

Relier les examens de langues au Cadre européen
commun de référence pour les langues:
apprendre, enseigner, évaluer (CECR)

Les points essentiels du Manuel

édité par l'équipe RelEx

José Noijons, Jana Béréšová, Gilles Breton et Gábor Szabó

Centre européen pour les langues vivantes

Editions du Conseil de l'Europe

Edition anglaise:

Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR):

Highlights from the Manual

ISBN 978-92-871-7169-6

Les vues exprimées dans cet ouvrage sont de la responsabilité des auteurs et ne reflètent pas nécessairement la ligne officielle du Conseil de l'Europe.

Tous droits réservés. Aucun extrait de cette publication ne peut être traduit, reproduit ou transmis, sous quelque forme et par quelque moyen que ce soit – électronique (CD-Rom, internet, etc.), mécanique, photocopie, enregistrement ou de toute autre manière – sans l'autorisation préalable écrite de la Division de l'information publique, Direction de la communication (F-67075 Strasbourg Cedex ou publishing@coe.int).

Couverture: Georg Gross

Mise en page: Stenner Medienproduktion

Traduction: Gilles Breton

<http://book.coe.int>

Editions du Conseil de l'Europe

F-67075 Strasbourg Cedex

Centre européen pour les langues vivantes / Conseil de l'Europe

Nikolaiplatz 4

A-8020 Graz

www.ecml.at

ISBN: 978-92-871-7168-9

© Conseil de l'Europe, 2011

Imprimé en Autriche

Sommaire

Avant-propos	7
1. Le CECR et le Manuel	9
1.1. Les objectifs du Manuel	9
1.2. Le contexte du Manuel	11
2. Le processus de mise en relation	15
2.1. Approche adoptée	15
2.2. Questions liées à la qualité	17
2.3. Etapes du processus	17
2.4. Utilisation du CECR	18
2.5. Utilisation du Manuel	19
3. Familiarisation	21
3.1. Introduction	21
3.2. Activités préalables au séminaire	22
3.3. Activités introductives pendant le séminaire	23
3.4. Analyse qualitative des échelles du CECR	25
3.5. Préparation à l'évaluation	26
4. Spécification	39
4.1. Introduction	39
4.2. Description générale de l'examen	40
4.3. Outils disponibles pour la spécification	42
4.4. Procédures	43
4.5. Déclaration du niveau: représentation graphique de la relation de l'examen avec le CECR	48
5. Formation à la standardisation et au calibrage	51
5.1. Introduction	51
5.2. La formation nécessaire	52

5.3.	Planification préalable	53
5.4.	Animation des stages	55
5.5.	Formation avec des performances orales et écrites	56
5.6.	Formation à l'aide des tâches et des items de capacités de réception écrite et orale et de compétences linguistiques	59
5.7.	De la formation au calibrage	61
6.	Procédures de définition des scores de césure	65
6.1.	Introduction	65
6.2.	Aspects généraux	66
6.3.	La méthode du corpus de productions: centrée sur le candidat	69
6.4.	La méthode de Tucker-Angoff: centrée sur le test	70
6.5.	La méthode du panier: centrée sur le test	71
6.6.	La méthode du marque-page: centrée sur le test	72
6.7.	Définition des scores de césure sur plusieurs capacités langagières	73
6.8.	Définition des scores de césure et ajustement de tests	75
6.9.	Définition des scores de césure sur plusieurs langues	77
6.10	Conclusion	77
7.	Validation	81
7.1.	Introduction	81
7.2.	Pré-requis: la qualité de l'examen	81
7.3.	Validité procédurale de la formation à la standardisation et à la définition des scores de césure	85
7.4.	Validité interne de la définition des scores de césure	86
7.5.	Validation externe	87
7.6.	Conclusion	88
	Bibliographie	91
	Glossaire	97

RelEx

Relating Language Examinations

Relier les examens de langues

Cette publication est issue du projet du Centre européen pour les langues vivantes intitulé «Evaluation en classe reliée au Cadre européen commun de référence pour les langues (RelEx)».

Pour obtenir de plus amples informations et consulter les matériels liés à cette publication, veuillez consulter les sites internet <http://relex.ecml.at> et http://www.coe.int/t/dg4/linguistic/Manuel1_FR.asp.

Avant-propos

Waldemar Martyniuk

Directeur exécutif, Centre européen pour les langues vivantes

Cette publication est un résultat du projet RelEx du CELV qui vise à relier les examens de langues au Cadre européen commun de référence pour les langues (CECR) élaboré par le Conseil de l'Europe. En réponse à un besoin croissant exprimé par les utilisateurs de cet outil, la Division des politiques linguistiques du Conseil de l'Europe a élaboré un Manuel et un groupe d'instruments d'accompagnement pouvant être utilisés pour établir un lien de manière fiable et responsable entre les examens de langues locaux et les Niveaux communs de référence concernant les compétences linguistiques. Une des réalisations prévues du projet RelEx consistait en la production d'une introduction facile d'utilisation au *Manuel pour relier les examens de langues au Cadre européen commun de référence pour les langues*. Cette présentation des principaux éléments du Manuel est destinée aux concepteurs de politiques, aux experts dans le domaine de l'évaluation dans les centres d'examen, aux concepteurs de curriculums et aux formateurs d'enseignants et autres pédagogues ayant une connaissance moins étendue des aspects techniques du processus de mise en relation. Le projet RelEx veut porter l'attention sur le fait que relier les examens de langues vivantes au CECR est une activité qui en vaut la peine. Cependant, il souligne également le fait que cette mise en relation implique de soigneusement suivre différentes étapes.

L'équipe du projet est d'avis qu'une introduction au Manuel telle que celle-ci, qui présente les aspects les plus importants du processus de mise en relation, peut se révéler utile en permettant aux acteurs clés de mieux se familiariser avec le Manuel. Les formateurs d'enseignants ayant participé au projet ont tous bénéficié avec succès de cette présentation. L'attention des enseignants, qu'ils soient en formation initiale ou déjà praticiens, est attirée sur le fait qu'ils doivent vérifier si un examen prétendant être relié au CECR l'est vraiment. Lorsqu'ils envisagent d'utiliser des instruments d'évaluation existant déjà, les enseignants peuvent trouver des informations et des preuves étayant cette vérification. Ils peuvent également appliquer certaines des procédures proposées dans le Manuel (ou dans la présentation de ses principaux aspects) lorsqu'ils développent leurs propres examens en classe.

Cette présentation des principaux aspects du Manuel a été utilisée lors d'ateliers organisés par le CELV pour des participants issus de tous ses Etats membres. Les participants ont émis des commentaires positifs concernant sa concision et son accessibilité pour les non-experts. Il faut noter que cette présentation des principaux aspects du Manuel est entièrement et uniquement tirée de ce dernier et qu'aucun texte n'a été ajouté ou modifié, excepté dans les quelques cas où un terme technique devait être expliqué.

Le CELV est convaincu que cette publication constituera un nouvel outil utile pour relier les instruments d'évaluation quels qu'ils soient au Cadre européen commun de référence pour les langues. Il existe également des synergies entre cette publication et d'autres instruments développés par le Centre dans ce domaine. La publication intitulée *Parcours d'évaluation, d'apprentissage et d'enseignement à travers le CECR* résultant du projet du CELV «Encourager la culture de l'évaluation chez les professionnels (ECEP)» représente une approche complémentaire. Les deux publications peuvent être téléchargées sur le site internet du CELV: <http://www.ecml.at>.

Chapitre 1: Le CECR et le Manuel

1.1. Les objectifs du Manuel

Ce Manuel a pour principal objectif d'aider les concepteurs d'examens à élaborer des procédures transparentes et concrètes pour situer leurs examens par rapport au CECR, à les appliquer et à en rendre compte dans un processus cumulatif de perfectionnement continu. Le Manuel n'est pas le seul guide permettant de relier un test au CECR et aucune institution n'est obligée d'entreprendre ce travail d'harmonisation. Cependant, les institutions qui affirment que leurs examens sont reliés aux niveaux du CECR trouveront les procédures proposées très utiles au moment de démontrer la validité de leur affirmation.

L'approche développée dans le Manuel propose des conseils aux utilisateurs pour:

- décrire ce que recouvre l'examen, ainsi que les procédures de passation et d'analyse;
- mettre en relation les résultats de l'examen et les niveaux communs de référence du CECR;
- apporter des preuves rendant compte des procédures suivies.

Toutefois, suivant en cela les meilleures traditions de l'action du Conseil de l'Europe pour le développement de l'enseignement des langues, le Manuel vise plus largement à fortement encourager et faciliter la coopération entre les organismes concernés et les spécialistes des pays membres. Le Manuel a pour objectif de:

- contribuer à l'élaboration d'une compétence dans le domaine de la relation des examens de langues avec le CECR;
- encourager une plus grande transparence de la part des organismes qui produisent des examens;
- encourager la constitution de réseaux d'organismes et d'experts, officiels ou non, tant sur le plan national qu'international.

La Division des politiques linguistiques recommande aux concepteurs d'examens utilisant les procédures proposées ou d'autres procédures visant les mêmes fins de faire un relevé d'expérience sous forme de rapport. Ces rapports devraient faire une description de la mise en œuvre des procédures, des points positifs et des difficultés et s'ils affirment que l'examen est relié aux niveaux du CECR, en apporter des preuves. On encourage fortement les utilisateurs à rédiger ces rapports afin:

- d'accroître la transparence du contenu (justification théorique, objectifs de l'examen, etc.);

- d'accroître la transparence du niveau attendu de l'examen;
- de donner aux candidats, aux utilisateurs et aux professionnels de l'enseignement et de l'évaluation l'occasion d'analyser la qualité de l'examen et de la relation affirmée avec le CECR;
- de procurer un argumentaire expliquant pourquoi certaines des procédures recommandées n'ont pas été suivies;
- de procurer à de futurs chercheurs un ensemble élargi de techniques pouvant venir en complément de celles indiquées dans ce Manuel.

Il faut souligner que, si ce Manuel recouvre un large éventail d'activités, son objectif est limité:

- c'est un guide tout particulièrement axé sur les procédures à mettre en œuvre pour justifier l'affirmation selon laquelle un examen ou un test donné est relié au CECR;
- ce n'est pas un guide général pour l'élaboration de tests ou d'examens de langue de qualité. Il existe plusieurs guides utiles pour ce faire et ce sont ceux-là qu'il faut consulter;
- il *ne prescrit pas* une approche particulière pour élaborer des tests ou des examens de langue. Si le CECR milite en faveur d'une approche actionnelle de l'apprentissage des langues, il admet, dans son effort d'exhaustivité, que des examens différents peuvent refléter des buts différents (concepts hypothétiques);
- il n'exige pas que les tests soient spécialement conçus pour évaluer des performances en relation avec le CECR, mais une utilisation évidente du CECR pendant le processus de formation, de types de tâches, de rédaction d'items et d'élaboration de grille d'évaluation renforce l'affirmation selon laquelle le contenu est relié au CECR;
- il ne fournit pas de label, ni de statut de validité ou d'accréditation selon laquelle tel ou tel examen est relié au CECR. De telles affirmations relèvent de la responsabilité des institutions. Des associations de professionnels travaillent sur les standards et les codes de bonnes pratiques (par exemple l'AERA, American Educational Research Association – AERA/APA/NCME: 1999); l'EALTA – www.ealta.org; ALTE – www.ALTE.org). Ces associations sont une source d'informations et de conseils pour l'évaluation des langues et les procédures de mise en relation.

Malgré tout, la version expérimentale du Manuel a été utilisée par les responsables d'examens de différentes façons:

- en appliquant les procédures à un test élaboré avant le CECR et par conséquent sans relation évidente avec le Cadre, afin de pouvoir donner des résultats en rapport avec les niveaux du CECR;
- pour confirmer la relation entre un test datant d'avant le CECR et le concept hypothétique du CECR, ainsi que les niveaux du CECR;
- pour confirmer la relation entre un test conçu après la publication du CECR mais précédant celle du manuel;
- en apportant à la révision des examens des informations qui permettent une relation plus étroite avec le concept hypothétique et les niveaux du CECR;
- en aidant les écoles à mettre en œuvre des procédures pour relier leurs examens au CECR.

Pour aider les utilisateurs à savoir s'il est pertinent d'utiliser les procédures dans leur propre contexte et ce qu'implique leur utilisation, des encadrés reprenant quelques uns des points essentiels et des enjeux sont proposés, sur le modèle du CECR, à la fin de chaque chapitre (les utilisateurs peuvent se demander si...).

En vue d'assurer la cohérence, le système de numérotation utilisé dans la publication pour les tableaux, les formulaires et les diagrammes se réfère à celui du Manuel.

1.2. Le contexte du Manuel

Le Cadre européen commun de référence pour les langues se fixe un objectif ambitieux, celui de fournir:

... une base commune à l'élaboration de programmes de langues vivantes, de référentiels, d'examens, de manuels, etc. partout 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 capacités langagières qu'ils doivent acquérir afin d'avoir un comportement langagier efficace. La description englobe aussi le contexte culturel d'utilisation de la langue. Le cadre définit aussi les niveaux de compétence qui permettent de mesurer les progrès de l'apprenant à chaque étape de l'apprentissage et à tout moment de la vie.

(Conseil de L'Europe, 2001a:1).

Le CECR traite cependant également de l'évaluation et des examens, et c'est à ce niveau que le Manuel peut servir de référence:

L'un des principaux objectifs du Cadre de référence est d'aider tous les partenaires de l'enseignement et de l'apprentissage des langues à décrire les niveaux de compétence exigés par les standards et les examens existants afin de faciliter les comparaisons entre

les différents systèmes de certification. C'est dans ce but qu'ont été élaborés le Schéma descriptif et les Niveaux communs de référence. Ceux-ci fournissent une grille de lecture conceptuelle que les utilisateurs peuvent utiliser pour décrire leur système.

(Conseil de l'Europe, 2001a:21)

L'objectif du CECR est de faciliter la réflexion, la communication et le travail en réseau dans le domaine de l'enseignement et de l'apprentissage des langues. Au niveau local, l'objectif de toute stratégie devrait être de répondre aux besoins propres à un contexte. La clé pour concilier les deux objectifs en un système cohérent est la souplesse. Le CECR est un outil de référence semblable à un accordéon, fournissant des catégories, des niveaux et des descripteurs que des professionnels de l'éducation peuvent regrouper ou subdiviser, détailler ou résumer – tout en gardant la structure hiérarchique commune. On encourage les utilisateurs à mettre en place des ensembles d'activités langagières, de compétences et de performance convenant à leur contexte local, mais qui soient aussi en rapport avec le schéma général, afin de permettre une communication plus aisée avec des collègues d'autres institutions et d'autres parties prenantes telles que les apprenants, les parents et les employeurs.

Il n'y a pas de contradiction entre, d'une part, un cadre commun de référence nécessaire à l'organisation de l'enseignement et facilitant les comparaisons et, d'autre part, des stratégies et des décisions locales nécessaires pour faciliter un apprentissage efficace et élaborer des examens convenant à tout type de contexte.

Par contre, la reconnaissance mutuelle de qualifications langagières octroyées par toutes les parties concernées est une question beaucoup plus compliquée. En Europe, les professionnels de l'évaluation en langue ont des traditions très différentes. D'un côté, on trouve les producteurs d'examens qui opèrent selon le mode classique d'examens annuels préparés par une commission de spécialistes et notés en fonction de la connaissance intuitive du standard exigé. Il existe de nombreux cas où l'examen ou le test débouchant sur une qualification reconnue est préparé par l'enseignant ou le personnel de l'école plutôt que par une commission externe, parfois sous le contrôle d'un expert extérieur. Il y a ensuite de nombreux examens qui se concentrent sur la mise en œuvre de spécifications de tâches, avec des critères écrits, un barème et une formation des examinateurs permettant d'assurer une cohérence; ils incluent ou excluent selon le cas une forme de pré-test ou de validation empirique. Enfin, de l'autre côté, on trouve des systèmes extrêmement centralisés qui utilisent essentiellement des questions à réponse fermée pour mesurer des capacités de réception. Les questions sont extraites de banques d'items. On y ajoute quelquefois des tâches de production (habituellement écrites) afin de mesurer la compétence et de délivrer les certifications. Les politiques nationales, les traditions et les cultures de l'évaluation autant que les politiques, les cultures et les intérêts légitimes des organismes spécialisés dans les tests et les examens de langue sont des facteurs qui peuvent être un frein à l'intérêt qu'il y a à une reconnaissance mutuelle des qualifications. Toutefois, il y va de l'intérêt de chacun que l'on applique des procédures convenables en matière d'évaluation.

Parallèlement à la question de la tradition se pose celle de la compétence et des ressources. Des établissements reconnus ont ou peuvent avoir les ressources à la fois humaines et matérielles qui leur permettent de mettre en œuvre et d'appliquer des procédures traduisant de bonnes pratiques, ainsi que des systèmes convenables de formation, d'assurance qualité et de contrôle. Dans d'autres cas, l'expérience de l'évaluation et les connaissances nécessaires sont moindres. Il peut n'y avoir qu'une familiarité limitée avec les techniques de travail en réseau et de formation des examinateurs à l'évaluation en fonction de standards, lesquels sont un préalable à toute évaluation cohérente de la performance. D'un autre côté, il peut n'y avoir que peu de familiarité avec des approches qualitatives et psychométriques, préalable nécessaire à la validation adéquate d'un examen. Mais surtout, il peut n'y avoir qu'une familiarité limitée avec les techniques de mise en relation des examens puisque, dans la plupart des cas, les groupes qui s'occupent d'évaluation ont l'habitude de travailler de manière isolée.

Il n'est nulle part envisagé une équivalence quelconque entre des examens différents qui auraient été reliés au CECR en suivant les procédures proposées dans le Manuel. Le contenu et la forme des examens varient en fonction des besoins liés au contexte et des traditions de la culture pédagogique qui ont déterminé leur conception. Deux examens peuvent très bien être au «niveau B2» tout en étant très différents. Des apprenants dans deux contextes différents peuvent obtenir des résultats différents (a) à un examen dont la forme et le contenu leur sont familiers et (b) à un examen du même niveau conçu dans un contexte différent. Ensuite, ce n'est pas parce qu'ils ont suivi les procédures pour relier les examens préconisées par ce Manuel que plusieurs examens peuvent, par exemple, se réclamer exactement du même niveau, par exemple B2. Le niveau B2, comme tout autre niveau, est à situer sur une «bande» de performance langagière qui est très large; le point de césure pour passer d'un niveau à un autre dans ces différents examens peut être déterminé à des endroits différents dans cette bande qui ne coïncident pas tous exactement à la même limite entre les niveaux B1 et B2.

Les programmes et les examens pour l'apprentissage des langues doivent être conçus et adaptés au contexte dans lequel ils seront utilisés. Les auteurs du CECR sont très clairs à ce propos: le CECR ne doit en aucune façon être considéré comme un projet d'harmonisation. Le CECR n'a aucune intention de dire aux professionnels du domaine des langues quels devraient être les objectifs. Il n'est pas dans notre intention de dire aux praticiens ce qu'il faut faire et comment le faire.

Il n'est pas non plus dans l'intention de ce Manuel de dire aux professionnels du domaine ce que devraient être les standards et la façon de prouver le lien établi avec eux. Le CECR et ce Manuel ont tous les deux comme objectifs d'encourager la réflexion, de faciliter la communication (entre les professionnels du domaine et entre les parties concernées par l'éducation) et de fournir des outils de référence concernant les processus et les techniques. Les Etats membres et les institutions concernés par l'enseignement et l'apprentissage des langues travaillent et coopèrent de façon autonome; c'est à eux et elles que reviennent le privilège et la responsabilité du choix de l'approche la plus appropriée à leur but et leur contexte.

Les utilisateurs peuvent se demander:

- si l'utilisation du CECR est pertinente dans leur évaluation et leur contexte;
- pourquoi et dans quels buts ils appliquent ce Manuel;
- quelles modifications sont à introduire dans leur contexte spécifique pour l'application du Manuel;
- quelles parties du Manuel les concernent le plus;
- comment ils pensent faire connaître leurs résultats afin de contribuer à l'amélioration de l'expertise dans le domaine de la mise en relation.

Chapitre 2: Le processus de mise en relation

2.1. Approche adoptée

Relier un examen ou un test au CECR est une entreprise complexe. L'existence d'une relation entre l'examen et le CECR n'est pas un fait directement observable, mais relève d'une affirmation pour laquelle les concepteurs d'examen devront apporter des preuves tant au plan théorique qu'empirique. La procédure par laquelle on obtient ces preuves est la «validation de l'affirmation».

Mettre en relation des examens ou des tests avec le CECR présuppose que l'on ait défini un ou plusieurs points de césure. Ces points de césure répartissent la distribution des performances des candidats sur deux ou plus de deux niveaux du CECR.

On peut garantir des normes convenables si on applique un processus approprié de définition des points de césure dès le début de la mise en relation. La définition de points de césure implique des prises de décision exigeant des données de haute qualité et un travail rigoureux. A partir du moment où ces décisions peuvent avoir des conséquences importantes, elles doivent être équitables, ouvertes, valides, efficaces et défendables. Ce sera le cas si des processus systématiques éprouvés et des critères explicites sont utilisés.

Lors de la définition de points de césure, il est fréquent de se référer à des contenus et des performances standards. Les contenus standards décrivent les contenus liés aux domaines à partir desquels l'examen peut être ou a été élaboré. Cette description renvoie très souvent aux niveaux de performance. De telles descriptions ont forcément un caractère général et sont habituellement formulées en termes qualitatifs.

Il faut toutefois souligner un point important. Le Cadre européen commun de référence pour les langues (CECR) fournit les contenus et les descripteurs du niveau de performance (DNP). Les DNP y sont donnés, contrairement à ce qui se passe dans les autres contextes de définition de points de césure, où les DNP doivent être définis en premier.

Cela signifie qu'on doit se référer au CECR à chaque étape du processus de mise en relation (cf. schéma 2.1). L'approche retenue dans ce Manuel rend indispensable une connaissance approfondie du CECR.

