'un test peuvent avoir de graves conséquences tant pour les candidats individuels que pour des groupes plus importants de personnes ou encore pour la société en général – notamment sur le plan des droits civiques et humains des candidats. Pour bien exploiter un test de langue à des fins de migration et de citoyenneté, dès que son utilisation est décidée, ceux qui définissent la politique doivent collaborer avec les fournisseurs de test sur plusieurs aspects ; notamment sur la définition de l'objectif précis du test et sur l'allocation des ressources nécessaires à toutes les étapes du processus de développement et d'utilisation. À tout moment, l'équité du test doit être considérée comme de première importance.