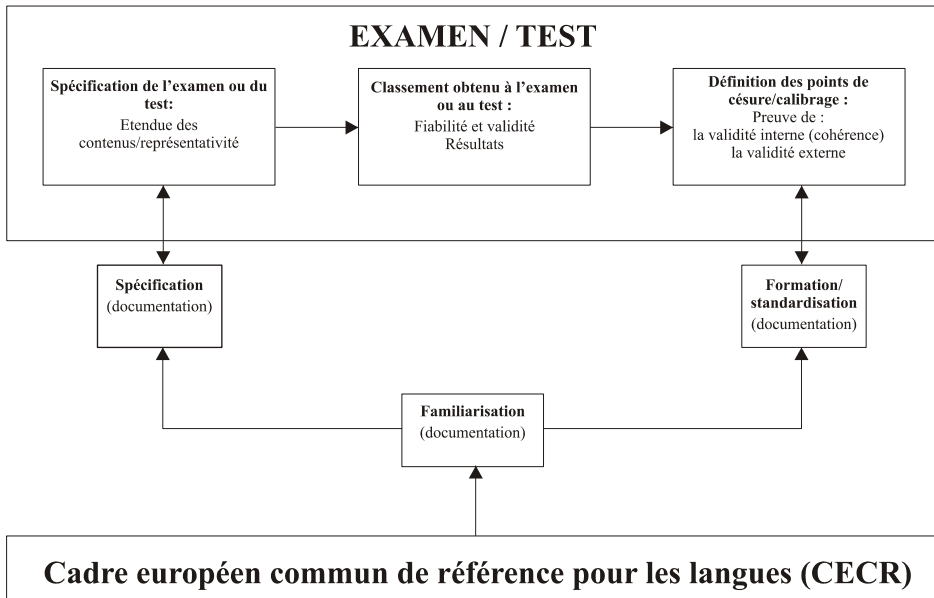


Schéma 2.1: Preuve de la validité de la mise en relation de l'examen ou du test avec le CECR

On peut envisager la mise en relation d'un examen ou d'un test avec le CECR comme un processus «d'élaboration d'une argumentation» basé sur un raisonnement théorique. Le concept de «validité» est au centre de ce processus. Le Manuel présente cinq séries de démarches étroitement liées. On conseille aux utilisateurs de suivre ces démarches une à une, ce qui leur permet de concevoir leur plan de mise en relation comme une suite d'activités indépendantes et réalisables:

- familiarisation;
- spécification;
- formation à la standardisation/calibrage;
- définition des points de césure;
- validation.

Il est important d'insister sur le fait que les cinq séries de procédures (ou «étapes») ne constituent pas uniquement des jalons isolés les uns des autres sur un processus linéaire. Il est primordial de vérifier, à l'issue de chaque étape, qu'on est sur la bonne voie: l'interprétation des niveaux correspond bien à l'interprétation courante, illustrée par des exemples représentatifs. Dans le cas de la révision ou du développement d'un examen, il

est conseillé d'appliquer les procédures recommandées à chaque étape du développement ou de la révision, de façon à ce que la mise en relation avec le CECR se fasse d'une façon organisée, cyclique, à mesure que l'équipe devient de plus en plus familiarisée avec le CECR – et que le projet ne soit pas remis en cours de route à une autre équipe, interne ou extérieure à l'institution, avant que le projet principal ne soit achevé.

Bien qu'elles ne doivent pas être considérées comme des jalons sur un parcours linéaire, les cinq étapes s'organisent selon un ordre logique. A chaque étape, on demande aux utilisateurs de commencer par les compétences de production (orale et écrite), car ces compétences peuvent être plus directement reliées aux riches descriptions du CECR, fournissant ainsi une base claire pour la formation, les jugements et les discussions.

2.2. Questions liées à la qualité

La mise en relation d'un examen ou d'un test avec le CECR ne peut être valide si l'examen ou le test en question ne peut démontrer une validité en lui-même. Un test qui ne convient pas à un certain contexte ne conviendra pas plus s'il est mis en relation avec le CECR; de même, un examen qui ne dispose pas de procédures permettant de s'assurer que les examinateurs et les correcteurs appliquent les mêmes normes de sévérité, ou que les versions d'un test administré lors de différentes sessions sont équivalentes, ne peut rendre crédible une affirmation de mise en relation avec le CECR, car il ne peut démontrer de cohérence interne dans l'opérationnalisation de ses normes.

2.3. Etapes du processus

Le processus de mise en relation d'un test avec le CECR consiste à mettre en œuvre les différentes étapes d'une série de procédures:

- **La familiarisation** (chapitre 3): il s'agit d'une sélection d'activités de formation visant à ce que ceux qui participent au processus de mise en relation parviennent à une bonne connaissance du CECR, de ses niveaux et de ses descripteurs.
- **La spécification** (chapitre 4): il s'agit d'un inventaire de l'étendue de ce que l'examen recouvre (contenu et types de tâches) par rapport aux catégories présentées dans le CECR au chapitre 4: «L'utilisation de la langue et l'apprenant/utilisateur» et au chapitre 5: «Les compétences de l'apprenant/utilisateur». Tout en faisant fonction de compte rendu, ces procédures servent également, dans une certaine mesure, à la prise de conscience qui pourra ultérieurement aider à l'amélioration de la qualité de

l'examen en question. Les fiches A2 et A8-A20 du chapitre 4 mettent l'accent sur l'analyse des contenus et la relation qu'ils entretiennent avec le CECR.

- **La formation à la standardisation et le calibrage** (chapitre 5): les démarches proposées facilitent la mise en œuvre d'une compréhension commune des «Niveaux communs de référence», à l'aide des exemples représentatifs des performances orales et écrites. Un calibrage réussi d'exemples locaux peut venir à l'appui d'une affirmation basée sur les résultats de la spécification. En effet, si les conclusions du processus de calibrage indiquent que les échantillons de performances du test ont été avec succès calibrés sur les niveaux pour lesquels ils avaient été conçus, cela confirme l'affirmation à laquelle on est arrivé dans la spécification.
- **La définition des points de césure** (chapitre 6): le point crucial dans le processus de mise en relation d'un examen avec le CECR est l'instauration d'une règle permettant de décider si on attribue l'un des niveaux du CECR à un candidat, à partir de la performance qu'il a réalisée lors de l'examen. On prend généralement une décision sur les points de césure, sur les performances à la limite du niveau. Les étapes précédentes, de familiarisation, spécification et standardisation peuvent être considérées comme des activités préparatoires à des décisions valides et logiques. Le chapitre 6 décrit les procédures qui mènent aux décisions finales permettant de définir les scores de césure.
- **La validation** (chapitre 7): bien que les étapes précédentes de familiarisation, spécification, standardisation et définition des points de césure puissent être réalisées dans un ordre chronologique, il serait imprudent d'attendre que tout soit terminé avant d'entreprendre les activités de validation, comme si elles constituaient l'ultime verdict sur la qualité du processus de mise en relation. La validation doit plutôt être considérée comme un processus continu de contrôle de la qualité, qui peut permettre de répondre à la question générale: «Avons-nous atteint notre but pour cette activité?».

2.4. Utilisation du CECR

Un cadre commun de référence permet à différents examens d'être reliés entre eux indirectement sans qu'ils prétendent être exactement équivalents. L'objectif d'un examen peut varier, mais ce qu'il recouvre peut être défini en relation directe avec les catégories et les niveaux du CECR. De même que deux étudiants de niveau B2 sont à ce niveau pour des raisons différentes, deux examens de niveau B2 auront des aspects qui ne seront pas totalement identiques.

2.5. Utilisation du Manuel

Les chapitres suivants concernent les différentes étapes du processus de mise en relation. Pour chaque étape, l'utilisateur peut choisir parmi l'ensemble de procédures proposées celles qui correspondent le mieux à leur contexte.

Le Manuel ne prétend pas être un modèle pour la conception d'un nouvel examen. En revanche, il a vocation à encourager une réflexion sur les bonnes pratiques.

Le Manuel propose un ensemble de procédures et de techniques de principe qui étayent le processus techniquement complexe et exigeant de mise en relation. Chaque étape suppose un jugement averti. C'est à l'organisme certificateur concerné qu'incombe la responsabilité de la mise en place d'un processus cohérent et approprié. Cette responsabilité comprend:

- une réflexion sur les besoins, les ressources et les priorités dans le contexte concerné;
- un choix des procédures adéquates parmi celles qui sont proposées ou parmi d'autres dont fait état la littérature;
- une gestion réaliste du projet selon une approche modulaire et par étape qui en assure la qualité;
- une collaboration et une mise en réseau avec des collègues d'autres domaines professionnels et d'autres pays;
- avant une utilisation réfléchie des procédures: la coordination des participants dans le processus de mise en relation établi localement ;
- une utilisation réfléchie des procédures;
- une communication fidèle des résultats;
- une communication précise, transparente et détaillée des conclusions.

Une affirmation selon laquelle un examen est relié au CECR ne peut être prise au sérieux qu'à partir du moment où une preuve existe que cette affirmation, fondée sur la spécification (contenus standards) et sur la définition des points de césure (performances standards), est confirmée par la validation.

Les utilisateurs du Manuel peuvent se demander avant d'entamer le processus de mise en relation:

- ce que l'approche proposée signifie, de manière générale, dans leur contexte;
- ce que l'approche proposée signifie, de façon plus spécifique (temps, ressources, etc.), dans leur contexte;

- si les différentes procédures sont praticables dans leur contexte;
- s'ils doivent se concentrer sur une ou plusieurs procédures ou bien appliquer les principes de chacun des cinq ensembles de procédures de façon limitée, en particulier si les ressources sont limitées;
- si la loi requiert la mise en relation de l'examen ou du test.

Chapitre 3: Familiarisation

3.1. Introduction

Avant d'entreprendre les activités de spécification et de standardisation, il faut organiser des tâches de «familiarisation» pour que les personnes impliquées dans une démarche de mise en relation de leurs examens avec les niveaux du Cadre aient une excellente connaissance de ce processus. Alors que la plupart des professionnels connaissent bien les tableaux du CECR plus globaux (tableau 1: échelle globale et tableau 2: grille d'auto-évaluation du Portfolio), beaucoup n'ont pas une idée très précise des caractéristiques de la compétence dans les différentes capacités langagières de l'apprenant aux différents niveaux.

Il faut faire une différence entre la familiarisation avec le CECR, avec les instruments d'évaluation à utiliser et avec les activités à entreprendre. Il n'y a pas de frontière nette entre la fin de la familiarisation et le début de la spécification ou de la standardisation; à chaque fois, les premières activités de la tâche principale s'inscrivent dans le continuum du processus de familiarisation.

Il faut aussi prendre en compte ce qui est en jeu et avoir à l'esprit le public et les applications: un panel sélectionné d'experts ou l'application du CECR par une équipe ou à l'échelle d'une institution. Il faut aussi se demander quelles activités de mise en relation pourront servir comme introduction à une session spécifique de familiarisation. Le temps que les individus vont consacrer aux activités de familiarisation dépend essentiellement du degré de familiarité qu'ils ont avec le CECR. La durée que prendra le processus de familiarisation dans sa totalité (repris avant les activités de spécification et de standardisation) dépendra de l'objectif et de l'importance du projet de mise en relation.

Les membres du groupe de travail peuvent aussi être nettement influencés par des normes institutionnelles locales données aux niveaux du CECR, ainsi que par leurs interprétations des descripteurs ou par les variantes locales des descripteurs du CECR. De plus, ils ignorent souvent qu'il existe une différence entre le niveau des descripteurs du CECR (dans toutes les sous-échelles, ainsi que dans les tableaux récapitulatifs 1, 2 et 3) et les «niveaux plus» du CECR (que l'on trouve uniquement dans les sous-échelles).

C'est avec ces éléments à l'esprit que ce chapitre propose des activités de familiarisation dans les quatre parties indiquées ci-dessous. Ces techniques sont expliquées plus en détail dans la suite du chapitre. Il est vivement conseillé aux utilisateurs de sélectionner des activités dans chaque partie au début des processus de spécification et de standardisation.

3.2. Activités préalables au séminaire

Les organisateurs d'activités de familiarisation doivent bien faire la différence entre une présentation du CECR et un séminaire/atelier de familiarisation. Alors que la première vise à présenter de façon générale l'importance et le contenu du CECR à des fins diverses, la familiarisation est supposée assurer une connaissance suffisamment fine des niveaux du CECR pour analyser et évaluer des tâches d'examens et des performances en rapport avec eux.

Le séminaire de familiarisation sera d'autant plus utile et réussi que les coordinateurs auront réuni les documents nécessaires et les informations permettant aux participants de s'y préparer en leur faisant parvenir un «paquet de pré-tâches» (par envoi postal ou par courrier électronique) deux à trois semaines avant le séminaire.

Après la première information sur le CECR, une des activités suivantes peut être choisie pour commencer le séminaire lui-même ou pour contribuer à la cohésion du groupe.

a) Lecture des parties importantes du CECR

On demande aux participants de prendre connaissance des niveaux pour pouvoir identifier les caractéristiques de chaque niveau et indiquer de façon sûre le niveau atteint par des apprenants avec lesquels ils travaillent. Le travail qui a été fait individuellement avant le séminaire peut être repris comme activité introductive ou pour «rompre la glace». Voir le tableau 3.1 à la fin de ce chapitre.

b) Prise en compte d'une sélection des questions de l'encadré

L'objectif de l'exercice est de faire prendre conscience aux participants des nombreux aspects à prendre en compte lors de la conception et de l'analyse des tâches d'examens, ainsi que de l'étendue de ce que le CECR recouvre.

On peut organiser cette activité suivant différentes modalités:

- On peut photocopier une liste de contrôle telle que celle qui est présentée ci-dessous, centrée sur la production orale, pour amener les participants à réfléchir aux différents aspects en jeu dans l'évaluation de la production orale.

Les personnes qui utilisent le Cadre pour analyser et évaluer les performances de production orale envisageront et expliciteront selon le cas:

- comment les conditions matérielles dans lesquelles l'apprenant sera amené à communiquer affecteront ce qu'il doit faire;

Approprié?
Pourquoi?

- | | |
|---|--|
| <ul style="list-style-type: none"> ▪ comment le nombre et la nature des interlocuteurs affecteront ce que l'apprenant doit faire; ▪ avec quelles contraintes de temps l'apprenant devra effectuer sa performance; ▪ dans quelle mesure les apprenants devront s'adapter au contexte mental de leur interlocuteur; ▪ comment tenir compte de la perception du niveau de difficulté d'une tâche pour l'évaluation de sa réalisation réussie et pour l'(auto-) évaluation de la compétence communicative de l'apprenant. | |
|---|--|

- Les coordinateurs peuvent sélectionner les questions des encadrés du CECR qui leur semblent pertinentes et élaborer une nouvelle liste de contrôle en fonction des capacités langagières sur lesquelles le groupe travaillera.
- Les coordinateurs peuvent s'inspirer du travail accompli par les participants lors de cette activité quand ils discuteront des exercices (f-g) consistant à trier dans la partie 3.4.

c) **Avoir accès au site de formation CEFTrain**

Le projet CEFTrain (www.helsinki.fi/project/ceftrain/index.php.35.html, www.webcef.eu/?q=node/49) a consisté à choisir des activités visant à familiariser les professeurs avec les niveaux du CECR. Il comprend des exercices avec les échelles du CECR, ainsi que les tâches et les performances (pour des enseignements dans le primaire, le secondaire et l'enseignement aux adultes) qui ont été analysées et mises en relation avec les niveaux du CECR en tenant compte des avis partagés des membres du projet.

3.3. **Activités introductives pendant le séminaire**

La première activité du séminaire consiste à présenter brièvement l'importance du CECR pour l'évaluation; par la suite, les coordinateurs organiseront une ou deux des activités présentées ci-dessous, tout en s'assurant que les participants réinvestissent le travail fait avant le séminaire.

d) **Tri du texte des différents niveaux du tableau A1**

C'est une activité qui permet de faire le lien avec le travail fait individuellement avant le séminaire.

- L'exercice consiste à demander aux participants de trier les traits caractéristiques du tableau 3.1 du Manuel qui est une simplification de la partie 3.6 du CECR. Il faut supprimer les références aux niveaux pour que les participants soient obligés de lire attentivement les descripteurs. Les coordinateurs distribuent une feuille avec les descripteurs dans le désordre et la tâche consiste à attribuer des niveaux A1 à C2 aux descripteurs.
- Une fois le travail achevé, les coordinateurs distribuent le tableau 3.1 avec les réponses.
- Les coordinateurs demandent ensuite aux participants d'échanger – en tandem ou en petits groupes – leurs points de vue sur les traits caractéristiques de chaque niveau du CECR. La meilleure façon de procéder est de demander aux participants de surligner les éléments clés.

e) **Auto-évaluation avec le tableau 2 du CECR**

Cette activité est un bon point de départ pour des groupes de participants qui connaissent déjà le Portfolio. Le tableau 3.2 représente une partie importante du PEL et on en parle souvent comme de «la grille du PEL».

- On demande aux participants de faire une auto-évaluation de leur compétence dans deux langues étrangères à l'aide de la grille du PEL (tableau 3.2 à la fin de ce chapitre). Ils en débattent ensuite avec leurs voisins. L'importance de cette discussion ne doit pas être sous-estimée. La discussion doit être dirigée de telle façon que les participants prennent conscience de l'existence de profils non uniformes. Les coordinateurs expliquent alors comment le CECR prend en compte cette non-uniformité et encourage sa reconnaissance.
- On peut également demander aux participants d'auto-évaluer leur niveau en termes de **qualité**: dans quelle mesure font-ils bien ce qu'ils disent savoir faire? Pour ce faire, on peut utiliser le tableau 3 du CECR qui définit chaque niveau pour:
 - l'étendue linguistique,
 - la correction grammaticale,
 - l'aisance,
 - la cohérence
 - et l'interaction.

Voir le tableau 3.3 à la fin de ce chapitre.

3.4. Analyse qualitative des échelles du CECR

Une fois les activités introductives achevées, il faut poursuivre la familiarisation par un travail d'approfondissement des niveaux du CECR et par des discussions sur les descripteurs spécifiques de la compétence.

f) Tri des descripteurs isolés d'une échelle du CECR

Cette tâche a le mérite d'obliger les participants à examiner les descripteurs indépendamment les uns des autres comme des critères autonomes.

- Les coordinateurs préparent des enveloppes à l'avance pour chaque participant ou pour un tandem. Chaque enveloppe contient une ou plusieurs échelles dont les descripteurs ont été découpés en bandes. Si l'on mélange des échelles apparentées (par exemple, Conversation, tours de parole dans la Discussion informelle), on doit s'assurer que le nombre de descripteurs isolés n'excède pas 40 ! En découpant les descripteurs, il faut veiller à supprimer la ligne de séparation entre deux descripteurs consécutifs afin de ne pas donner d'indication sur la capacité ou l'incapacité du coordinateur à couper droit ! On demande aussi aux participants de ne rien écrire sur les bandes pour pouvoir les réutiliser.
- Individuellement ou par deux, les participants trient alors les descripteurs selon leur niveau. Ils peuvent commencer avec «A», «B» ou «C» qu'ils divisent ensuite ou se lancer tout de suite dans les six niveaux, s'ils le souhaitent.
- Ils en discutent ensuite avec les autres participants afin d'arriver à un consensus.
- Puis ils comparent avec la bonne réponse.

Il faut s'attendre à ce que certains descripteurs ne se retrouvent pas à leur place mais, en règle générale, si l'on a pris le temps nécessaire pour atteindre un consensus, l'ordre trouvé sera plus ou moins le même que celui des échelles du CECR.

g) Reconstitution du tableau 2 du CECR

Cette activité est une variante de la précédente, mais elle utilise le tableau 3.2 de ce Manuel. Là aussi, la meilleure façon de procéder est de mettre les cases découpées dans une enveloppe.

- On distribue la grille vierge agrandie au format A3 du tableau 3.2 dont les cases ont été vidées de leur contenu. On demande aux participants de replacer les descripteurs dans les cases convenables.
- Pour éviter de faire perdre du temps aux participants, on peut affecter les descripteurs de symboles correspondant aux différentes capacités langagières. Il est en effet inutile de leur faire trouver que «Je peux utiliser des expressions et

des phrases simples pour décrire l'endroit où je vis et les gens que je connais» est un descripteur de Production orale.

- Cette activité peut aussi être menée en vidant de leur contenu la moitié seulement des cases. Il est recommandé de procéder ainsi avec des grands groupes ou dans des salles dans lesquelles il n'y a que de petites tables.

On a constaté que la combinaison de cette activité de reconstitution avec l'auto-évaluation de son propre niveau de langue ((c), voir ci-dessus) était particulièrement efficace si on la pratiquait comme suit:

- En petits groupes, les participants lisent attentivement chaque descripteur et en discutent pour reconstituer le tableau. Les coordinateurs contrôlent le travail de groupe et aident à clarifier les doutes sur l'interprétation des différents descripteurs.
- Les coordinateurs distribuent une copie du tableau 3.2 achevé et «complet» pour que les participants vérifient leur exercice de reconstitution et pour faciliter la discussion.
- On demande aux participants de faire une auto-évaluation de leur propre connaissance des langues étrangères (d'abord individuellement), puis d'en discuter avec le groupe en se référant au tableau 3.2 «Niveaux communs de compétences – Grille d'auto-évaluation».

3.5. Préparation à l'évaluation

Une fois que l'on s'est assuré que les participants se sont familiarisés avec les niveaux du CECR, on peut entamer la dernière étape de familiarisation. Cela suppose une préparation plus poussée à l'évaluation de tâches et de performances dans les différentes capacités langagières.

h) Reconstitution de la grille du CECR à utiliser

L'exercice est organisé exactement de la même façon qu'en (f) (tri des descripteurs du CECR).

Au lieu de trier les descripteurs découpés et mis dans une enveloppe, on peut utiliser une fiche type d'une liste de contrôle avec, dans le désordre, les descripteurs de la capacité langagière. C'est ensuite aux participants de rattacher chaque descripteur au niveau correspondant (comme cela est décrit dans le (d) ci-dessus).

A l'issue des discussions sur les descripteurs et les «corrections» apportées en grand groupe, les coordinateurs distribuent une liste de contrôle complétée avec les réponses.

i) Exemples filmés représentatifs de performances d'étudiants

Cette activité donne une très bonne idée correspondant à la réalité des niveaux du CECR. Elle est tout à fait appropriée, même si les participants ne vont pas travailler sur la production orale.

Les coordinateurs ne peuvent mener à bien cette activité que s'ils peuvent avoir accès aux échantillons de performances du CECR (www.ciep.fr/publi_evalcert/dvd-productions-orales-CECR/index.php). Il faut choisir avec soin les performances les plus appropriées, en termes de niveau et d'âge. La procédure à suivre peut être la suivante:

- Les coordinateurs font visionner la performance et demandent aux participants d'attribuer un niveau en utilisant le tableau 3.1.
- On distribue ensuite aux participants, avant qu'ils ne discutent entre eux, le tableau 3 du CECR (tableau 3.3) et on leur demande de confirmer le niveau choisi individuellement.
- Les coordinateurs demandent ensuite aux participants de discuter, en petits groupes, du niveau qu'ils ont attribué en se référant au tableau 3 du CECR (tableau 3.3).
- Les coordinateurs annoncent ensuite le niveau attribué à la performance et distribuent les commentaires qui justifient du niveau attribué (voir le site ci-dessus), toujours en se référant aux descripteurs du tableau 3 du CECR (tableau C2).

Les utilisateurs du Manuel peuvent se demander:

- jusqu'à quel point les participants se sont familiarisés avec les finalités et les fonctions du CECR;
- quelle est la meilleure stratégie pour renforcer la familiarisation avec le CECR;
- s'il est nécessaire de demander aux groupes de lire ou relire certains chapitres ou des parties en supplément de la section 3.6 du CECR;
- quelles questions de l'encadré peuvent être utiles;
- s'il serait judicieux de donner une tâche préliminaire sur le CECR, de recueillir le travail et l'analyser ou le faire de façon informelle;
- quelles seraient les échelles de niveaux le plus utiles pour effectuer les exercices de tri;

- s'il faut montrer des échantillons représentatifs du DVD à cette étape;
- si un moyen tel qu'un quiz serait approprié pour savoir s'il est nécessaire de renforcer la familiarisation;
- si les résultats de cette étape de familiarisation entraînent une modification de l'organisation.

Niveau		Principales caractéristiques des niveaux du CECR
Utilisateur expérimenté		On ne saurait trop insister sur le fait qu'au niveau C2 on n'a pas l'ambition d'égaliser la compétence du locuteur natif ou presque. La recherche initiale autant qu'un projet utilisant les descripteurs du CECR pour évaluer la compétence en langue maternelle (North, 2002: CECR Etudes de cas) ont montré l'existence de locuteurs bilingues bien au-dessus du niveau le plus élevé défini (C2). Wilkins a identifié un septième niveau de «Compétence ambilingue» dans sa proposition de 1978 pour une échelle européenne d'unités de crédits.
	C2	Le niveau C2 a pour but de caractériser le degré de précision, d'adéquation et d'aisance de la langue que l'on trouve dans le discours de ceux qui ont été des apprenants de haut niveau. Les descripteurs inventoriés ici comprennent: <i>transmettre les subtilités de sens avec précision en utilisant, avec une raisonnable exactitude, une gamme étendue de modalisateurs; avoir une bonne maîtrise des expressions idiomatiques et familières accompagnée de la conscience des connotations; revenir en arrière et reformuler une difficulté sans heurts de sorte que l'interlocuteur s'en aperçoive à peine.</i>
	C1	Le niveau C1 semble être caractérisé par le bon accès à une large gamme de discours qui permet une communication aisée et spontanée comme on le verra dans les exemples suivants: <i>peut s'exprimer avec aisance et spontanéité presque sans effort. A une bonne maîtrise d'un répertoire lexical large dont les lacunes sont facilement comblées par des périphrases. Il y a peu de recherche notable de certaines expressions ou de stratégies d'évitement; seul un sujet conceptuellement difficile peut empêcher que le discours ne se déroule naturellement.</i> Les capacités discursives qui caractérisent le niveau précédent se retrouvent au Niveau C1 avec encore plus d'aisance, par exemple: <i>peut choisir une expression adéquate dans un répertoire disponible de fonctions du discours pour introduire ses commentaires afin de mobiliser l'attention de l'auditoire ou de gagner du temps en gardant cette attention pendant qu'il/elle réfléchit; produit un discours clair, bien construit et sans hésitation qui montre l'utilisation bien maîtrisée des structures, des connecteurs et des articulateurs.</i>

Niveau		Principales caractéristiques des niveaux du CECR
Utilisateur indépendant	B2+	Ce Niveau B2+ correspond au degré supérieur de L'utilisateur indépendant (Vantage) . L'accent y est mis sur l'argumentation, et la conscience de la langue qui apparaît en B2 se poursuit ici. Néanmoins, on peut aussi interpréter l'accent mis sur l'argumentation et le discours social comme une importance nouvelle accordée aux capacités discursives. Ce nouveau degré de compétence discursive apparaît dans la gestion de la conversation (stratégies de coopération): <i>est capable de donner un feed-back et une suite aux déclarations et aux déductions des autres locuteurs et, ce faisant, de faciliter l'évolution de la discussion; de mettre en relation adroitement sa propre contribution et celle des autres locuteurs</i> . Il apparaît également dans la relation logique/cohésion: <i>utilise une variété de mots de liaison efficacement pour indiquer le lien entre les idées; soutient systématiquement une argumentation qui met en valeur les points significatifs et les points secondaires pertinents</i> . Enfin, c'est à ce niveau que se concentrent les descripteurs portant sur la négociation.
	B2	Les descripteurs calibrés au niveau B2 marquent une coupure importante avec ceux qui les précèdent. Par exemple, le degré élémentaire de ce niveau se concentre sur l'efficacité de l'argumentation. <i>Rend compte de ses opinions et les défend au cours d'une discussion en apportant des explications appropriées, des arguments et des commentaires; développe un point de vue en soutenant tour à tour les avantages et les inconvénients des différentes options; développe une argumentation en défendant ou en accablant un point de vue donné, prend une part active dans une discussion informelle dans un contexte familial, fait des commentaires, exprime clairement son point de vue, évalue les choix possibles, fait des hypothèses et y répond</i> . En second lieu, si l'on parcourt le niveau, on constate deux nouveaux points de convergence. Le premier est d'être capable de faire mieux que se débrouiller dans le discours social, par exemple: <i>comprendre dans le détail ce que l'on vous dit dans une langue standard courante même dans un environnement bruyant; prendre l'initiative de la parole, prendre son tour de parole au moment voulu et clore la conversation lorsqu'il faut, même si cela n'est pas toujours fait avec élégance; intervenir avec un niveau d'aisance et de spontanéité qui rend possibles les échanges avec les locuteurs natifs sans imposer de contrainte à l'une ou l'autre des parties</i> . Le second point de convergence porte sur un nouveau degré de conscience de la langue: <i>corriger les fautes qui ont débouché sur des malentendus; prendre note des «fautes préférées» et contrôler consciemment le discours pour les traquer. En règle générale, corriger les fautes et les erreurs aussitôt qu'on en prend conscience</i> .
	B1+	Ce niveau (B1+) semble correspondre à un degré élevé du niveau seuil . On y retrouve les deux mêmes traits caractéristiques auxquels s'ajoute un certain nombre de descripteurs qui se concentrent sur la quantité d'information échangée, par exemple: <i>apporte l'information concrète exigée dans un entretien ou une consultation (par exemple, décrit des symptômes à un médecin), mais avec une précision limitée; explique pourquoi quelque chose pose problème; donne son opinion sur une nouvelle, un article, un exposé, une discussion, un entretien, un documentaire et répond à des questions de détail complémentaires – les résume; mène à bien un entretien préparé en vérifiant et confirmant l'information même s'il doit parfois faire répéter l'interlocuteur dans le cas où sa réponse est longue ou rapidement énoncée; décrit comment faire quelque chose et donne des instructions détaillées; échange avec une certaine assurance une grande quantité d'informations factuelles sur des questions habituelles ou non dans son domaine</i> .

Niveau		Principales caractéristiques des niveaux du CECR
	B1	Le niveau B1 correspond aux spécifications du niveau seuil . Deux traits le caractérisent particulièrement. Le premier est la capacité de poursuivre une interaction et à obtenir ce que l'on veut dans des situations différentes, par exemple: <i>en règle générale, suit les points principaux d'une discussion assez longue à son sujet, à condition que la diction soit claire et la langue standard; reste compréhensible même si la recherche des mots et des formes grammaticales ainsi que la remédiation sont évidentes, notamment au cours de longs énoncés</i> . Le deuxième trait est la capacité de faire face habilement aux problèmes de la vie quotidienne, par exemple: <i>se débrouiller dans une situation imprévue dans les transports en commun; faire face à l'essentiel de ce qui peut arriver chez un voyageur ou au cours du voyage; intervenir sans préparation dans des conversations sur des sujets familiers</i> .
Utilisateur élémentaire	A2+	Ce niveau A2+ correspond à une capacité supérieure au niveau du Waystage (Niveau intermédiaire ou de survie) . On remarquera ici une participation plus active encore que limitée et accompagnée d'aide, par exemple: <i>comprend assez bien pour se débrouiller dans des échanges simples et courants sans effort excessif; se fait comprendre pour échanger des idées et des informations sur des sujets familiers dans des situations quotidiennes prévisibles à condition que l'interlocuteur aide, le cas échéant; se débrouille dans les situations quotidiennes dont le contenu est prévisible bien qu'en devant adapter le message et chercher ses mots; de manière plus significative, une meilleure capacité à poursuivre un monologue, par exemple, exprime ses impressions en termes simples; fait une longue description des données quotidiennes de son environnement comme les gens, les lieux, une expérience professionnelle ou académique; décrit des activités passées et des expériences personnelles; décrit des occupations quotidiennes et des habitudes; décrit des projets et leur organisation; explique ce qu'il/elle aime ou n'aime pas</i> .
	A2	Le niveau A2 (intermédiaire ou de survie) semble correspondre à la spécification du niveau Waystage . C'est à ce niveau que l'on trouvera la plupart des descripteurs qui indiquent les rapports sociaux tels que: <i>utilise les formes quotidiennes de politesse et d'adresse; accueille quelqu'un, lui demande de ses nouvelles et réagit à la réponse; invite et répond à une invitation; discute de ce qu'il veut faire, où, et fait les arrangements nécessaires; fait une proposition et en accepte une</i> . C'est ici que l'on trouvera également les descripteurs relatifs aux sorties et aux déplacements, version simplifiée de l'ensemble des spécifications transactionnelles du Niveau seuil pour adultes vivant à l'étranger telles que: <i>mener à bien un échange simple dans un magasin, un bureau de poste ou une banque; se renseigner sur un voyage; utiliser les transports en commun: bus, trains et taxis, demander des informations de base, demander son chemin et l'indiquer, acheter des billets; fournir les produits et les services nécessaires au quotidien et les demander</i> .
	A1	Le niveau A1 est le niveau le plus élémentaire d'utilisation de la langue à titre personnel – celui où l'apprenant <i>est capable d'interactions simples; peut répondre à des questions simples sur lui-même, l'endroit où il vit, les gens qu'il connaît et les choses qu'il a et en poser; peut intervenir avec des énoncés simples dans les domaines qui le concernent ou qui lui sont familiers et y répondre également</i> en ne se contentant pas de répéter des expressions toutes faites et préorganisées.

Niveau		Principales caractéristiques des niveaux du CECR
Utilisateur élémentaire	A2	C'est au niveau A2 que l'on trouvera la plupart des descripteurs qui indiquent les rapports sociaux tels que: <i>utilise les formes quotidiennes de politesse et d'adresse; accueille quelqu'un, lui demande de ses nouvelles et réagit à la réponse; invite et répond à une invitation; discute de ce qu'il veut faire, où, et fait les arrangements nécessaires; fait une proposition et en accepte une.</i> C'est ici que l'on trouvera également les descripteurs relatifs aux sorties et aux déplacements, version simplifiée de l'ensemble des spécifications transactionnelles du niveau seuil pour adultes vivant à l'étranger telles que: <i>mener à bien un échange simple dans un magasin, un bureau de poste ou une banque; se renseigner sur un voyage; utiliser les transports en commun: bus, trains et taxis, demander des informations de base, demander son chemin et l'indiquer, acheter des billets; fournir les produits et les services nécessaires au quotidien et les demander.</i>
	A1	Le niveau A1 est le niveau le plus élémentaire d'utilisation de la langue à titre personnel – celui où l'apprenant est <i>capable d'interactions simples; peut répondre à des questions simples sur lui-même, l'endroit où il vit, les gens qu'il connaît et les choses qu'il a et en poser; peut intervenir avec des énoncés simples dans les domaines qui le concernent ou qui lui sont familiers et y répondre également</i> en ne se contentant pas de répéter des expressions toutes faites et préorganisées.

Tableau 3.2. Grille d'auto-évaluation

Comprendre

	A1	A2	B1	B2	C1	C2
Écouter	Je peux comprendre des mots familiers et des expressions très courantes au sujet de famille et de moi-même, de ma famille et de l'environnement concret et immédiat, si les gens parlent lentement et distinctement.	Je peux comprendre des expressions et un vocabulaire très fréquent relatifs à ce qui me concerne de très près (par exemple moi-même, ma famille, les achats, l'environnement proche, le travail). Je peux saisir l'essentiel d'annonces et de messages simples et clairs.	Je peux comprendre les points essentiels quand un langage clair et standard est utilisé et s'il s'agit de sujets familiers concernant le travail, l'école, les loisirs, etc. Je peux comprendre l'essentiel de nombreuses émissions de radio ou de télévision sur l'actualité ou sur des sujets qui m'intéressent à titre personnel ou professionnel si l'on parle d'une façon relativement lente et distincte.	Je peux comprendre des conférences et des discours assez longs et même suivre une argumentation complexe si le sujet m'en est relativement familier. Je peux comprendre la plupart des émissions de télévision sur l'actualité et les informations. Je peux comprendre la plupart des films en langue standard.	Je peux comprendre un long discours même s'il n'est pas clairement structuré et que les articulations sont seulement implicites et pas signalées de façon explicite. Je peux comprendre les émissions de télévision et les films sans trop d'effort.	Je n'ai aucune difficulté à comprendre le langage oral, que ce soit dans les conditions du direct ou dans les médias et quand on parle vite, à condition d'avoir du temps pour me familiariser avec un accent particulier.
Lire	Je peux comprendre des noms familiers, des phrases très simples, par exemple dans des annonces, des affiches ou des catalogues.	Je peux lire des textes courts très simples. Je peux trouver une information particulière prévisible dans des documents courants comme les petites publicités, les prospectus, les menus et les horaires, et je peux comprendre des lettres personnelles courtes et simples.	Je peux comprendre des textes rédigés essentiellement dans une langue courante ou relative à mon travail. Je peux comprendre la description d'événements, l'expression de sentiments et de souhaits dans des lettres personnelles.	Je peux lire des articles et des rapports sur des questions contemporaines dans lesquels les auteurs adoptent une attitude particulière ou un certain point de vue. Je peux comprendre un texte littéraire contemporain en prose.	Je peux comprendre des textes factuels ou littéraires longs et complexes et en apprécier les différences de style. Je peux comprendre des articles spécialisés et de longues instructions techniques même lorsqu'ils ne sont pas en relation avec mon domaine.	Je peux lire sans effort tout type de texte, même abstrait ou complexe quant au fond ou à la forme, par exemple un manuel, un article spécialisé ou une œuvre littéraire.

Parler

	A1	A2	B1	B2	C1	C2
Prende part à une conversation	Je peux communiquer, de façon simple, à condition que l'interlocuteur soit disposé à répéter ou à reformuler ses phrases plus lentement et à m'aider à formuler ce que j'essaie de dire. Je peux poser des questions simples sur des sujets familiers ou sur ce dont j'ai immédiatement besoin, ainsi que répondre à de telles questions.	Je peux communiquer lors de tâches simples et habituelles ne demandant qu'un échange d'informations simple et direct sur des sujets et des activités familiers. Je peux avoir des échanges très brefs même si, en règle générale, je ne comprends pas assez pour poursuivre une conversation.	Je peux faire face à la majorité des situations que l'on peut rencontrer au cours d'un voyage dans un pays où la langue est parlée. Je peux prendre part sans préparation à une conversation sur des sujets familiers ou d'intérêt personnel qui concernent la vie quotidienne (par exemple famille, loisirs, travail, voyage et actualité).	Je peux communiquer avec un degré de spontanéité et d'aisance qui rend possible une interaction normale avec un locuteur natif. Je peux participer activement à une conversation dans des situations familières, présenter et défendre mes opinions.	Je peux m'exprimer spontanément et couramment sans trop apparemment devoir chercher mes mots. Je peux utiliser la langue de manière souple et efficace pour des relations sociales ou professionnelles. Je peux exprimer mes idées et opinions et lier mes interventions de manière adroite à celles de mes interlocuteurs.	Je peux présenter une description ou une argumentation claire et fluide dans un style adapté au contexte, construire une présentation de façon logique et aider mon auditeur à remarquer et à se rappeler les points importants.
S'exprimer oralement en continu	Je peux utiliser des expressions et des phrases simples pour décrire mon lieu d'habitation et les gens que je connais.	Je peux utiliser une série de phrases ou d'expressions pour décrire en termes simples ma famille et d'autres gens, mes conditions de vie, ma formation et mon activité professionnelle actuelle ou récente.	Je peux articuler des expressions de manière simple afin de raconter des expériences et des événements, mes rêves, mes espoirs ou mes buts. Je peux brièvement donner les raisons et explications de mes opinions ou projets. Je peux raconter une histoire ou l'intrigue d'un livre ou d'un film et exprimer mes réactions.	Je peux m'exprimer de façon claire et détaillée sur une grande gamme de sujets relatifs à mes centres d'intérêt. Je peux développer un point de vue sur un sujet d'actualité et expliquer les avantages et les inconvénients de différentes possibilités.	Je peux présenter des descriptions claires et détaillées de sujets complexes, en intégrant des thèmes qui leur sont liés, en développant certains points et en terminant mon intervention de façon appropriée.	Je peux présenter une description ou une argumentation claire et fluide dans un style adapté au contexte, construire une présentation de façon logique et aider mon auditeur à remarquer et à se rappeler les points importants.

Écrire	A1	A2	B1	B2	C1	C2
	Je peux écrire une courte carte postale simple, par exemple de vacances. Je peux porter des détails personnels dans un questionnaire, inscrire par exemple mon nom, ma nationalité et mon adresse sur une fiche d'hôtel.	Je peux écrire des notes et messages simples et courts. Je peux écrire une lettre personnelle très simple, par exemple de remerciements.	Je peux écrire un texte simple et cohérent sur des sujets familiers ou qui m'intéressent personnellement. Je peux écrire des lettres personnelles pour décrire expériences et impressions.	Je peux écrire des textes clairs et détaillés sur une grande gamme de sujets relatifs à mes intérêts. Je peux écrire un essai ou un rapport en transmettant une information ou en exposant des raisons pour ou contre une opinion donnée. Je peux écrire des lettres qui mettent en valeur le sens que j'attribue personnellement aux événements et aux expériences.	Je peux m'exprimer dans un texte clair et bien structuré et développer mon point de vue. Je peux écrire sur des sujets complexes dans une lettre, un essai ou un rapport, en soulignant les points que je juge importants. Je peux adopter un style adapté au destinataire.	Je peux écrire un texte clair, fluide et stylistiquement adapté aux circonstances. Je peux rédiger des lettres, rapports ou articles complexes, avec une construction claire et logique permettant au lecteur d'en saisir et de mémoriser les points importants. Je peux résumer et critiquer par professionnel ou une œuvre littéraire.

Tableau 3.2. Grille d'auto-évaluation

	Étendue	Correction	Aisance	Interaction	Cohérence
C2	Montre une grande souplesse dans la reformulation des idées sous des formes linguistiques différentes lui permettant de transmettre avec précision des nuances fines de sens afin d'insister, de discriminer ou de lever l'ambiguïté. A aussi une bonne maîtrise des expressions idiomatiques et familières.	Maintient constamment un haut degré de correction grammaticale dans une langue complexe, même lorsque l'attention est ailleurs (par exemple, la planification ou l'observation des réactions des autres).	Peut s'exprimer longuement, spontanément dans un discours naturel en évitant les difficultés ou en les rattrapant avec assez d'habileté pour que l'interlocuteur ne s'en rende presque pas compte.	Peut interagir avec aisance et habileté en relevant et utilisant les indices non verbaux et intonatifs sans effort apparent. Peut intervenir dans la construction de l'échange de façon tout à fait naturelle, que ce soit au plan des tours de parole, des références ou des allusions, etc.	Peut produire un discours soutenu cohérent en utilisant de manière complète et appropriée des structures organisationnelles variées ainsi qu'une gamme étendue de mots de liaison et autres articulateurs.
C1	A une bonne maîtrise d'une grande gamme de discours parmi lesquels il peut choisir la formulation lui permettant de s'exprimer clairement et dans le registre convenable sur une grande variété de sujets d'ordre général, éducationnel, professionnel ou de loisirs, sans devoir restreindre ce qu'il/elle veut dire.	Maintient constamment un haut degré de correction grammaticale; les erreurs sont rares, difficiles à repérer et généralement auto-correctées quand elles surviennent.	Peut s'exprimer avec aisance et spontanéité presque sans effort. Seul un sujet conceptuellement difficile est susceptible de gêner le flot naturel et fluide du discours.	Peut choisir une expression adéquate dans un répertoire courant de fonctions discursives, en préambule à ses propos, pour obtenir la parole ou pour gagner du temps pour la garder pendant qu'il/elle réfléchit.	Peut produire un texte clair, fluide et bien structuré, démontrant un usage contrôlé de moyens linguistiques de structuration et d'articulation.
B2+					

	Étendue	Correction	Aisance	Interaction	Cohérence
B2	Possède une gamme assez étendue de langue pour pouvoir faire des descriptions claires, exprimer son point de vue et développer une argumentation sans chercher ses mots de manière évidente.	Montre un degré assez élevé de contrôle grammatical. Ne fait pas de fautes conduisant à des malentendus et peut le plus souvent les corriger lui/elle-même.	Peut parler relativement longtemps avec un débit assez régulier; bien qu'il /elle puisse hésiter en cherchant structures ou expressions, l'on remarque peu de longues pauses.	Peut prendre l'initiative de la parole et son tour quand il convient et peut clore une conversation quand il le faut, encore qu'éventuellement sans élégance. Peut faciliter la poursuite d'une discussion sur un terrain familier en confirmant sa compréhension, en sollicitant les autres, etc.	Peut utiliser un nombre limité d'articulateurs pour lier ses phrases en un discours clair et cohérent bien qu'il puisse y avoir quelques «sauts» dans une longue intervention.
B1+					
B1	Possède assez de moyens linguistiques et un vocabulaire suffisant pour s'en sortir avec quelques hésitations et quelques périphrases sur des sujets tels que la famille, les loisirs et centres d'intérêt, le travail, les voyages et l'actualité.	Utilise de façon assez exacte un répertoire de structures et «schémas» fréquents, courants dans des situations prévisibles.	Peut discourir de manière compréhensible, même si les pauses pour chercher ses mots et ses phrases et pour faire ses corrections sont très évidentes, particulièrement dans les séquences plus longues de production libre.	Peut engager, soutenir et clore une conversation simple en tête-à-tête sur des sujets familiers ou d'intérêt personnel. Peut répéter une partie de ce que quelqu'un a dit pour confirmer une compréhension mutuelle.	Peut relier une série d'éléments courts, simples et distincts en une suite linéaire de points qui s'enchaînent.
A2+					

	Étendue	Correction	Aisance	Interaction	Cohérence
A2	Utilise des structures élémentaires constituées d'expressions mémorisées, de groupes de quelques mots et d'expressions toutes faites afin de communiquer une information limitée dans des situations simples de la vie quotidienne actualité.	Utilise des structures simples correctement, mais commet encore systématiquement des erreurs élémentaires.	Peut se faire comprendre dans une brève intervention même si la reformulation, les pauses et les faux démarrages sont évidents.	Peut répondre à des questions et réagir à des déclarations simples. Peut indiquer qu'il/elle suit, mais est rarement capable de comprendre assez pour soutenir la conversation de son propre chef.	Peut relier des groupes de mots avec des connecteurs simples tels que «et», «mais» et «parce que».
A1	Possède un répertoire élémentaire de mots et d'expressions simples relatifs à des situations concrètes particulières.	A un contrôle limité de quelques structures syntaxiques et de formes grammaticales simples appartenant à un répertoire mémorisé.	Peut se débrouiller avec des énoncés très courts, isolés, généralement stéréotypés, avec de nombreuses pauses pour chercher ses mots, pour prononcer les moins familiers et pour remédier à la communication.	Peut répondre à des questions simples et en poser sur des détails personnels. Peut interagir de façon simple, mais la communication dépend totalement de la répétition avec un débit plus lent, de la reformulation et des corrections.	Peut relier des mots ou groupes de mots avec des connecteurs très élémentaires tels que «et» ou «alors».

Tableau 3.3. Grille des critères d'évaluation de l'oral (Tableau 3 du CECR)

Chapitre 4: Spécification

4.1. Introduction

Ce chapitre traite de l'analyse du contenu d'un examen ou d'un test dans le but de décrire le ou les niveaux du CECR qu'ils recouvrent. La procédure proposée peut prendre la forme d'un débat ou d'une analyse individuelle suivie d'un débat.

Au final, en se fondant sur les spécifications, l'institution disposera de descriptions détaillées lui permettant de déclarer le degré de relation de ses examens avec les catégories et les niveaux du CECR.

Toutefois, comme cela a été précisé dans le chapitre 2, la déclaration du degré de relation n'est recevable que si, parallèlement, sont apportées des preuves de bonnes pratiques, d'une validité interne convenable et de procédures qualité adéquates pour toutes les étapes du développement et de l'administration de l'examen ou du test.

Les procédures de spécifications exposées dans ce chapitre impliquent quatre étapes:

- assurer une familiarisation convenable avec le CECR (chapitre 3);
- analyser le contenu de l'examen ou du test en question par rapport aux catégories pertinentes du CECR; l'utilisateur devra décrire un domaine évalué dans son examen ou son test et qui s'avérerait non traité dans le CECR;
- mettre en relation l'examen ou le test avec l'échelle de descripteurs pertinente du CECR, sur la base de l'analyse de contenus;
- faire une première déclaration sur le degré de mise en relation de l'examen ou du test avec l'un des niveaux du CECR, en se fondant sur l'analyse de contenus.

Ces procédures impliquent trois types d'activités:

- les activités de familiarisation décrites dans le chapitre 3;
- la description détaillée du contenu de l'examen de langue, consignée dans un certain nombre de fiches complétées;
- l'utilisation des descripteurs adéquats du CECR afin de relier l'examen de langue à ses niveaux et à ses catégories.

Ces procédures liées aux spécifications donnent aux concepteurs d'examens l'occasion:

- d'être encore plus sensibles à l'importance d'une bonne analyse du contenu d'un examen;
- de se familiariser avec l'utilisation du CECR pour la planification et la description des examens de langue;
- de décrire et d'analyser en détail le contenu d'un examen ou d'un test;

- de fournir la preuve de la qualité de leur examen ou de leur test;
- de fournir la preuve de la relation de leur examen ou de leur test avec les niveaux du CECR;
- d'apporter des conseils aux rédacteurs d'items;
- d'accroître, pour les enseignants, les évaluateurs, les utilisateurs d'examens et les candidats, la transparence des contenus, de la qualité et de la relation d'un examen ou d'un test avec le CECR. Les fiches à compléter ont une fonction de sensibilisation (processus) et seront utilisées pour étayer la déclaration qui sera faite (produit final).

Les procédures décrites ici ont été spécialement conçues pour ce Manuel. Il en existe cependant d'autres. Les utilisateurs de ce Manuel peuvent consulter des procédures d'analyses descriptives permettant de relier un examen à un cadre de référence (par exemple Alderson *et al.*, 1995, Chapitre 2; Davidson et Lynch, 1993, 2002; Lynch et Davidson, 1994, 1998).

4.2. Description générale de l'examen

La première étape consiste en une définition et une description claires de l'examen ou du test que l'on va relier au CECR. Ce processus de sensibilisation ne peut être entrepris par une seule personne (chercheur ou membre de l'équipe). Cet exercice met parfois en évidence un manque de cohérence entre les spécifications officielles de l'examen – qui n'ont peut-être pas été modifiées depuis des années – et l'examen lui-même – tel qu'il a été administré récemment. L'exercice est assurément plus facile s'il existe des spécifications formelles de l'examen. S'il n'en existe pas, le procédé consistant à compléter les fiches de ce chapitre aidera les utilisateurs à prendre en compte certains aspects qui devraient faire partie intégrante de ces spécifications.

On trouvera les fiches suivantes en annexe, partie 2:

A1	Description générale de l'examen
A2	Conception de l'examen
A3	Correction
A4	Notation
A5	Communication des résultats
A6	Analyses des données
A7	Justification des décisions

Avant de compléter les fiches, les utilisateurs doivent se munir, d'une part, des spécifications et, d'autre part, des copies des trois derniers examens administrés aux candidats. S'il s'agit de relier au CECR une suite d'examens de différents niveaux, une fiche par examen devra être complétée.

La fiche A1 permet de définir les buts et les objectifs de l'examen, ainsi que sa population cible. Elle permet également d'avoir une vue d'ensemble des activités communicatives évaluées, des différentes épreuves, ainsi que des renseignements fournis et de la façon dont les résultats sont communiqués aux utilisateurs (candidats et centres d'examens).

Les fiches A2 à A6 décrivent les étapes les plus importantes du cycle de conception, développement et administration d'un examen. On y consignera des informations sur la conception, la correction, la notation, la façon de communiquer les résultats et les analyses de données:

- fiche A2: processus de développement;
- fiche A3: critères de correction et barèmes de notation pour chaque épreuve;
- fiche A4: notation et procédures de définition des points de césure pour chaque épreuve;
- fiche A5: communication des résultats;
- fiche A6: analyses et procédures de révision;
- fiche A7: (justification des décisions). Le concepteur d'examen pourra ici expliquer et justifier ses décisions. Par exemple, pour quelles raisons certains domaines sont-ils évalués et d'autres non? Pourquoi une pondération particulière est-elle utilisée? Pourquoi la double correction n'est-elle qu'exceptionnellement mise en œuvre? Pour quelle raison ne fournit-on pas les résultats par épreuve ou par compétence? Cela relève-t-il d'un problème de fiabilité ou d'une décision politique?
- fiche A8: elle permet de consigner l'estimation initiale de l'institution quant au niveau global du CECR évalué par l'examen.

Estimation initiale du niveau global du CECR		
<input type="checkbox"/> A1 <input type="checkbox"/> <input type="checkbox"/> A2 <input type="checkbox"/>	<input type="checkbox"/> B1 <input type="checkbox"/> <input type="checkbox"/> B2 <input type="checkbox"/>	<input type="checkbox"/> C1 <input type="checkbox"/> <input type="checkbox"/> C2 <input type="checkbox"/>
Brève justification, références à de la documentation		

Fiche A8: Estimation initiale du niveau global du CECR

Le processus détaillé de spécification est exposé dans les fiches A9 à A22.

La fiche A23 présente les résultats du processus de spécification sous la forme d'un graphique illustrant les catégories et les niveaux pertinents du CECR couverts par l'examen analysé. Cette fiche est traitée et illustrée au chapitre 4.4.

Les procédures sont strictement les mêmes pour un examen de langue générale et pour un examen sur objectifs spécifiques. Le CECR prend en effet les différents domaines en compte (public, personnel, éducationnel et professionnel). De même, si les activités de communication langagière sont regroupées dans les catégories «Réception, interaction, production et médiation» plutôt que sous les quatre capacités langagières traditionnelles, c'est afin de pouvoir prendre en compte efficacement les objectifs spécifiques éducationnels et professionnels.

4.3. Outils disponibles pour la spécification

Ces outils liés au CECR sont de trois types. Outre le CECR lui-même, traduit, à la date de cette publication, en 36 langues, on trouvera :

- les tableaux et les fiches annexés à ce Manuel;
- les grilles d'analyses de contenus qui permettent de détailler de façon extrêmement fine les tâches proposées dans l'examen, en les classant selon des critères standards;
- les référentiels pour les différentes langues, particulièrement utiles pour les spécifications linguistiques.

4.3.1. Tableaux et fiches

Cette partie propose une série de tableaux tirés des échelles de descripteurs du CECR et accompagnés de fiches à compléter. Le CECR étant extrêmement détaillé, le nombre de fiches est considérable. Elles sont disponibles et peuvent être téléchargées sur le site www.coe.int/lang.

Dans les études de cas qui ont conduit à la rédaction de ce Manuel, plusieurs utilisateurs ont indiqué que compléter ces fiches s'avérait être une bonne méthode pour reconsidérer l'étendue de ce que recouvre un examen et pour réévaluer sa fiabilité.

4.3.2. Grilles d'analyse de contenus

Les grilles d'analyse de contenus du CECR pour la réception orale et écrite, ainsi que pour la production orale et écrite, ont été conçues pour que les utilisateurs de ce Manuel

puissent décrire leur examen de façon bien plus détaillée que ce que permettent de faire les sous-échelles du CECR et les tableaux de l'annexe A.

Les *grilles d'analyse de contenus pour la réception orale et écrite du CECR* sont en ligne et permettent aux concepteurs d'examens et de tests d'analyser les épreuves de réception orale et écrite afin de les relier au CECR.

La grille permet de consigner, à partir d'une série de choix tirés directement ou indirectement du CECR, les caractéristiques de chaque tâche, de chaque support, de chaque item de l'examen ou du test: source, type de discours, niveau de difficulté estimé, etc.). Une excellente connaissance du CECR est naturellement nécessaire pour pouvoir utiliser les grilles de façon totalement efficace. Une composante «familiarisation» avec le CECR est par conséquent comprise dans cette procédure afin d'apporter des conseils plus approfondis.

Les *grilles d'analyse des tâches de production orale et écrite du CECR* ont été aussi conçues pour aider les utilisateurs à décrire de façon standardisée les caractéristiques des tâches de leurs examens et leur liaison avec le CECR. Les grilles, modifiables en tant que de besoin, sont toutes disponibles sur le site du Conseil de l'Europe. Deux modes d'utilisation sont possibles pour chacune des deux grilles: un pour l'analyse et l'autre pour la présentation du rapport.

Les versions les plus récentes des grilles, accompagnées d'échantillons représentatifs utilisant des grilles complétées, peuvent être téléchargées sur le site www.coe.int/portfolio.

4.3.3. Ouvrages de référence

Dans les procédures de spécification, l'analyse de contenus se réfère principalement au CECR lui-même. Cependant, en tant que *cadre commun*, le CECR ne traite par définition d'aucune langue en particulier.

4.4. Procédures

Avant de compléter les fiches proposées dans l'annexe A que l'on trouve aussi sur le site www.coe.int/lang, les procédures impliquent que vous consultiez le CECR, les annexes de ce Manuel et les autres ouvrages de référence cités ci-dessus.

1. **Choix de la commission:** la première étape est la mise en place d'une commission d'experts, si possible mixte (appartenant à l'institution/organisation et extérieurs), et la désignation d'un coordinateur ou d'une coordinatrice. Ce groupe d'experts internes et externes devrait être constitué de représentants des différentes étapes du développement d'un examen ou d'un test de langue.

2. **Familiarisation:** avant de mettre en œuvre les procédures de spécification, il est essentiel que la commission se familiarise avec le CECR lui-même. La commission doit donc commencer son travail par les activités de familiarisation du chapitre 3.
3. **Choix de la méthode:** une fois cette étape effectuée, le groupe doit prendre connaissance des multiples fiches et tableaux associés, ainsi que des outils de spécification cités au paragraphe 4.2. Il décidera alors du choix de la méthode et des fiches et tableaux qui seront complétés. Il n'est pas prévu que toutes les fiches de l'annexe A soient complétées. Il est rappelé que seules les fiches correspondant aux contenus de l'examen doivent être complétées; le groupe doit choisir les fiches pertinentes pour l'analyse de l'examen en question. Exemple: si un examen comporte uniquement des tâches lexicales, seules les fiches correspondantes seront complétées et seule l'échelle du niveau de vocabulaire sera examinée. Autre exemple: si un examen mesure plusieurs compétences linguistiques dans différentes capacités langagières, on devra alors compléter un plus grand nombre de fiches et examiner plus d'échelles.

La norme minimale est que les fiches suivantes soient complétées:

- les fiches de la phase 1 (Description générale: A1 à A7);
 - la fiche A8 (Première estimation du niveau global de l'examen);
 - certaines des fiches numérotées de A9 à A22 – qui correspondent à l'examen ou au test en question;
 - la fiche A23 (Représentation graphique de la relation de l'examen avec les niveaux du CECR);
 - la fiche A24 (Confirmation de l'estimation du niveau global de l'examen);
 - les preuves pertinentes qui permettent d'étayer la déclaration.
4. **Activités communicatives langagières:** on complètera normalement en premier les fiches portant sur les activités langagières communicatives (fiches A9-A18). Comme cela a été précisé ci-dessus, chacune des fiches peut être complétée par la personne appropriée de l'institution impliquée. On peut cependant souhaiter procéder de façon plus interactive. L'information consignée dans les fiches sera plus fiable si plus d'une personne est impliquée. Chaque membre de la commission va donc compléter tout ou partie des fiches sélectionnées. Un consensus devra ensuite être obtenu grâce à la confrontation des fiches complétées.

Le tableau 4.1 présente une vue d'ensemble des fiches et des échelles du CECR qui y sont reliées. A la fin de la plupart des fiches, il est demandé aux utilisateurs de comparer l'épreuve en question avec la sous-échelle correspondante du CECR.

Fiche	Activité communicative langagière	Fiche	Echelle
A9	Compréhension de l'oral	✓	✓
A10	Compréhension de l'écrit	✓	✓
A11	Interaction orale	✓	✓
A12	Interaction écrite	✓	✓
A13	Production orale	✓	✓
A14	Production écrite	✓	✓
A15	Combinaisons de compétences intégrées	✓	
A16	Compétences intégrées	✓	✓
A17	Médiation orale	✓	
A18	Médiation écrite	✓	

*Tableau 4.1: Fiches et échelles du CECR
pour les activités langagières communicatives*

	Réception		Interaction		Production		Médiation	
	Compréhension de l'oral	Compréhension de l'écrit	Interaction orale	Interaction écrite	Production orale	Production écrite	Médiation orale	Médiation écrite
Compétence linguistique								
Etendue linguistique générale	✓	✓	✓	✓	✓	✓	✓	✓
Etendue du vocabulaire	✓	✓	✓	✓	✓	✓	✓	✓
Maîtrise du vocabulaire			✓	✓	✓	✓	✓	✓
Correction grammaticale			✓	✓	✓	✓	✓	✓
Maîtrise du système phonologique			✓		✓		✓	
Maîtrise de l'orthographe				✓		✓		✓
Compétence sociolinguistique								
Correction sociolinguistique	✓	✓	✓	✓	✓	✓	✓	✓
Compétence pragmatique								
Souplesse			✓	✓			✓	✓
Tours de parole			✓					
Développement thématique	✓	✓		✓	✓	✓	✓	✓
Cohésion et cohérence	✓	✓			✓	✓	✓	✓
Aisance à l'oral			✓		✓		✓	

	Réception		Interaction		Production		Médiation	
Précision	✓	✓			✓	✓	✓	✓
Compétence stratégique								
Reconnaître des indices et faire des déductions	✓	✓					✓	✓
Tours de parole (reprise)			✓					
Coopérer			✓	✓				
Faire clarifier			✓	✓				
Planifier					✓	✓		✓
Compenser			✓	✓	✓	✓	✓	✓
Contrôler et corriger			✓	✓	✓	✓	✓	✓

Tableau 4.2: Echelles du CECR pour les aspects de la compétence communicative

5. **Compétence communicative langagière:** on complétera ensuite les fiches qui concernent les aspects de la compétence communicative langagière (fiches A19-A22). Le tableau 4.2 donne une vue d'ensemble des différentes compétences communicatives pour lesquelles il est possible de consigner des informations. Cette partie est organisée différemment. Un tableau des descripteurs du CECR est fourni. Les utilisateurs doivent ensuite renseigner la fiche correspondante sur la base d'une analyse des épreuves de l'examen ou du test en question. A la fin de chaque fiche, les utilisateurs comparent l'examen et l'échelle correspondante du CECR. Une description, ainsi qu'une indication du niveau de chacun des aspects pertinents des compétences retenues dans le CECR sont demandées. Le même groupe d'experts peut compléter les fiches de façon interactive.

Les fiches sont proposées dans cet ordre:

- Réception: fiche A19;
- Interaction: fiche A20;
- Production: fiche A21;
- Médiation: fiche A22.

Aucune échelle du CECR n'est fournie pour la médiation. Les utilisateurs se référeront aux descripteurs pour la réception et la production.

4.5. Déclaration du niveau: représentation graphique de la relation de l'examen avec le CECR

Une fois l'examen analysé en fonction des catégories du CECR, le résultat obtenu doit être présenté sous la forme d'un graphique montrant clairement la relation avec les niveaux du CECR. Cette représentation permet de visualiser le contenu de l'examen étudié, rapporté aux sous-échelles appropriées du CECR pour ce qui concerne les activités de communication langagière et les aspects de la compétence linguistique (cf. ci-dessous un exemple de fiche A23 complétée).

C2								
C1								
B2.2								
B2								
B1.2								
B1								
A2.2								
A2								
A1								
Panorama	Compréhension de l'oral	Compréhension de l'écrit	Conversation sociale	Echange d'information	Notes, messages et formulaires	Socio-linguistique	Pragmatique	Linguistique

Fiche A23: Représentation graphique de la relation de l'examen aux niveaux du CECR (exemple)

Dans le graphique ci-dessus, l'axe Y (vertical, à gauche) représente les niveaux du CECR. Sur l'axe X, on représentera la compétence langagière générale et les activités communicatives langagières, ainsi que les aspects de la compétence linguistique. Chaque colonne a comme intitulé une catégorie pertinente du CECR. Les cases qui représentent l'examen ou les épreuves traités seront ombrées. Si l'examen est d'un niveau plus élevé dans certaines catégories, on le montrera en ombrant les cases correspondantes comme dans l'exemple de la fiche A23 ci-dessus.

L'intitulé des colonnes de la fiche A23 peut ne pas correspondre à celui qui a été donné aux épreuves de l'examen. Quelques intitulés peuvent correspondre aux épreuves, mais il est possible d'en ajouter d'autres, en tant que de besoin. Il se peut, par exemple, que l'examen étudié ne propose pas d'épreuve spécifique pour la compétence linguistique, mais que les concepteurs de l'examen veuillent cependant indiquer aux utilisateurs le niveau de compétence linguistique attendu.

Les démarches présentées dans ce chapitre mettent l'accent à la fois sur le *processus* et sur le *résultat*. On encourage les praticiens à suivre un processus d'analyse de contenus et de mise en relation avec le CECR. On recommande vivement de réexaminer chaque hypothèse sur le niveau avancée au cours du processus. Il est fort probable que

l'estimation initiale donnée dans la fiche A8 doit être modifiée. Les utilisateurs doivent reconsidérer les analyses et proposer un jugement raisonné. L'estimation (fiche A8) est confirmée ou révisée dans la fiche A24.

Les chapitres suivants fournissent des outils qui permettent de renforcer la déclaration de niveau. Une recherche plus poussée et une analyse plus approfondie lors d'étapes ultérieures peuvent entraîner une révision de la déclaration avancée. L'exactitude de la déclaration est subordonnée à un large processus de vérification argumentée. On recommande vivement aux concepteurs d'examens d'impliquer leurs collègues dans des débats et des échanges tout au long du processus.

Estimation confirmée (déclaration) du niveau global du CECR		
<input type="checkbox"/> A1	<input type="checkbox"/> B1	<input type="checkbox"/> C1
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/> A2	<input type="checkbox"/> B2	<input type="checkbox"/> C2
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Brève justification, références à de la documentation. Si cette fiche présente une conclusion différente de l'estimation initiale consignée dans la fiche A8, merci de commenter les raisons principales de ce changement.</p>		

Fiche A24: Estimation confirmée (déclaration) du niveau global du CECR

<p><i>Les utilisateurs de ce Manuel peuvent se demander:</i></p> <ul style="list-style-type: none"> ▪ s'il est important de réunir et/ou d'analyser des informations ou des données avant d'entreprendre l'étape de spécification; ▪ s'ils utiliseront les grilles d'analyse de contenus du CECR; ▪ si tous les examens ou les tests peuvent être reliés au CECR; ▪ si le fait d'achever l'étape de spécification présage des changements dans le plan initial d'utilisation de ce Manuel; ▪ si l'expérience acquise à l'issue de l'étape de spécification implique, dans l'examen ou le test analysé, des changements qui pourraient intervenir lors de la prochaine réforme programmée; ▪ comment ils décideront que l'étape de spécification a été achevée de façon satisfaisante.

Chapitre 5: Formation à la standardisation et au calibrage

5.1. Introduction

Le but de la démarche de mise en relation des examens avec les niveaux du CECR est de permettre une catégorisation des candidats en termes de niveaux de compétences du CECR, de telle façon que cette catégorisation reflète de façon fiable ce que signifient les niveaux du CECR. Si on considère qu'un étudiant est au niveau B1, il faut être tout à fait certain que cet étudiant est vraiment représentatif des descripteurs de ce niveau. Il s'agit là de la validité. Les procédures qui suivent renvoient à la définition des points de césure (*standard setting*).

Il existe deux grandes façons d'attribuer des niveaux à des candidats. Il peut s'agir soit d'un simple jugement global de la part du professeur ou de l'examineur, soit des notes qui sont attribuées au résultat de l'examen. La première option est en général choisie pour les capacités de production, alors que la deuxième concerne généralement les capacités de réception. La distinction n'est pourtant pas aussi tranchée. Dans des épreuves de production écrite, parmi les deux ou trois tâches proposées, chaque tâche peut être notée en fonction de critères analytiques. La totalité des notes obtenues par un candidat peut être traitée de la même façon que le résultat d'une épreuve de réception écrite comportant un certain nombre d'items séparés. Pour éviter tout malentendu, on utilisera respectivement les termes d'examen indirect (examens avec des résultats à base de notes) et d'examen direct (examens évalués de façon globale).

- **Examens directs.** Dans des examens évalués de façon globale, le jugement sur le niveau (les six niveaux du CECR) est direct et il est pour cette raison important d'aider les évaluateurs à émettre des jugements valides. Le principal outil utilisé pour ce genre particulier de définition des points de césure (*standard setting*) est appelé **calibrage**. Le calibrage consiste à proposer un (ou plusieurs) échantillon(s) représentatif(s) illustrant des performances à un niveau donné à la fois pour la formation à la standardisation et comme outil de référence pour les décisions ultérieures concernant des performances de candidats.
- **Examens indirects.** Pour les examens avec des résultats à partir de notes, il faut établir des performances standards. La performance standard est la limite entre deux niveaux de l'échelle continue, indiquée par un examen, et qui est représentée par une note de césure. Une note de césure de 30, par exemple, signifie qu'une note de 30 ou plus acquise à l'examen prouve qu'un certain niveau ou un niveau plus élevé est atteint (par exemple B1), alors qu'un résultat moins élevé correspondra à un niveau plus bas que le niveau de la note de césure (dans ce cas B1). On appelle généralement le processus pour arriver à une note de césure la «**définition des points de césure**» (*standard setting*). Dans le cas des capacités de réception (écrite et orale) ou des compétences sous-

jacentes (grammaire, lexicque), il est important de prendre des décisions sur ces notes de césure.

Les procédures de **calibrage** et de **définition de points de césure** (*standard setting*) supposent des décisions collectives qui doivent être soigneusement préparées par une formation adéquate. Le but principal de ce chapitre est d'aider à cette formation.

5.2. La formation nécessaire

L'objectif de cette partie est de décrire une suite de procédures:

- a) pour aider le groupe d'experts à atteindre une compréhension commune des niveaux du CECR;
- b) pour vérifier que la compréhension commune est vraiment atteinte;
- c) pour maintenir cette norme dans le temps.

La formation à la standardisation liée aux niveaux du CECR comprend quatre étapes:

- effectuer les activités de familiarisation décrites dans le chapitre 3;
- travailler avec des performances et des tâches d'examens représentatives afin d'atteindre une compréhension adéquate des niveaux du CECR;
- transmettre une compétence à relier les tâches d'examens locaux et des performances à ces niveaux;
- s'assurer que cette compréhension est partagée par l'ensemble du groupe et se déroule de façon cohérente.

Avant de commencer la formation, le facilitateur/coordonateur désigné (appelé désormais coordinateur) doit lire attentivement ce Manuel et prendre en compte les ouvrages de référence recommandés et considérés comme étant pertinents dans ce contexte.

L'ordre dans lequel sont présentées les étapes du processus de standardisation n'est pas aléatoire. La formation avec des échantillons de performances orales et écrites – qui sont évaluées directement – est plus aisée pour les participants que la formation avec les items de réception orale et écrite. La réception écrite est l'aptitude la plus difficile à évaluer et devrait donc être traitée à la fin. Plusieurs études de cas lors de l'expérimentation du Manuel montrent un niveau d'accord entre les experts et un éventail de résultats plus réduit avec des échantillons de production qu'avec des items de réception. Nous considérons que cet ordre est le plus efficace et le recommandons, mais il est bien entendu possible de le modifier selon les besoins et les contraintes de la situation.

Une fois la formation terminée et un consensus adéquat obtenu sur l'évaluation des échantillons illustratifs (avec une fourchette s'étalant au maximum sur un niveau et demi A2+ à B1+), le travail de calibrage (échantillons de production) ou la définition de points de césure (pour des examens indirects avec des résultats à base de notes) peut commencer avec des performances d'apprenants locaux.

5.3. Planification préalable

Les coordinateurs sont responsables:

- de la logique à suivre, basée sur ce Manuel et sur les références appropriées;
- des décisions quant aux types d'expertises auxquels il faut faire appel, quant aux personnes à impliquer et à leurs rôles, ainsi que l'étape du processus à laquelle elles interviendront;
- des décisions quant au nombre et à la composition du groupe d'évaluateurs. Un groupe de 12 à 15 personnes est un minimum. L'expérience tirée de l'expérimentation du Manuel et d'autres projets de définition de points de césure montre qu'il est intéressant de faire appel à des évaluateurs externes à l'institution, ainsi qu'à des experts/parties prenantes représentant des points de vue différents;
- de la mobilisation d'experts locaux habitués à:
 - travailler avec le CECR;
 - produire des programmes et des spécifications d'examens;
 - évaluer des capacités langagières de production en fonction de critères définis;
 - concevoir des examens de langue et rédiger des items;
 - coordonner et former des groupes d'enseignants et d'examineurs;
- de la collecte de copies d'échantillons représentatifs du CECR et de la documentation appropriée;
- des instructions qu'ils donneront pour recueillir, dans un format défini localement, le matériel qui sera utilisé:
 - les échantillons locaux d'écrits et les vidéos de performances orales d'étudiants qui seront utilisées pour calibrer les performances locales sur des échantillons standards du CECR et sur le CECR lui-même;
 - les tâches d'examens locaux qui serviront de documents de travail dans les stages sur l'évaluation;

- de la décision d'utiliser ou non les niveaux plus du CECR. Des descripteurs calibrés pour les niveaux A2+, B1+ et B2+ sont disponibles;
- de la préparation, de l'élaboration et de la reproduction du matériel qui sera utilisé aux différentes étapes de la démarche (voir tableau 5.5 pour les détails):
 - les descripteurs de niveau du CECR;
 - les tableaux du CECR et les outils d'évaluation (par exemple le tableau 3 du CECR – tableau C2 du Manuel);
 - une sélection d'échantillons de performances et de tâches représentatives du CECR;
 - une sélection d'échantillons de performances et/ou d'items d'examens locaux;
 - les fiches de compte rendu et les documents utilisés pour recueillir l'information sur les stages;
- de la vérification du nombre de salles disponibles pour les travaux de groupes, ainsi que des moyens – tables et matériel audio – pour pouvoir travailler sur des échantillons d'écrits ou des items de réception orale;
- du recueil et de l'analyse des données venant des stages de formation à la standardisation, de la présentation et de la reproduction de résultats significatifs (par exemple la difficulté empirique de la valeur des items; les évaluations d'échantillons par d'autres groupes) afin de les réutiliser éventuellement dans d'autres stages au moment approprié;
- de l'organisation même des stages de la façon la plus adaptée à la situation locale. Le coordinateur devra décider du nombre de participants par stage, ainsi que des dates et du programme les plus appropriés. Cela comprend:
 - une décision concernant le statut des participants (enseignants/examineurs/rédacteurs d'items), les stages auxquels ils participeront et les implications sur la préparation des stages selon le public concerné;
 - la nécessité de s'assurer d'une bonne ambiance et du regroupement d'experts adéquat;
 - la planification appropriée du temps (voir ci-dessous) afin de donner l'occasion d'une réflexion et d'une discussion vaste et approfondie qui contribuera à l'obtention d'un consensus sur les évaluations;
 - le résumé des conclusions.
- de l'organisation de la documentation et du compte rendu du travail effectué durant les stages de formation afin d'assurer la crédibilité du système et de fournir un support servant à la diffusion de stages et à des sessions ultérieures.

- de la planification du contrôle continu, de la diffusion et des actions de suivi.

5.4. Animation des stages

La formation devrait se dérouler au cours de stages pendant lesquels les participants se familiarisent avec le CECR, analysent et évaluent des performances ou des items d'examens et parviennent à un consensus sur le placement à un niveau du CECR.

Pendant les stages, les coordinateurs désignés doivent:

- s'assurer que les participants arrivent à une bonne compréhension de ce qu'est le CECR et vérifier jusqu'à quel point ils prennent conscience de la contribution du CECR à l'amélioration de leur travail. On utilisera dans ce but les activités de familiarisation du chapitre 3;
- s'assurer, lors de l'évaluation d'échantillons de performances, qu'une progression logique est suivie afin de parvenir à un consensus et de le renforcer:
- recueillir des informations et donner régulièrement un retour d'information de manière aussi claire et visuelle que possible;
- vérifier, comme cela est précisé dans les instructions, qu'un consensus satisfaisant sur l'interprétation des niveaux du CECR est atteint, d'une part en ce qui concerne les descripteurs et d'autre part en ce qui concerne les performances ou les tâches qui les rendent opérationnels.

A l'issue de la formation, les coordinateurs ont la responsabilité de s'assurer que les participants ont à leur disposition tout le matériel nécessaire avant le début du processus de calibrage/définition de points de césure.

5.4.1. Arriver à un consensus et le vérifier

Tout au long du stage, on recommande aux coordinateurs de susciter les commentaires et les discussions et de faire une synthèse des évaluations en tenant compte du contexte afin de parvenir à un véritable consensus.

Comme dans tout stage de formation d'évaluateurs, on demande aux stagiaires d'évaluer le niveau correct d'un échantillon standard. Ce niveau est connu des coordinateurs, mais n'est pas donné aux stagiaires avant leur évaluation. Il ne sera donné qu'à la fin du stage, par les coordinateurs. Contrairement aux activités de calibrage et de définition de points de césure qui suivent, dans cette étape, le groupe n'est pas invité à trouver un consensus sur le niveau sans tenir compte de preuve antérieure, mais doit plutôt arriver à la réponse correcte déjà trouvée en appliquant les critères.

Cela suppose un certain savoir-faire de la part des coordinateurs qui doivent (a) conduire le groupe vers la réponse correcte au cours de ces expériences initiales importantes et (b) éviter de mettre en cause les participants trop sévères ou trop indulgents dans leur interprétation avant qu'ils n'aient eu le temps de s'investir dans la formation – car cela pourrait les perturber et déstabiliser leurs jugements ultérieurs. Il ne faut pas sous-estimer le temps que demande cette évolution. Il est essentiel de prendre tout le temps nécessaire à la formation avant de passer au travail sur les échantillons locaux non encore standardisés.

Quel que soit le type d'approche choisie, les coordinateurs devront calculer le pourcentage de participants qui s'accordent sur les différentes évaluations ou les coefficients de corrélation entre évaluateurs. Les coordinateurs devront décider de l'opportunité de communiquer ces chiffres aux participants s'ils considèrent que cela contribue à la formation et à une meilleure convergence des évaluations.

Il est également intéressant de présenter un schéma de dispersion des évaluations. Des graphiques sont facilement produits avec le vote électronique.

5.5. Formation avec des performances orales et écrites

Il se peut que des échantillons de performances et/ou de tâches d'examens représentatifs ne soient pas encore disponibles dans la langue concernée. Dans ce cas, nous recommandons de travailler avec les échantillons de la langue que le groupe a en commun – à condition que les groupes aient un niveau minimum B2/C1 de compétence dans cette langue. Dans ce cas, il faut indiquer dans la documentation qu'il s'agit d'une formation indirecte.

La première étape de la démarche est l'analyse et l'évaluation de performances orales représentatives du CECR. Elle est suivie (si cela convient) par des performances écrites représentatives. La plupart des échantillons oraux ont un format identique qui comprend, pour chaque candidat, une phase de production orale (un monologue suivi au cours duquel un candidat explique quelque chose à un autre candidat qui lui pose des questions) suivie d'une phase d'interaction (au cours de laquelle les deux candidats discutent d'un sujet de façon spontanée).

Pour l'évaluation de la performance écrite, il est important d'examiner des échantillons à la fois d'interaction écrite (par exemple des notes, des lettres) et de production écrite (par exemple des descriptions, des histoires, des critiques) d'un candidat. Cela est plus particulièrement important pour les niveaux élémentaires.

Il est important de noter que dans les échantillons représentatifs, c'est la compétence du candidat dans son ensemble, à partir de la performance dans sa totalité, que l'on évalue et non pas les performances séparées (monologue/interaction). Dans la documentation, on trouve des raisons argumentées justifiant tel ou tel niveau d'un candidat, avec des références explicites aux critères du CECR (CECR, tableau 3/tableau C2 pour la

performance orale; tableau B4 pour la performance écrite). Cela signifie que les tâches d'évaluation ont pour but de générer des échantillons représentatifs et complémentaires de la capacité du candidat à s'exprimer oralement dans la langue. Sur la base de *toutes* les preuves à disposition, l'expert utilise les descripteurs génériques basés sur les critères (CECR, tableau 3/tableau C2) pour juger de la compétence du candidat dans la mesure où elle peut être déduite d'un échantillonnage immanquablement limité et incomplet. Le résultat – la compétence apparaissant à travers la performance – est ce qu'on appelle habituellement en français la compétence.

5.5.1. La performance orale

Il est essentiel, pour ce stage, que les participants utilisent une grille d'évaluation comportant les descripteurs du CECR.

Ce stage est organisé en trois étapes:

- **Phase 1: Illustration.** Les coordinateurs commencent le stage par deux ou trois performances orales représentatives pour exemplifier les niveaux. Ils projettent l'échantillon et invitent les participants à commenter la performance avec leurs voisins. Au moment opportun, les coordinateurs reconstituent le grand groupe et lui font expliciter pourquoi cette performance illustre le niveau décrit sur la grille du tableau 3 du CECR (tableau C2) et non un niveau inférieur ou supérieur.
- **Phase 2: Pratique.** Dans cette seconde phase, le rôle des coordinateurs est d'aider les stagiaires à voir s'ils ont encore tendance à être trop sévères ou trop indulgents. Si le vote s'est fait par bulletin, les coordinateurs utiliseront une fiche de synthèse sur transparent (par exemple la fiche B3) ou un graphique pour enregistrer les évaluations. Tout au long de cette étape, les coordinateurs doivent faire visualiser aux participants leur comportement en tant que groupe et animer la discussion comme indiqué plus haut, sans embarrasser les individus. Si l'on n'a pas utilisé le vote anonyme, une technique efficace consiste à écouter les discussions des groupes et, lorsque tout le monde est regroupé, à faire donner «la réponse» par les groupes avec lesquels on a la meilleure chance qu'elle soit correcte.
- **Phase 3: Evaluation individuelle.** Les stagiaires évaluent individuellement le reste des performances, rendent leurs bulletins d'évaluation et discutent ensuite sur ce que représentent les niveaux du CECR auxquels ces performances ont été affectées. On recommande vivement de continuer à analyser les performances par blocs de trois. De la sorte, on focalisera mieux la discussion sur la standardisation – plutôt que d'entrer dans une discussion sur les mérites de certaines performances. Le dernier bloc de trois devrait faire l'objet d'un accord presque général.

Le stage se terminera lorsqu'on aura atteint ce niveau d'accord dans le groupe et que les coordinateurs (et les participants) seront satisfaits du degré de consensus atteint pour l'évaluation d'échantillons standards de performances orales.

5.5.2 La performance écrite

On recommande une démarche semblable à celle qui a été préconisée pour la performance orale.

- **Phase 1: Illustration.** Le stage commence par deux ou trois performances écrites standards que les coordinateurs utilisent pour exemplifier les niveaux. Pour chaque échantillon, à un moment donné, les coordinateurs reconstituent le grand groupe et lui font expliciter comment cette performance illustre le niveau décrit sur la grille du tableau C4 et pourquoi elle n'est pas du niveau inférieur ou supérieur.
- **Phase 2: Pratique.** Dans cette seconde phase – où l'on utilisera de nouveau trois échantillons – le rôle des coordinateurs est d'aider les stagiaires à voir s'ils ont encore tendance à être trop sévères ou trop indulgents. Si le vote s'est fait par bulletin, les coordinateurs utiliseront une fiche de synthèse (par exemple la fiche C3) pour rapporter les évaluations sur transparent. Tout au long de cette phase, les coordinateurs doivent faire visualiser aux participants leur comportement en tant que groupe et animer la discussion comme indiqué plus haut, sans embarrasser les personnes. Si l'on n'a pas utilisé le vote anonyme, une technique efficace consiste à écouter les discussions des groupes et, lorsque tout le monde est regroupé, à faire donner «la réponse» par les groupes avec lesquels on a la meilleure chance qu'elle soit correcte.
- **Phase 3: Evaluation individuelle.** Les stagiaires évaluent individuellement le reste des performances et discutent ensuite des niveaux du CECR auxquels ces performances ont été affectées.

On recommande très vivement de continuer à analyser les performances par blocs de trois. De la sorte, on focalisera mieux la discussion sur la standardisation – plutôt que d'entrer dans une discussion sur les mérites de certaines performances. Le dernier bloc de trois devrait faire l'objet d'un accord. Cela signifie que la grande majorité des participants devraient être d'accord sur le niveau avec une dispersion inférieure à un niveau et demi.

Le stage se terminera lorsqu'on aura atteint ce niveau d'accord dans le groupe.

5.6. Formation à l'aide des tâches et des items de capacités de réception écrite et orale et de compétences linguistiques

L'objectif des activités décrites dans cette partie est de s'assurer que les participants peuvent établir le lien entre leur interprétation des niveaux du CECR et les items d'examens représentatifs afin de pouvoir ultérieurement utiliser cette compréhension commune pour:

- relier les épreuves ou les items pertinents produits localement aux niveaux du CECR;
- acquérir, comme une plus-value, une compétence pour l'élaboration d'items d'examens pouvant éventuellement être considérés comme étant reliés aux niveaux du CECR.

Les techniques décrites peuvent être utilisées pour les items et les tâches d'examens évaluant des capacités de réception et peuvent être, le cas échéant, transférées à l'évaluation d'autres aspects de l'utilisation de la langue tels que la grammaire et le vocabulaire.

Les tâches qui impliquent des capacités intégrées (par exemple écouter un texte et répondre à des questions, puis utiliser l'information donnée pour faire un résumé) devront être considérées du double point de vue de la difficulté des aspects réceptifs et productifs de la tâche. Il y a généralement une différence délibérée de difficulté entre les deux parties de la tâche, et il faut traiter cette question au cours de la formation. La difficulté des items peut varier (et on peut la faire varier systématiquement si on le souhaite) en fonction du texte lu ou écouté, de l'aptitude à la compréhension que l'on teste et de la réponse que le candidat doit donner pour manifester sa compréhension.

Comme pour les échantillons de performance, une formation avec des tâches et des items représentatifs affectés de valeurs de difficulté connues doit d'abord être mise en place et être ensuite suivie du processus d'analyse d'items produits localement (chapitre 6).

La formation avec des items calibrés prépare, dans l'ordre suivant, à:

1. se rendre pleinement compte de l'étendue des sous-échelles de descripteurs du CECR pour des domaines particuliers disponibles dans le CECR (voir chapitre 4);
2. identifier la pertinence du contenu des items analysés en fonction de ce que recouvre le construit par rapport aux niveaux et aux échelles du CECR;
3. estimer le niveau de chaque tâche et item en fonction des descripteurs pertinents du CECR;
4. examiner les raisons possibles de divergences entre les niveaux estimés et les niveaux établis empiriquement;

5. confirmer le niveau de difficulté en les confrontant aux données empiriques.

Il est essentiel de commencer la formation avec la réception écrite. De même qu'il est plus facile de travailler sur des performances orale et écrite (que l'on peut observer en direct) que de travailler sur des compétences de réception (qu'on ne peut pas observer), il est de loin beaucoup plus facile d'organiser un travail de groupe sur la lecture et la relecture de textes et d'items imprimés (que l'on peut voir) que d'écouter à de nombreuses reprises des items et des textes (que l'on ne peut pas observer).

Une fois le processus d'évaluation des items de compréhension de l'écrit achevé, il sera plus facile d'organiser le stage sur la capacité à la compréhension de l'oral et de travailler sur des textes de compréhension de l'oral, car les stagiaires auront déjà l'habitude de la tâche à accomplir.

Les coordinateurs doivent décider de l'organisation des stages et estimer leur durée, laquelle dépendra du contexte et de la formation antérieure des participants.

5.6.1 La familiarisation nécessaire

Même si les stagiaires ont déjà participé au stage de familiarisation décrit dans le chapitre 3, avant de commencer l'évaluation de la difficulté et la définition des points de césure, une activité consistant à trier les descripteurs de la capacité que l'on étudie doit être organisée.

Le CECR fournit des échelles globales générales (par exemple «Réception», «Compréhension générale de l'écrit», «Compréhension générale de l'oral»), mais aussi des échelles spécifiques qui décrivent les différentes activités langagières de réception (par exemple «Comprendre en tant qu'auditeur») et de stratégies («Reconnaître des indices et faire des déductions»).

5.6.2. Formation à la définition des points de césure

Le processus de standardisation se déroule en trois étapes suivant des procédures pour la formation semblables à celles utilisées avec les échantillons de performances standards:

- **Phase 1: Illustration:** première évaluation du niveau d'un texte et des tâches et des items qui y correspondent. Cette activité préliminaire aidera les participants à s'entendre sur les niveaux du CECR pour la compétence évaluée. Il est essentiel d'examiner à la fois le **niveau du texte d'origine** et la **difficulté de chaque item** qui l'accompagne. Un texte n'a pas un «niveau». C'est la compétence des candidats, telle qu'elle se manifeste dans leurs réponses aux questions, que l'on peut relier à un niveau du CECR. Ce qu'on peut dire au mieux d'un texte, c'est qu'il convient pour son utilisation dans un examen visant un niveau donné.

- **Phase 2: Pratique suivie:** une fois les étapes d'illustration et de discussion achevées et un accord obtenu sur la façon d'envisager le processus, on demandera aux participants d'évaluer individuellement différents textes accompagnés des tâches et des items qui leur correspondent, de les relier aux niveaux du CECR et d'identifier les descripteurs du CECR que chaque tâche ou item met en œuvre.
- **Phase 3: Evaluation individuelle:** les stagiaires continuent à travailler individuellement avec le reste des items, puis discutent des niveaux du CECR auxquels ils ont été calibrés.

Une fois que l'on a achevé la formation à la standardisation (sections 5.4 et 5.5) et que l'on considère que le consensus sur l'évaluation des échantillons standards est satisfaisant, l'étape de travail sur les performances locales peut commencer. La section qui suit (5.6.) fait un compte rendu pour chaque étape de la façon de calibrer des échantillons locaux de performances orale et écrite. Les procédures à suivre sont très semblables à celles de la formation (section 5.4).

Pour établir des seuils fonctionnels sur des examens conçus localement pour les réceptions écrite et orale ou pour des compétences sous-jacentes, le choix des procédures de définition des points de césure parmi celles qui sont décrites dans le chapitre 6 du Manuel (ou d'autres écrits sur la définition des scores de césure) aura une influence sur les procédures à suivre. On recommande aux utilisateurs du Manuel de lire le chapitre 6 et de choisir une méthode ou plus d'une et, en suivant le canevas de la formation décrite dans cette partie, d'élaborer leurs propres procédures, étape par étape, qui soit appropriée au contexte. La documentation disponible pourra être très utile pour rédiger les procédures, mais il est nécessaire de prendre en compte les points décrits dans la partie suivante pour le calibrage en rapport avec la sélection d'items, l'analyse de données et la documentation.

5.7. De la formation au calibrage

L'application de la compréhension des niveaux du CECR au calibrage des échantillons locaux (de performances orale ou écrite) ou de tâches/d'items locaux (pour les examens portant sur la réception orale et écrite et la compétence linguistique évalués avec des notes) doit avoir lieu dès que possible après la formation à la standardisation. On recommande très vivement qu'elle ait lieu au cours du même stage, l'après-midi même ou le jour suivant. Le coordinateur sera le mieux placé pour juger si cela est faisable ou s'il vaut mieux le faire plus tard.

Si le calibrage d'échantillons locaux se fait au cours d'un stage à part, on recommande, au cours d'une phase d'harmonisation, de montrer aux stagiaires des extraits d'une ou deux performances standards évaluées au cours de la session précédente et on leur rappelle la discussion qui a eu lieu.

5.7.1 Echantillons nécessaires

Même si cela retarde le projet, il est important d'investir du temps et de l'énergie pour recueillir un jeu d'échantillons locaux de bonne qualité.

Le processus de collecte peut être très semblable à celui du processus de production d'items:

- définition des critères de sélection;
- identification des échantillons de candidats;
- travail en atelier pour étudier et filtrer les échantillons en fonction de leur qualité;
- sélection;
- vérification de la représentativité du jeu d'échantillons sélectionnés;
- apport éventuel d'échantillons supplémentaires pour «compléter» l'ensemble;
- documentation des caractéristiques des échantillons pour le calibrage grâce à un outil tel que les grilles du CECR pour les tâches de productions écrite et orale (annexe B2).

Il est essentiel que les échantillons locaux de performances utilisés pour le calibrage comprennent, pour les mêmes candidats, des discours de types différents couvrant l'éventail des activités décrites dans le CECR.

5.7.2 Arriver à un consensus et le vérifier

En général, les procédures à suivre sont celles décrites dans les sections 5.3 et 5.4 pour la formation à la standardisation avec des échantillons représentatifs. Cela comprendra:

- l'utilisation des mêmes outils que ceux utilisés pour la formation;
- une évaluation individuelle suivie d'une discussion en petits groupes conduisant le grand groupe au consensus;
- une discussion sur la dispersion dans les évaluations individuelles renouvelée jusqu'à ce que l'on parvienne à un accord acceptable (dispersion égale à un niveau et demi).

Un point important est ici à souligner: les évaluations individuelles doivent être enregistrées **avant** toute discussion. En effet, l'expérience des séminaires de calibrage qui ont débouché sur l'édition de DVD représentatifs montre que c'est la dispersion des évaluations qui est affectée par les discussions (les marginaux se conformant à la norme) et non la moyenne et donc le résultat. Néanmoins, le signe du succès d'un séminaire de calibrage est que les évaluations d'individus combinés et le consensus

final arrivent aux mêmes niveaux du CECR pour un échantillon ou un item. La publication de données non biaisées fait partie des preuves qui peuvent être fournies.

Si l'on ne parvient pas à un accord, les coordinateurs doivent discuter avec les stagiaires de la raison de ce problème incompatible avec leur maîtrise de l'évaluation des échantillons standards. Les coordinateurs devront se prononcer sur la cause du problème et faire le nécessaire pour le résoudre.

5.7.3. Analyse des données

Les évaluations des échantillons standards du CECR devraient être analysées statistiquement afin de (a) confirmer la relation avec les niveaux et (b) calculer la fiabilité d'un même évaluateur (cohérence) et des évaluateurs entre eux (cohérence).

L'avantage principal est que les évaluateurs dont le comportement n'est pas cohérent peuvent être identifiés et qu'on peut exclure leurs évaluations de l'analyse.

5.7.4. Documentation

A la fin du stage, il est essentiel que le jeu d'échantillons calibrés soit archivé, accompagné des comptes rendus du stage. Lors d'un stage ultérieur de formation, il sera extrêmement utile de pouvoir donner une explication justifiant qu'un échantillon donné ait été classé à un certain niveau. A cet égard, la documentation qui accompagne les échantillons représentatifs des DVD peut servir de modèle.

L'enregistrement sonore des débats lors du stage peut être un document utile pour préparer des notes de ce type sur chaque échantillon calibré. Les coordinateurs peuvent aussi décider de demander à l'un des stagiaires de l'aider en prenant des notes sur la raison du classement de certains échantillons à des niveaux donnés. On peut alors standardiser ces notes et en faire un ensemble cohérent pour la documentation et les distribuer aux participants à l'issue du stage.

Les utilisateurs du Manuel peuvent se demander:

- comment s'assurer de la constitution d'un panel équilibré et représentatif pour le projet;
- quelle taille un panel peut et doit raisonnablement avoir;
- quelle stratégie est la plus appropriée au contexte (en termes de ressources, de planification, d'application, d'analyse);
- si le projet a pour but de calibrer des échantillons «locaux» pour une utilisation ultérieure comme échantillons représentatifs d'un contexte spécifique;

- comment s'assurer de la qualité d'un tel matériel «local» en vue du calibrage (et de formations ultérieures);
- sous quelle forme présenter la documentation sur le matériel local et comment la distribuer;
- quelle durée de formation est nécessaire;
- si tous les participants doivent être mis au même niveau au départ ou s'il est possible de donner à certains des tâches à accomplir avant le stage;
- s'ils vont utiliser les «niveaux plus» du CECR (il y a des arguments pour et contre; l'important est de ne pas modifier l'approche une fois que le processus est en cours);
- s'ils vont utiliser les grilles d'évaluation du CECR dans l'annexe C ou élaborer d'autres outils plus spécifiques du CECR;
- comment publier et diffuser les résultats du processus de standardisation dans le champ de l'évaluation;
- comment s'assurer d'une bonne diffusion locale et du suivi

Chapitre 6: Procédures de définition des scores de césure

6.1. Introduction

Le résultat élémentaire de la participation à un test est un score numérique. Dans le cadre de tests constitués d'une forte proportion d'items, en compréhension écrite et en compréhension orale par exemple, ce score correspond généralement au nombre de bonnes réponses. Dans le cadre des compétences productives, la performance de la tâche est principalement évaluée à partir d'un nombre défini de critères pour lesquels le candidat reçoit un nombre de points (par exemple de zéro à quatre ou cinq). Le cas échéant, le score au test est le nombre total de point acquis par le candidat sur l'ensemble des critères et l'ensemble des tâches qu'il ou elle a accompli. Sur la base de ce score, une décision est prise quant aux compétences du candidat, dont, la plus importante, celle relative à l'échec/réussite: la performance du candidat au test est-elle satisfaisante?

Si la certification est reliée au CECR, une autre décision doit alors être prise: savoir si le candidat a atteint ou non un niveau particulier du CECR (B2 par exemple). Ces décisions (échec/réussite et niveau du CECR) impliquent la détermination d'un *score de césure* qui définit une *performance normée*. Pour la décision échec/réussite, le score de césure est le score minimal au test qui conduit à la décision «réussite»; les scores inférieurs à ce score de césure conduisent eux à la décision «échec». De même, un score de césure pour le niveau B2 correspond au score minimal qui conduira à positionner la compétence du candidat au niveau B2 ou plus; les scores inférieurs sont alors interprétés comme infra-B2 (c'est-à-dire B1 ou moins que B1).

Certains tests nécessitent plusieurs points de césure. En reliant l'examen au CECR, on pourrait par exemple souhaiter disposer d'un score de césure pour A2, B1 et B2. Il est important de comprendre ce qui est précisément signifié par la précédente phrase. Un score de césure doit être conçu comme une frontière entre deux catégories adjacentes d'une seule et même échelle. Ainsi, dans l'exemple dont il est ici question, il faudra considérer que chaque candidat sera classé soit en A2, en B1 ou en B2 et que deux scores de césure sont alors nécessaires: l'un qui marque la frontière entre les niveaux A2 et B1 et l'autre pour la seconde frontière entre les niveaux B1 et B2. En général, le nombre de points de césure est égale au nombre de classification moins un.

Pour éviter toute confusion entre les catégories (niveaux) et les scores de césure (les limites entre ces niveaux), on dénomme souvent le point de césure par les deux catégories adjacentes qu'il sépare. Dans l'exemple du paragraphe précédent avec trois catégories, les points de césure pourront être indiqués comme A2/B1 et B1/B2. Il est primordial de rester vigilant à l'égard de la labellisation des deux catégories aux extrémités de l'échelle: la labellisation de la catégorie la plus faible, dans cet exemple en A2, pourrait impliquer que tout candidat dont le score est inférieur au score de césure A2/B1 est de niveau A2, incluant également les candidats ayant un score de

zéro. C'est pourquoi il est préférable de rendre la labellisation explicite, pour l'exemple ci-avant, il conviendrait de retenir «A2 ou inférieur à A2». De même, l'utilisation de «B2 ou supérieur à B2» serait plus judicieuse pour la catégorie supérieure de cet exemple.

La détermination du score de césure ou de la performance normée relève souvent d'une décision collégiale. Le groupe qui réalise une pareille décision est généralement appelé panel. Les approches qui impliquent la participation d'un panel prennent classiquement plusieurs jours. La plupart du temps est consacrée à des activités qui sont décrites dans les chapitres précédents. Pour relier les examens au CECR, les panélistes doivent être familiers du CECR (chapitre 3), ils doivent s'assurer que l'examen recouvre lui-même les spécifications du CECR (chapitre 4) et, enfin, ils doivent être entraînés à la façon d'appliquer les descripteurs du CECR à l'examen (chapitre 5). Dans le présent chapitre, l'attention sera portée sur des aspects plus formels du groupe de décision: le type de jugement établi par les panélistes, le type d'information disponible et la manière dont les jugements sont traités et compilés pour parvenir à un ou plusieurs scores de césure. De telles procédures ont souvent été formalisées et sont connues sous le nom de procédures de détermination des scores de césure.

La détermination des scores de césure peut avoir des conséquences importantes pour les individus et pour les décisionnaires politiques. Cette détermination exige un jugement prudent; autrement dit, «la définition des scores de césure est probablement le pan de la psychométrie qui associe plus d'aspects culturels, politiques et artistiques en un mélange de ses produits que n'importe quel autre» (Cizek, 2001: 5).

6.2. Aspects généraux

Une part essentielle des procédures de détermination des points de césure tient en l'organisation efficace des rencontres. Généralement, une partie, voire la totalité des phases de familiarisation, de spécification et de standardisation décrites dans les chapitres précédents de ce Manuel forment un ensemble cohérent avec les procédures de définition des scores de césure (au sens strict du terme) qui sont discutées dans ce chapitre. Ainsi, la procédure considérée dans son ensemble nécessite des ressources et exige une organisation efficace.

6.2.1. Organisation

Les procédures de définition des points de césure basées sur des panels durent généralement deux à trois jours et commencent avec une ou deux sessions de familiarisation, de discussion sur les spécifications du test et d'entraînement avec du matériel servant d'illustration et passent par une étape cruciale au cours de laquelle tous les experts du panel jugent le test constitué des items considérés. Après la remise d'instructions appropriées, les membres du jury donnent leur jugement, généralement

en deux ou trois phases (tours) séparées par des phases de discussions, puis de retour de d'information et de données supplémentaires.

Pendant les sessions entre les phases d'évaluation, deux types d'informations principales sont fournies. Après la première phase d'évaluation, une information indiquant le comportement des membres du jury est remise, montrant que certains d'entre eux offrent de véritables jugements déviants. Ce type d'information est appelé information *normative* et doit en principe permettre, en premier lieu, de détecter et d'éliminer les malentendus au sujet des instructions. C'est une bonne expérience que de permettre aux membres du jury de discuter de cette information en petit groupe. Après le deuxième tour, une information de nature différente nommée *impact* est souvent donnée. Cette information indique les conséquences des jugements des panélistes et repose sur le calcul de la proportion des candidats qui atteindraient ou non chaque catégorie selon les scores de césure provisoires déterminés par le résultat des tours précédents.

On ne devrait cependant pas oublier que la procédure conduisant à la définition des scores de césure en situation de fort enjeu est souvent ancrée dans un contexte social et politique et qu'il est alors prudent de confronter les panélistes aux conséquences sociales de leurs décisions. Après avoir informé les panélistes, il est possible qu'un certain nombre d'entre eux changent d'avis et deviennent plus stricts ou plus indulgents, comparativement à leurs jugements précédents, et ce pour des raisons opportunistes. Si cela se produit, cela n'implique pas nécessairement que ce changement d'opinion devienne la décision finale. Au contraire, une déviation importante dans les standards après mesure de l'impact pourrait être utilisée pour engager une discussion plus approfondie dans le but de trouver un consensus raisonnable et rationnel entre deux états décisifs très différents, ce qui pourrait justifier l'organisation d'un quatrième tour de jugement.

Pour la grande majorité des procédures d'établissement des points de césure décrites dans la littérature, de nombreuses variations ont été testées, adaptées à des besoins spécifiques ou inspirées par des carences d'expériences antérieures. Les applications illustrent ce qui tient essentiellement en la même procédure: l'organisation des échanges (en séance plénière ou en petits groupes), etc. Pour jauger la validité et l'efficacité d'une procédure appliquée à un projet donné, il est cependant crucial qu'une documentation détaillée et adéquate de l'ensemble des étapes de la procédure soit disponible. Sans cette description technique détaillée, l'évaluation professionnelle des résultats devient délicate et l'on ne peut plus prétendre avoir élaboré un argumentaire. Cela est d'autant plus important que de plus en plus d'étudiants ont besoin de preuves de leur compétence langagière dans une langue étrangère. Il faut qu'ils puissent produire des certificats de compétence valides. Il est enfin nécessaire que les procédures de définition de points de césure soient explicites.

6.2.2. Concepts

En insistant sur le fait que les scores de césure ne peuvent être correctement définis en se contentant de suivre mécaniquement une méthode donnée, ce chapitre proposera une discussion de quelques aspects fondamentaux qui sont soulevés par une variété de méthodes de détermination des scores de césure.

Les méthodes pour établir les points de césure sont parfois divisées en deux sous-ensembles; d'une part, celles centrées sur le test et, d'autre part, celles centrées sur le candidat. La caractéristique importante de ces méthodes centrées sur le candidat tient au fait que les candidats *spécifiques* sont reportés dans des catégories (échec ou réussite, niveau B1, B2 ou en cas limite) par un jugement holistique.

Dans les méthodes centrées sur le test, il est demandé aux panélistes d'effectuer un jugement sur chaque item. Ces jugements reposent sur les caractéristiques perçues des items par le panel d'experts. La procédure, dans son ensemble, peut être appliquée sans aucune donnée empirique de candidats. Pour ces méthodes, la mention «centrée sur le test» est tout à fait appropriée. Avec la popularité grandissante de la théorie de réponse à l'item (TRI), des méthodes ont été développées. Pour celles-ci, la distinction entre les méthodes centrées sur le test et celles centrées sur le candidat est moins claire. Dans ces méthodes, l'information disponible pour les panélistes est directement issue des performances d'un groupe de candidats. Généralement, cette information est formalisée par la mesure de difficulté de l'item. La disponibilité d'une telle information est censée aider le panel d'experts et le dispenser de la délicate tâche de fournir une estimation de la difficulté qui repose exclusivement sur les caractéristiques perçues d'un item.

Les méthodes discutées dans ce chapitre pourraient ainsi être catégorisées en trois groupes. Le premier serait relatif aux méthodes «centrées sur le candidat», le deuxième serait relatif aux méthodes «centrées sur le test» dans la mesure où elles peuvent être mises en œuvre sans (ou avec) donnée empirique et le troisième serait relatif aux méthodes de la «TRI» en ce sens où le panel d'experts utilise un résumé des données empiriques (classiquement fourni par l'analyse dans le cadre de la TRI).

La qualité de la définition des scores de césure est sujette à de grandes variations. Quelle que soit la méthode retenue ou la combinaison de plusieurs d'entre elles, nous ne pouvons pas considérer que les scores de césure ont été correctement définis uniquement parce que certaines procédures auraient été respectées. Il est nécessaire de rassembler des *preuves évidentes de la qualité des résultats* des procédures et d'en faire part de façon suffisamment détaillée et transparente. Cette question concernant la validité sera traitée plus longuement dans le dernier chapitre de ce Manuel.

6.3. La méthode du corpus de productions: centrée sur le candidat

La méthode du corpus de productions (Kingston *et al.*, 2001) est peut-être la plus appropriée pour les jugements holistiques, bien qu'elle puisse être utilisée avec toutes les combinaisons de types d'items et de tâches. Elle est centrée sur les candidats et n'utilise pas la TRI. Vous trouverez ci-après un brève liste de ce qui est nécessaire pour appliquer cette méthode:

- une collection de travail d'un échantillon de candidats. Ce travail peut consister seulement en des réponses à des questions à choix multiples ou en un mélange de questions à choix multiples, de questions ouvertes et de rédactions, voire en un portfolio. Une condition d'application nécessaire est que le travail (la performance au test, le portfolio) soit validé par un *score numérique*;
- l'échantillon n'a pas besoin d'être représentatif d'une population cible du test. Il doit néanmoins couvrir la plupart de l'étendue des scores possibles, indépendamment de la fréquence relative de ces scores avant la mise en place de la procédure pour déterminer les scores de césure;
- la tâche des experts du panel est de fournir un *jugement holistique* sur chaque échantillon de travail qui leur est présenté. Dans le cadre du CECR, un tel jugement consistera en l'attribution aux candidats de l'un des niveaux prédéfinis que l'on vise dans la procédure de définition des points de césure. Supposons que l'on veuille définir les points de césure pour A1/A2 et A2/B1, le jugement des membres experts du panel devra catégoriser chaque production de candidat soit en A1, A2 ou B1 (ou plus);
- le type de jugement requis de la part des panélistes est le même que celui demandé dans la méthode des groupes contrastés ou dans la méthode des cas limites. La différence essentielle avec ces deux méthodes tient au fait qu'ici, tous les panélistes évaluent la même collection d'échantillons de production, de telle sorte que les discussions en groupe entre les tours aient du sens. La méthode du corpus de productions nécessite deux tours, bien qu'il puisse être nécessaire d'en ajouter un troisième;
- les scores des échantillons de travail des candidats ne sont pas connus des membres experts du panel;
- pour convertir les jugements des panélistes en un score de césure, on doit avoir recours à une technique particulière, appelée la régression logistique. La raison est liée à la haute sélection de l'échantillon des travaux utilisés. En effet, l'application des méthodes usuelles (par exemple rechercher le point central entre les moyennes dans le cas de la méthode des groupes contrastés) pourrait conduire à de sévères biais.

6.4. La méthode de Tucker-Angoff: centrée sur le test

C'est une des méthodes les plus répandues pour déterminer les scores de césure. De nombreuses variations de cette méthode ont été proposées.

Un concept de base, qui apparaît également dans de nombreuses autres procédures de détermination des points de césure, est le concept du «candidat aux compétences minimales», également désigné parfois comme le «candidat limite», le «candidat à la frontière» ou encore le «candidat réussissant à peine». Là où un point de césure doit être utilisé, par exemple pour le CECR au niveau B1, le candidat aux capacités minimales est celui qui a les compétences pour être apparié au niveau B1, mais de telle sorte que la perte, si infime soit-elle, d'une partie de ses compétences suffirait à ne plus le catégoriser dans ce niveau de qualification. La tâche des panélistes est de conserver à l'esprit un tel profil de candidat ou d'un ensemble de candidats durant tout le travail de jugement qu'ils doivent effectuer.

Pour chaque item du test, le panel d'experts doit indiquer avec quelle probabilité un candidat aux compétences minimales répondrait correctement.

L'étape suivante de la procédure consiste en l'**addition** des probabilités sur l'ensemble des items et pour tous les juges. Le plus souvent, c'est le calcul de la moyenne des sommes qui est opéré et la moyenne est considérée comme le point de césure recherché. Voir le tableau ci-dessous.

	Juge 1	Juge 2	...	Juge 15	
Item 1	25%	32%	...	35%	
Item 2	48%	55%	...	45%	
Item 3	33%	38%	...	28%	
...	
Item 49	21%	30%	...	35%	
Item 50	72%	80%	...	90%	
Moyenne	65%	72%	...	78%	Standard: 75%

*Tableau 6.1: Données de base dans la méthode de Tucker-Angoff.
Pourcentage de réponses correctes pour un candidat limite*

En résumé, trois éléments sont essentiels: le concept de personne limite, la probabilité d'une réponse correcte pour cette personne (donnée pour chaque item par chaque participant) et le calcul de la somme de ces probabilités pour tous les membres du panel.

6.5. La méthode du panier: centrée sur le test

La méthode suppose la comparaison des ressources exigées par un item en termes de PLD, c'est-à-dire au sens des descripteurs «être capable de» («can do») du CECR. La question élémentaire posée aux panélistes n'est pas un jugement sur l'item, mais se centre sur un candidat abstrait ayant les compétences d'un niveau défini. La formulation élémentaire de la question est la suivante: *«A quel niveau du cadre un candidat peut-il déjà répondre correctement à cet item?»*.

Les experts du panel doivent mettre chaque item dans un panier correspondant aux niveaux du CECR. Si un item est placé dans le panier B1, cela signifie qu'une personne de ce niveau devrait donner une réponse correcte à cet item. On suppose ici que si c'est le cas, les personnes des niveaux supérieurs devraient également répondre correctement à l'item. Notez que ce jugement n'implique pas que les personnes de niveaux inférieurs ne devraient pas fournir une réponse correcte; cela signifie simplement (pour les membres du panel) qu'une réponse correcte ne devrait pas être exigée pour les candidats de niveaux inférieurs.

La méthode pour convertir les jugements en score de césure suppose qu'avec la méthode du panier, le panéliste propose les exigences minimales requises pour chacun des niveaux. Supposons que pour un test constitué de 50 items, deux items sont placés dans le panier A1, sept dans le panier A2, 12 dans le panier B1. Pour ce panéliste, ces 21 (= 2+7+12) items devraient être traités correctement par n'importe quel candidat de niveau B1 ou supérieur. Ce nombre, qui correspond à l'exigence minimale, est interprété comme le score de césure.

Nous proposons maintenant une courte note technique. Pour un des panélistes, un item pourrait être jugé comme étant si difficile qu'il ne pourrait pas exiger qu'un candidat du niveau supérieur le réussisse. Au regard de la procédure, cela signifie que l'item ne s'ajuste à aucun des paniers envisagés. On peut anticiper de telles situations en ajoutant un panier supplémentaire qui serait étiqueté «supérieur à C2». Bien entendu, si un test vise le niveau B1, il n'est pas nécessaire de disposer de paniers pour tous les niveaux. Les trois paniers de niveaux les plus forts pourraient être nommés B1, B2 et supérieur à B2.

Il est possible que l'ajustement de l'exigence minimale au point de césure conduise à des points de césure trop indulgents. Il serait alors raisonnable de penser qu'une personne d'un niveau donné soit également en mesure de répondre correctement à des items qui exigent un niveau supérieur. Ce point n'est pas pris en compte dans la méthode, mais des études comparatives (non encore publiées) indiquent que la méthode du panier tend à produire des points de césure minorés (indulgents) par rapport à ceux obtenus avec d'autres méthodes.

6.6. La méthode du marque-page: centrée sur le test

La méthode est centrée sur le test et est applicable aussi bien pour les items dichotomiques que les items polytomiques (questions à réponse ouverte).

Les membres experts du panel utilisent le concept du candidat aux compétences minimales acceptables ou du candidat à la limite de deux niveaux. La procédure doit être répétée autant de fois qu'il y a de points de césure à fixer (par exemple A1/A2, A2/B1 et B1/B2 d'un même test). La charge de travail est néanmoins plus légère que celle exigée par la méthode de Tucker-Angoff. La raison est expliquée au point suivant.

Les items ou les tâches sont présentés aux membres du panel par ordre de difficulté croissante. Les tâches à réponse ouverte apparaîtront plusieurs fois dans cette liste. Par exemple, si le score peut être 0, 1 ou 2 points, la tâche apparaîtra à deux reprises, une fois avec une réponse permettant d'obtenir un point et la seconde fois avec une réponse permettant d'obtenir deux points. Les items et les tâches sont disposés dans un livret. Chaque page contient un item (dans le cas d'items dichotomiques) ou une combinaison d'une tâche à notation partielle pour les questions à réponse ouverte.

Le concept de maîtrise d'une tâche ou d'un item. La maîtrise est ici définie en termes probabilistes. Si un candidat maîtrise un item, on peut s'attendre à ce qu'il réponde correctement avec une probabilité associée élevée. La définition exacte de cette probabilité associée élevée est arbitraire, mais, dans la plupart des cas, est fixée à 2/3, même si certains autres préfèrent la fixer à 50%, alors que d'autres la fixent à 80%. Dans la littérature sur les points de césure, le critère de maîtrise se réfère à la *probabilité de réponse*. Les membres du panel doivent décider si, pour un item donné, un candidat à la frontière des niveaux (pour un point de césure défini) maîtrise ou non cet item. Pour une probabilité de réponse fixée à 2/3, cela signifie qu'ils doivent décider si la personne répondra correctement dans au moins deux cas sur trois (si la probabilité de réponse est fixée à 80 %, il faudra considérer une réponse correcte dans 4 cas sur 5). Il est important de s'assurer que les membres du panel aient bien intégré cette notion de probabilité de réponse et une attention particulière doit être allouée à cette compréhension au cours de la phase d'entraînement.

Il est demandé aux membres experts du panel de commencer avec le point de césure le plus bas (par exemple A1/A2), de progresser dans le livret en allant du plus facile vers le plus difficile et de décider pour chaque item si la probabilité d'une réponse a atteint le seuil fixé ou si elle est supérieure. Lorsque la réponse est affirmative, cela signifie que le candidat limite maîtrise l'item, selon le point de vue du panéliste. Parce que les jugements s'opèrent en premier lieu sur les items les plus faciles, on s'attend à ce que les réponses soient affirmatives pour quelques items à la suite, mais qu'à partir d'un item donné, la réponse devienne négative. Supposons que ce soit le cas à l'item 11, alors un marque-page (ou un symbole similaire) doit être placé à cet endroit. Immédiatement, le membre du panel doit changer de point de césure en traitant le suivant (par exemple A2/B1 ici) et continuer son travail de jugement à partir de l'item où il se trouve.

S'il y a trois points de césure à définir, le travail est en principe finalisé quand les trois marque-pages sont placés dans le livret. Cette opération devrait être réalisée bien avant le dernier item. Il est cependant d'usage d'inviter les panélistes à examiner l'ensemble des items et même à considérer la possibilité de déplacer les marque-pages précédents au fur et à mesure qu'ils progressent.

A chaque tour, chaque membre du panel indique son point de césure provisoire dans un tableau comme celui présenté dans la figure 6.2 pour une situation correspondant à trois points de césure. Il est préférable de laisser aux participants la possibilité d'indiquer deux numéros de page, comme dans la figure 6.2. Les pages 11/12 pour le point de césure A1/A2 signifient (pour le participant) qu'un candidat aux compétences minimales acceptables au niveau A1/A2 a au moins une probabilité égale à la probabilité de réponse de répondre à l'item 11 correctement, mais pas pour l'item 12.

Les informations collectées, à l'issue du premier tour, par l'équipe organisatrice de la procédure de définition des points de césure vont être utilisées pour le tour suivant et la décision finale.

Tour 1			
Points de césure	A1/A2	A2/B1	B1/B2
Numéro de pages	11/12	24/25	38/39

Figure 6.2: Formulaire d'enregistrement des jugements des panélistes dans le cadre de la méthode du marque-page

6.7. Définition des scores de césure sur plusieurs capacités langagières

Dans certains cas, il peut être exigé de définir un seul point de césure, le résultat global comme le niveau du CECR d'un candidat, alors que le test lui-même peut comprendre trois, voire plus de trois parties, où chacune permet de tester les performances dans des compétences différentes.

Il y a alors plusieurs façons de voir les choses. Deux points de vue seront ici discutés, une approche compensatoire et une approche conjonctive. Lorsque ces deux approches sont appliquées de façon stricte, elles peuvent conduire à des résultats inacceptables; une solution raisonnable est donc également discutée.

Approche compensatoire

D'une part, selon une position extrême, on peut considérer toutes les tâches et tous les items comme un mélange des capacités et alors appliquer l'une des méthodes abordées précédemment sur l'ensemble des items et des tâches simultanément. En procédant de la sorte, on doit concevoir que les scores au test sont, par définition, compensatoires, puisqu'ils sont les sommes des scores aux items et aux tâches. Echouer à certaines tâches peut être compensé par une bonne performance à d'autres tâches. Dans la mesure où le test est homogène du point de vue de la nature des tâches, un tel mécanisme compensatoire est légitime et les items et les tâches auxquels le candidat a échoué ou réussi sont sans importance.

Néanmoins, avec un test plus hétérogène, ce point de vue compensatoire pourrait être inadéquat. Par exemple, supposons qu'un examen national pour l'anglais, composé d'une épreuve de réception écrite, d'une épreuve de réception orale, d'une épreuve de production orale et d'une épreuve de production écrite, ait un score maximal de 100 points sur l'ensemble des quatre parties. En outre, supposons que la méthode du corpus de productions soit appliquée pour fixer les scores de césure et qu'on ait pris le soin de collecter des échantillons de production en provenance de différentes régions dans le pays. Si les régions diffèrent significativement dans leurs méthodes et du point de vue de leur expertise pour une ou plusieurs compétences, des profils typiques de compétences devraient révéler les différents patterns de régions. Si, dans certaines régions, une faible attention est allouée à l'expression orale, même les meilleurs étudiants de cette région pourraient être qualifiés comme faibles dans cette compétence et réussir au même niveau que l'étudiant moyen des régions où une attention plus importante aura été allouée à cette compétence. Prendre en compte l'ensemble des compétences pourrait masquer des différences importantes dans les profils.

Par conséquent, il est important qu'une étude minutieuse soit entreprise pour examiner dans quelle mesure une approche unidimensionnelle est appropriée. En plus de l'étude de la structure des différentes compétences, des différences structurelles possibles entre écoles, régions ou méthodes utilisées et qui pourraient révéler des items à fonctionnement différentiel (DIF) devraient être examinées avant que l'approche unidimensionnelle puisse être justifiée. Si des différences marquées ou des corrélations moyennes entre compétence étaient avérées, on devrait être confronté à plusieurs problèmes. Deux d'entre eux sont discutés ci-après:

1. Une décision rationnelle doit être prise sur la pondération qui sera attribuée à chaque compétence pour le score total. S'il y a une disposition légale qui stipule que chacune des compétences doit avoir le même poids, le problème est alors résolu.
2. Toutefois, même avec une pondération imposée, nous ne disposons d'aucune garantie, pour les méthodes centrées sur les candidats, comme avec la méthode du corpus de productions par exemple, que les panélistes utiliseront

effectivement cette pondération définie a priori pour fournir un jugement holistique du niveau de l'étudiant.

Approche conjonctive

L'alternative est une approche qui prend en compte séparément chacune des compétences et qui implique que chaque point de césure soit défini indépendamment pour chaque compétence. La règle de décision conjonctive stipule que l'on a globalement atteint un niveau donné si l'on a atteint ce niveau pour chacune des compétences. L'application stricte de cette règle pourrait conduire à des résultats inacceptables. Par exemple, un étudiant pourrait se voir refuser le niveau B1, même s'il a atteint le niveau B2 dans trois des quatre compétences et pas le point de césure A2/B1 dans la quatrième.

Dans ce cadre, un compromis entre les règles de compensation et les règles conjonctives semble raisonnable. Une règle conjonctive générale pourrait être fixée à laquelle pourraient être adjointes des exceptions compensatrices, comme pour l'exemple ci-avant où il apparaîtrait raisonnable d'attribuer le niveau B1 à cet étudiant. La nature exacte des exceptions compensatrices doit être considérée avec une grande vigilance. Une bonne façon de les appréhender serait d'en discuter avec les membres experts du panel après qu'ils aient statué sur les points de césure pour chacune des compétences séparément.

6.8. Définition des scores de césure et ajustement de tests

Etant donné que la procédure pour déterminer les scores de césure est un dispositif onéreux, il serait plus prudent de prendre la peine de chercher comment éviter une somme importante de travail, en particulier pour les examens cycliques où les spécifications des tests se répètent généralement d'année en année sans modifications majeures.

Si une procédure de définition des points de césure a été effectuée selon les règles de l'art pour un examen d'une année, les résultats de la détermination du score de césure pourraient être transférés tels quels à une même forme d'examen (par exemple de l'année suivante) en appliquant la technique dite de l'ajustement de tests. L'ajustement de tests désigne un ensemble de techniques permettant d'avoir pour chaque score dans un test un score équivalent dans un autre test. Supposons que le point de césure A2/B1 ait été fixé pour la première année de l'examen à 35 points. Si le score équivalent à 35 est 37 pour l'examen de la deuxième année, cela implique nécessairement que le score de césure soit alors de 37 points.

La mise en œuvre des ajustements de tests présente deux aspects auxquels il faut accorder une attention toute particulière. Le premier est quasi exclusivement de nature technique, le second est conceptuel.

Pour appliquer les techniques d'ajustement de tests, il est essentiel que les deux échantillons de candidats ayant pris part aux examens soient comparables. Une telle comparabilité peut être garantie soit par l'utilisation d'items en commun dans les deux examens ou en prenant des dispositions de telle sorte que les deux échantillons soit statistiquement équivalents. Aucune approche ne peut être mise en œuvre aisément dans un contexte d'examen: généralement, il n'est pas possible de répéter l'examen de l'année précédente lors de l'année en cours pour des raisons de confidentialité et l'équivalence des échantillons n'est pas simple à obtenir, étant donné que les étudiants ne peuvent être assignés à un examen de façon aléatoire. Une population légèrement plus compétente qu'une autre (que la précédente ou que la suivante) donnera le sentiment que l'examen était plus facile qu'il ne l'est en réalité. Si cela n'est pas clairement identifié et si les populations sont considérées comme identiques au regard de leurs compétences, cela conduirait à des points de césure absolus.

Un autre problème concerne la validité du concept hypothétique. Même si on peut admettre que des spécifications identiques renvoient à des concepts équivalents, on peut considérer que cela n'est pas suffisant. En effet les gens n'interprètent pas nécessairement la signification d'un concept de façon identique.

Toutefois, la façon la plus prudente de garantir la validité du transfert des points de césure par ajustement est d'accomplir une définition des points de césure sur le nouvel examen, pour vérifier si les points de césure obtenus par application de l'équation d'ajustement correspondent effectivement aux points de césure fixés par un panel indépendant de juges experts.

6.9. Définition des scores de césure sur plusieurs langues

L'aspect probablement le plus stimulant dans le fait de relier les examens au CECR est de trouver des méthodes qui montrent que les examens dans différentes langues sont liés d'une manière comparable à des standards communs.

Bien qu'il soit théoriquement possible d'administrer deux examens dans des langues différentes à un même échantillon de candidats, cela supposerait que chaque candidat de l'échantillon a le même niveau de compétences dans chacune des langues, ce qui est clairement impossible. Par conséquent, les méthodes qui doivent être recherchées doivent considérer que chaque candidat n'a participé qu'à un seul des deux examens et que les performances de chaque candidat dans les différentes langues seront traitées comme celles de candidats sans lien particulier.

Pour relier ces deux examens aux CECR, on peut faire appel à des panélistes plurilingues, qui pourront offrir des jugements sérieux et dignes de confiance à la fois sur les items (pour les méthodes centrées sur les tests) et sur le travail des candidats dans les deux langues. La méthode du corpus de productions pourrait être l'une des méthodes à privilégier pour ce dernier cas ou une méthode centrée sur le test comme la méthode Tucker-Angoff.

6.10. Conclusion

Ce chapitre a offert une revue d'un nombre de procédures de définition des scores de césure, mais ne prétend pas l'avoir fait de façon exhaustive. Une revue accessible peut être consultée dans la section B du supplément au Manuel. Dans ce chapitre, l'accent a porté sur la faisabilité et l'adéquation des méthodes sélectionnées aux tests langagiers et pour relier les examens au CECR en soulignant l'importance d'une bonne compréhension des notions de base.

Bien entendu, au cours et après la mise en œuvre de ces procédures, il sera nécessaire d'en suivre la qualité en se centrant sur plusieurs questions:

- Est-ce que la procédure de détermination des points de césure a eu les effets attendus? La formation a-t-elle été efficace? Est-ce que les panélistes se sont sentis libres de suivre leurs propres intuitions? Des questions similaires sont ici bienvenues. Ce sont les questions liées à la validité procédurale.
- Est-ce que les évaluations des experts du panel sont fiables? Est-ce que chaque membre du panel a été régulier au cours des différentes tâches qu'il a réalisées? Est-ce que les membres du panel ont été en accord avec les autres dans leurs jugements et dans quelle mesure un consensus a permis de considérer le point de césure comme définitif? Est-ce que des erreurs ont été commises dans les scores

au test? Ces questions, et leurs réponses, constituent la validité interne de la procédure de définition des scores de césure.

- La question la plus importante est de savoir si les résultats de la procédure conduisant aux points de césure – qui attribuent un niveau du CECR aux étudiants sur la base de leur score au test – sont dignes de confiance. La réponse à cette question vient de la preuve indépendante qui corrobore les résultats d’une procédure particulière de détermination des scores de césure. C’est la tâche de tout un chacun que d’appliquer une telle procédure pour fournir une réponse à cette question; c’est précisément ce qui est signifié par le terme «validation». Une telle preuve peut provenir de différentes sources, comme:
 - **la validation croisée:** la répétition des procédures de détermination des scores de césure avec des groupes indépendants de panélistes;
 - **la détermination complémentaire** des scores de césure: mettre en place des méthodes de détermination des points de césure indépendantes en utilisant une procédure différente et appropriée au contexte;
 - **la validation externe:** en conduisant une étude indépendante pour vérifier les résultats de la procédure de détermination des scores de césure, en les rapprochant d’un critère externe. Ce critère externe peut être un test pour la/les même(s) compétence(s), connu pour être fidèlement calibré au CECR. Cela pourrait également être les jugements d’enseignants ou d’apprenants formés aux descripteurs du Cadre.

Ces questions sont abordées dans le prochain chapitre.

Les utilisateurs du Manuel devraient considérer:

- la nécessité de lectures supplémentaires sur les procédures de détermination des scores de césure;
- quelle(s) méthode(s) est/sont la/les mieux adaptée(s) au contexte;
- s’il faut adopter une méthode centrée sur le candidat ou sur l’examen;
- s’il faut opter pour une méthode évaluant la difficulté des items (par exemple, l’appariement au descripteur ou la méthode du panier) ou pour une méthode évaluant le score de césure sur l’échelle du pré-test (par exemple, méthode du marque-page, méthode du corpus de productions);
- si deux méthodes devraient être utilisées pour la validation de leurs résultats respectifs;
- comment les panélistes proposeront leurs évaluations sur les points de césure après le premier tour; est-ce que le vote électronique est réalisable?

- si les paramètres de difficulté de la TRI seront disponibles pour renseigner le procédé permettant la précision des points de césure ou si les valeurs de probabilité devront être utilisées;
- quels types de données d'impact sur les effets provisoires du point de césure devraient être disponibles pour enrichir les derniers tours de discussion;
- quel(s) type(s) de moyens devrai(en)t être nécessaire(s) pour appliquer la/les méthode(s) retenue(s).

Chapitre 7: Validation

7.1. Introduction

Relier un examen au CECR est un processus complexe qui implique plusieurs étapes, qui toutes exigent du professionnalisme. La validation a trait au corpus de preuves proposé pour convaincre les utilisateurs du test que le processus, dans sa globalité, et ses résultats sont dignes de confiance. Les utilisateurs du test doivent ici être compris dans un sens très large; ils comprennent les élèves (ou leurs représentants légaux, comme les parents) qui passent le test, les autorités éducatives et politiques qui utilisent les résultats du test pour prendre des décisions politiques, les éditeurs de manuels et les enseignants, les organismes de test, les employeurs et les formations syndicales, la communauté scientifique impliquée dans les tests de langue et, si les enjeux sont véritablement forts, également les autorités légales. Bien que le présent Manuel se concentre sur le procédé pour relier les examens au Cadre, dans un sens plutôt strict, en mettant l'accent sur l'application d'une ou plusieurs procédures de détermination des scores de césure, il serait erroné de considérer que le processus de validation peut être totalement restreint aux activités et résultats décrits au cours des chapitres 3 à 6. Dans ce présent chapitre, la plupart des procédures et techniques discutées se concentrent sur l'adaptabilité du procédé qui permet de relier les examens au Cadre.

La validité n'obéit pas à une loi de type tout ou rien, mais s'établit plutôt sur un continuum. Pour un rapport sur la validité, il faudra porter l'attention sur les nombreuses facettes impliquées, en mettant en avant de solides arguments et des preuves empiriques pour faire face aux critiques relatives à la généralisabilité. Il est ainsi indispensable, pour une bonne étude de la validation, de disposer d'une documentation conséquente sur l'ensemble des activités entreprises.

7.2. Pré-requis: la qualité de l'examen

Relier au Cadre un examen qualitativement pauvre est une entreprise vouée à l'échec et qui ne peut être sauvée même par une détermination attentive des scores de césure. Dans cette section, un nombre important d'aspects de l'examen lui-même seront discutés brièvement, en gardant un seul objectif en tête, celui de relier correctement l'examen au Cadre. Ces aspects se réfèrent au contenu de l'examen, à ses aspects opérationnels et psychométriques.

7.2.1. Validité de contenu

D'une manière générale, le contenu d'un examen est dicté par des prescriptions curriculaires qui laissent peu de marges de liberté. Bien que les descripteurs de compétences («can do») du CECR soient formulés de façon abstraite, il est possible d'entrevoir des zones de conflits entre les exigences curriculaires et la façon dont le CECR est articulé. Il se pourrait que certains items de l'examen soient si complexes qu'une correspondance univoque à l'un des niveaux du CECR soit impossible; toutefois, ne pas prendre en compte le caractère équivoque pourrait également introduire des conflits avec les exigences curriculaires.

Pour solder ce problème, considérons différents points:

- La position la plus extrême est de s'abstenir totalement de lien au CECR. Bien que cela ne puisse probablement pas solder le problème à court terme, une publication à cet égard pourrait s'avérer utile pour une révision (ou une extension) du CECR, ou pour une révision des exigences curriculaires pour les rendre plus compatibles avec le CECR.
- Une approche plus nuancée pourrait être de rechercher un compromis et de relier l'examen au Cadre sur une seule partie de l'examen, en laissant de côté par exemple 25% des tâches et des items utilisés de l'examen, parce qu'ils sont très difficilement appariables aux catégories ou niveaux du CECR.
- Une autre alternative serait de sélectionner une méthode de détermination des scores de césure moins analytique, pour laquelle aucune référence spécifique aux descripteurs du CECR n'est nécessaire. Quelques méthodes de détermination des scores de césure reposent sur des jugements globaux, holistiques (par exemple la méthode du corpus de productions; voir la section 6.6), alors que d'autres impliquent des jugements globaux sur la localisation du point de césure entre les niveaux d'un test, renseignées par une somme notable d'informations psychométriques (par exemple la méthode du marque-page ou sa variante selon le Cito; voir les sections 6.8 et 6.9).

Un autre aspect de ce problème est de savoir dans quelle mesure les activités pertinentes et les compétences décrites dans le CECR sont couvertes par l'examen. Les spécifications de l'examen (chapitre 4) détaillent ce qui est inclus dans l'examen, mais pas ce qui a été laissé de côté. L'omission de parties et d'aspects importants du construit du CECR peut conduire à un caractère unilatéral et engendrer des critiques quant à la généralisabilité d'un adossement injustifié de l'examen au Cadre. Pour éviter tout danger d'une «sur-généralisation», il est préférable de mentionner explicitement le contenu couvert par l'examen (représentativité du contenu).

7.2.2. Aspects opérationnels: le test pilote

En amont de l'administration d'un examen en contexte réel, les données peuvent être collectées au cours de plusieurs étapes. D'une façon générale, on distingue la phase pilote et la phase de pré-test.

Le plus souvent, on entend par test pilote l'expérimentation du matériel de test de manière à éliminer les ambiguïtés, à vérifier la clarté et la compréhension des questions et de leurs consignes, à disposer d'une première estimation de la difficulté des tâches et des items et à estimer la durée nécessaire. Un test pilote peut être conduit sur un petit échantillon (une ou deux classes suffisent généralement); il est cependant utile de ne pas présenter le matériel exclusivement comme un test, mais d'essayer de disposer d'un maximum de retour d'information sur la qualité du matériel de test. Des méthodes qualitatives, comme les interviews et les «labos cognitifs», peuvent révéler de nombreuses informations intéressantes pour ce qui concerne l'examen planifié; les participants au test pilote peuvent être des élèves et des enseignants. Un bon pilotage permet d'éviter les mauvaises surprises lors de la phase de pré-test et de l'examen réel.

7.2.3. Aspects opérationnels: le pré-test

Un pré-test est généralement construit pour obtenir des informations sur les principales caractéristiques de l'examen planifié. En plus des paramètres psychométriques (qui seront discutés par la suite), les caractéristiques opérationnelles doivent aussi être observées. Le temps attribué et le temps nécessaire pour le pré-test est une source majeure d'information qui doit être collectée à cette fin.

En dehors du fait d'être une répétition de l'examen à venir, le pré-test permet également la réalisation d'une fonction centrale, en l'occurrence celle de relier les examens entre eux. Etant donné que les examens tendent à être uniques du point de vue de leur composition d'une année sur l'autre et que les populations cibles n'ont pas d'élèves en commun, les données recueillies sur les deux examens ne peuvent être comparées; les différences au niveau du score moyen pourraient être dues à des différences systématiques entre les deux groupes de candidats ou à une différence en termes de difficulté des contenus des deux examens ou encore par un mixte de ces deux raisons. Il n'y a aucune manière de savoir dans quelle mesure l'une et/ou l'autre de ces deux causes sont avérées, sauf si les données sont liées d'une certaine façon.

Puisque présenter des items aux mêmes candidats dans un pré-test que lors d'un examen a des conséquences imprévisibles en regard des effets mnésiques, les bonnes pratiques exigent que le pré-test soit conduit deux ans en avance (ou sur une période de deux rotations d'examens). Si les examens des années 1 et 2 doivent être liés, le pré-test qui les lie devra alors être organisé deux ans avant l'examen 2, en l'occurrence en l'année 0.

7.2.4. Considérations psychométriques

Il est primordial que suffisamment d'aspects psychométriques puissent être fournis concernant l'examen. Le premier aspect concerne les paramètres de l'item, comme la difficulté (valeur p) et le pouvoir discriminant. Si l'on s'en tient aux indices de la Théorie Classique des Tests, on doit considérer que ces indices sont dépendants de la population et que leurs valeurs sont simplement une indication des valeurs qu'ils ont au niveau de la population parente, sous l'hypothèse que l'échantillon du pré-test est représentatif de la population cible. Conduire un pré-test uniquement dans un nombre restreint de centres pour des raisons de commodités (par exemple les centres où les enseignants sont membres de l'équipe d'élaboration du test) pourrait conduire à de sérieux biais au niveau des estimations.

Ensuite, la fidélité de l'examen est un aspect important si l'on souhaite le relier correctement au CECR. En effet, elle a un impact sur la précision et la consistance de la classification en termes de niveaux du CECR, comme ce sera démontré ci-après. En estimant la fidélité, il faut avoir à l'esprit que, souvent, le KR20 (ou l'alpha de Cronbach) est mentionné comme un indice de fidélité. En fait, il ne l'est pas exactement. Il est un estimateur par défaut de la fidélité. Ainsi, avec des tests hétérogènes, la fidélité est substantiellement sous-estimée. Le Manuel propose des façons d'évaluer la fidélité de tests.

7.2.5. Le bon moment pour déterminer les scores de césure

Si l'adossement au CECR relève d'une situation à forts enjeux, le temps est généralement insuffisant pour collecter les données de l'administration de l'examen, remettre les résultats, organiser de façon complète une procédure de détermination des scores de césure et évaluer la validité de cette procédure.

Comme l'utilisation de données réelles de candidats est conseillée, y compris pour les méthodes de détermination des scores de césure centrées sur les tests (étude d'impact, retour d'information réaliste; voir le chapitre 6), le laps de temps entre le pré-test et l'administration finale de l'examen sera probablement le plus adapté pour déterminer les scores de césure.

Dans cette section, la discussion sera focalisée sur les conséquences de ce qui est parfois nommé «l'effet pré-test». Cette appellation fait référence à toutes les différences systématiques entre le pré-test et le véritable examen, différences qui pourraient moduler les performances des candidats. L'influence principale provient d'une différence en termes de motivation et de l'ensemble des facteurs directement liés à la motivation, comme le sérieux de la préparation et l'anxiété. S'il s'agit d'un examen à fort enjeu et d'un pré-test à faible enjeu, tous ces facteurs pourraient suivre la même tendance, en l'occurrence diminuer la performance dans le pré-test comparativement à la situation d'examen. Le cas échéant, la mesure d'impact présentée aux panélistes au cours de la procédure de détermination des scores de césure sera biaisée et pourrait

avoir un effet systématique sur les scores de césure proposés; suite à cette information biaisée, si les panélistes se considèrent eux-mêmes trop stricts, cela pourra conduire à minimiser les scores de césure. Le Manuel donne quelques pistes qui pourraient permettre d'éviter – ou du moins de contrôler – l'effet pré-test.

7.3. Validité procédurale de la formation à la standardisation et à la définition des scores de césure

Au cours du chapitre précédent, nous avons décrit plusieurs procédures pour familiariser les panélistes au CECR, pour comprendre les spécifications d'un examen, pour déterminer des critères pertinents et pour définir les scores de césure. Les sessions de détermination des scores de césure exigent de débiter avec de telles explications et instructions; les panélistes doivent se sentir en confiance pour réaliser leurs tâches. L'ensemble de ces procédures peut être considéré comme une étape, un pas supplémentaire vers les «bonnes pratiques»; si on les ignorait, on se dirigerait vers des situations risquées. Le respect de telles procédures est une garantie *nécessaire* pour obtenir de bons résultats, en un mot: instructions correctes, résultats corrects.

Le problème de la validité est à mettre en relation avec le caractère nécessaire des procédures. Par exemple, en ce qui concerne la familiarisation (chapitre 3) et la formation à la standardisation (chapitre 5), s'il n'y a aucune phase préparatoire relative à la compréhension du CECR, on ne peut pas espérer aboutir à un résultat valide. Par ailleurs, même si la procédure de formation suggérée est mise en œuvre, rien ne garantit que le résultat obtenu soit un succès; la phase d'entraînement (formation) est nécessaire, mais est-elle suffisante? La validation de cet aspect exige que la formation ait été efficace: si l'on forme les gens à comprendre quelque chose, on doit aussi s'assurer qu'ils l'ont réellement compris à l'issue de la formation.

Plusieurs aspects relatifs à cette validité procédurale seront exposés ci-après. Il s'agit du caractère explicite, du caractère pratique, de la mise en œuvre, du retour d'information et de la documentation.

Le caractère explicite: il s'agit du degré avec lequel l'objectif de la procédure de détermination des scores de césure et la procédure elle-même sont clairement et explicitement articulés. En d'autres termes, le processus est défini dans son intégralité avant qu'il soit conduit; les étapes sont clairement décrites, les conditions de déroulement et les résultats attendus après chaque étape sont décrits comme un scénario immuable.

Le caractère pratique: même si certaines procédures sont compliquées, la préparation doit être pratique (voir Berk, 1986), ainsi:

- la méthode de détermination des scores de césure doit pouvoir être mise en œuvre sans grande difficulté;

- l'analyse des données doit pouvoir être réalisée sans calculs laborieux. Cela ne signifie pas pour autant que les calculs ne sont pas compliqués, mais que le travail de préparation doit être accompli bien en amont de la session;
- les procédures doivent être crédibles et interprétables par des non-techniciens.

La mise en œuvre: cet aspect fait référence à la manière, du point de vue de la rigueur, dont le panel est sélectionné et formé, à la manière dont les niveaux du CECR sont intégrés et à celle dont les données de jugement sont effectivement traitées et analysées. Des informations relatives à ces points doivent être fournies.

Le retour d'information: cet aspect se réfère au niveau de confiance des panélistes à l'égard de la procédure de détermination des scores de césure et aux résultats qui y sont liés. Est-ce que les panélistes estiment qu'ils ont trouvé les bons résultats? Des informations relatives à ces points doivent être collectées et rapportées.

Documentation: cet aspect se réfère à la manière dont la procédure de détermination des scores de césure est documentée, en particulier à l'égard des objectifs d'évaluation et de communication.

7.4. Validité interne de la définition des scores de césure

Les questions relatives à la validité interne doivent permettre de se prononcer sur la *précision*, au sens de l'exactitude, et sur la *consistance* du résultat de la procédure de détermination des scores de césure. Un défaut de consistance peut provenir d'une faiblesse générale de la méthodologie mise en œuvre ou avoir une origine plus locale en reposant sur un ou deux juges ou quelques items. Le cas échéant, on pourrait: i) pour ce qui concerne les panélistes, supprimer certains d'entre eux (ou l'analyse faisant suite à la procédure de détermination) ou ii) pour ce qui concerne les items, ne retenir qu'un sous-ensemble d'items et de tâches dans le test, en excluant ceux qui posent problème.

- En supprimant des panélistes, on doit prendre garde de ne pas influencer le résultat relatif aux scores de césure dans une direction souhaitée par l'organisateur. Si l'on dispose de preuves quant à l'incompréhension des instructions à suivre par un panéliste, ou s'il les ignore volontairement, on dispose alors d'une raison valide pour le retirer des données à analyser.
- Supprimer des items ou des tâches est un problème bien plus délicat. Lorsque le premier souhait est d'adosser son examen sur le CECR (par exemple en appliquant une règle qui associe un échec à l'examen au fait ne pas avoir atteint le niveau B1/B2), retirer certains items pourrait sérieusement biaiser la validité de contenu du test.

La suite de cette section traitera des points relatifs à la consistance et à la précision:

- la consistance intra-juge consiste à rechercher les informations qui montrent qu'un juge est cohérent dans son jugement;
- la consistance inter-juges consiste à rechercher dans quelle mesure les panélistes s'accordent les uns avec les autres dans leurs jugements;
- la stabilité des résultats est exprimée par l'erreur standard des points de césure;
- la précision et la consistance de la classification reposent sur la procédure de détermination des scores de césure.

7.5. Validation externe

Le principal résultat d'une procédure de détermination des scores de césure est une règle de décision pour assigner les candidats à un petit nombre de niveaux du CECR sur la base de leurs performances à un examen. Généralement, la performance au test a déjà été résumée par un nombre unique, le score au test.

Dans ce guide, l'accent a été mis sur le fait que les procédures permettant de parvenir à une telle règle de décision sont complexes et chronophages, qu'il y a de nombreux pièges possibles et que le résultat n'est jamais parfait; notamment en raison de l'erreur de mesure dans le test et la variance résiduelle dans le jugement des panélistes. Si toutes les procédures ont été suivies attentivement, si l'examen dispose d'une validité de contenu adéquate et d'un haut degré de fidélité et si l'erreur standard des scores de césure est faible, on pourra penser que le travail est accompli et résumer les résultats par un tableau indiquant l'exactitude des décisions, comme dans la partie gauche du tableau 7.9, tout en tenant compte des limites.

Selon ce raisonnement, le point faible est qu'un tel résultat dépend totalement des procédures mises en œuvre par la même personne ou le même groupe de personnes et des données collectées en une seule occasion sur un seul groupe de candidats et sur une seule situation d'examen. Cela pourrait être considéré comme étant trop restreint pour garantir la véracité, c'est-à-dire la validité, d'une affirmation telle que: «si un étudiant obtient un score de 39 ou plus à mon test, il peut à juste titre être considéré du niveau B2». En général, la faiblesse réside dans le contraste entre la particularité des procédures et la généralité des affirmations.

La validation externe vise à fournir des preuves en provenance de sources indépendantes et qui corroborent les résultats et les conclusions de ses propres procédures. Parmi l'ensemble des preuves fournies, toutes ne sont pas indépendantes de la même façon vis-à-vis de l'information que l'on doit utiliser dans la détermination des scores de césure; de même, parmi l'ensemble des preuves fournies, toutes ne sont pas convaincantes avec le même poids.

- Les preuves pourraient provenir de résultats des mêmes candidats sur un autre test ou une autre procédure d'évaluation.

- Les preuves pourraient être fournies par une autre procédure de détermination des scores de césure en utilisant le même panel ou un panel indépendant, conduit par les mêmes organisateurs ou par une équipe indépendante.

Voici un résumé du type de preuves qui pourrait être fournies pour justifier l'affirmation relative aux règles de décision qui émanent de ses propres procédures pour relier son examen au Cadre. On pourrait tenter de tout faire, mais ce serait irréaliste parce que la collection de preuves serait particulièrement coûteuse; en outre, toutes les études ne corroboreraient pas de façon comparable les résultats.

Dans cette section, quelques exemples de validation externe seront discutés et des arguments, ainsi que leurs limites et leur caractère persuasif (ou l'absence de ce caractère), seront présentés. Cependant, en premier lieu, une remarque générale doit être faite. Dans la théorie des tests, le problème de la validité externe est généralement considéré en montrant la correspondance entre les résultats au test et des critères externes. Parfois, les mesures du critère externe sont considérées comme absolues d'une certaine manière. Cependant, en réalité, aucun critère n'est parfaitement valide. Prenons le succès académique comme exemple. Obtenir un master à l'université peut être considéré sans erreur de mesure. Un master est alors certainement utile, mais non absolu, en termes de critères des habiletés mentales. En effet, quelques étudiants pourraient échouer à l'université pour des raisons largement indépendantes de leur habileté mentale et quelques étudiants pourraient réussir sans que cela soit mérité; aucun système d'examen n'est infaillible. Ainsi, il est préférable de considérer toutes les mesures de critères comme faillibles de la même façon que les tests le sont, c'est-à-dire qu'une part de leur variance est indésirable et non pertinente pour montrer la validité de la procédure d'un test, comme avec les résultats de la détermination des scores de césure.

7.6. Conclusion

La discussion relative à la validation externe dans ce chapitre pourrait apparaître décevante à l'égard de nombreux points. En fait, elle ne propose pas de distinction claire entre ce qui relève du bon et du mauvais. Elle ne prescrit non plus de façon claire et univoque ce qu'il convient de faire pour une situation donnée.

Voici ci-après, en deux points, quelques raisons à cela.

Premièrement, il n'y a aucune autorité qui détient la vérité et la divulgation reste problématique. Les organismes de test aspirent à découvrir cette vérité encore non connue en effectuant un choix méthodologique (et/ou des méthodes psychométriques) approprié. C'est en faisant part de ces travaux à la communauté que, dans le futur, nous pourrions nous approcher de la vérité de si près que nous pourrions considérer que nous avons résolu le problème. A l'opposé, nous croyons que ce qui constitue un «B1» est essentiellement une convention pratique, mais la formulation est si claire et si consistante que deux professionnels du monde des langues s'y référant signifieront

essentiellement la même chose, même si leur culture, leur formation et leur expérience sont différentes et se réfèrent à des langues cibles différentes. Le CECR constitue un système de référence dont l'objectif est de rendre de telles affirmations possibles. Du point de vue des études de validation, cela signifie que toute étude de validation peut, en principe, offrir une critique constructive pouvant conduire à une référence plus affinée, équilibrée et élaborée; ce qui est vrai de toute expérimentation d'hypothèses, de construits et de théories.

Deuxièmement, même dans le cas d'un système de référence largement accepté, les éléments déterminants des performances à un test de langue ou à un examen sont si variés (et pas toujours totalement compris) que toute tentative pour classer les études pour relier les performances au CECR (en les qualifiant de bonnes ou de mauvaises) doit être considérée comme simpliste et catégorique. En fait, nous tentons de développer un système qui offre un éclairage sur les points forts et les points faibles de toute tentative; il ne serait donc pas réaliste de dresser un verdict catégorique et fini.

Est-ce une bonne ou une mauvaise nouvelle? Nous pensons qu'il s'agit simplement d'un état de fait. Des conclusions plus fines devraient être dressées à partir d'une méta-analyse rigoureuse qui pourrait résumer les résultats d'un large nombre d'études de validation, rigoureusement conduites sur les prochaines années. Il appartient à la génération présente de fournir les données nécessaires et la documentation pour cette méta-analyse (voir Plake, 2008 pour une revue des enjeux et des recommandations).

Il faut ainsi espérer que les acteurs des procédures de détermination des scores de césure feront leur maximum pour prendre connaissance de l'information fournie dans ce Manuel, le supplément au Manuel et les autres sources d'informations pertinentes. Il faut également espérer que ces procédures seront conduites et les rapports établis de façon transparente. En les analysant et en les comparant, les savoir-faire pour déterminer les scores de césure progresseront. La crédibilité accordée aux décisions sur les points de césure progressera également, ainsi que la portée des conséquences en découlant.

Les utilisateurs du Manuel devraient considérer:

- la meilleure manière d'obtenir les preuves de validité exigées;
- quelles techniques ils seront capables de mettre en œuvre et dans quelle mesure ils auront besoin d'un support technique;
- s'ils peuvent élaborer un argumentaire sur la validité à propos de la qualité du test et des procédures qui y sont associées (validité interne), de la qualité des procédures suivies pour relier l'examen au Cadre et en particulier pour la détermination des scores de césure (validité procédurale), de l'existence de résultats corroborés par des analyses indépendantes (validité externe);
- comment ils s'assurent, le cas échéant, que les points de césure sont comparables à travers les langues;
- s'il y a, en particulier, suffisamment de preuves pour soutenir la validité des scores de césure;
- comment ils mettront à la disposition de leurs collègues les détails de leurs conclusions.

Bibliographie

- AERA/APA/NCME (1999): American Educational Research Association, American Psychological Association, National Council on Measurement in Education: *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association. (ISBN 0-935302-25-5).
- Alderson, J. C. (2005): *Diagnosing Foreign Language Proficiency*. London: Continuum.
- Alderson, J. C., Clapham, C. and Wall, D. (1995): *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alderson, J. C., Figueras, N., Kuijpers, H., Nold, G., Takala, S. and Tardieu, C. (2006): Analysing Tests of Reading and Listening in relation to the CEFR: the experience of the Dutch CEFR Construct Project. *Language Assessment Quarterly* 3 (1): 3-30.
- American Educational Research Association (1999): *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Angoff, W. H. (1971): Scales, Norms and Equivalent Scores. In: Thorndike, R. L. (ed.) *Educational Measurement* (2nd Edition), p. 508-600. Washington, D.C.: American Council on Education.
- Beacco, J-C. and Porquier, R. (2008): *Niveau A2 pour le français: Un référentiel*. Paris: Didier.
- Beacco, J-C., Porquier, R. and Bouquet, S. (2004): *Niveau B2 pour le français: Un référentiel*. Paris: Didier. (2 vols).
- Beacco, J-C., De Ferrari, M., Lhote, G. and Tagliante, C. (2006): *Niveau A1.1 pour le français / référentiel DILF livre*. Paris: Didier.
- Beacco, J-C., Porquier, R. and Bouquet, S. (2007): *Niveau A1 pour le français: Un référentiel*. Paris: Didier.
- Berk, R.A. (1986): A Consumer's Guide to Setting Performance Standards on Criterion Referenced Tests. *Review of Educational Research*, 56, p. 137-172.
- Bolton, S., Glaboniat, M., Lorenz, H., Müller, M., Perlmann-Balme, M. and Steiner, S. (2008): *Mündlich: Mündliche Produktion und Interaktion Deutsch: Illustration der Niveaustufen des Gemeinsamen Europäischen Referenzrahmens*. Berlin: Langenscheidt.
- Breton, Jones, Lalanne, Lepage and North.: Séminaire interlangues / Cross Language Benchmarking Seminar, CIEP Sèvres, 23-25 June 2008: Report. *Strasbourg: Council of Europe*.

- Cizek, G. J. (ed.) (2001): *Setting performance standards: concepts, methods and perspectives*. Mahwah, NJ: Lawrence Erlbaum.
- Cizek, G.J. and Bunch, M.B. (2007): *Standard setting: a guide to establishing and evaluating performance standards on tests*. Thousand Oaks: Sage.
- Cohen, A., Kane, M. and Crooks, T. (1999): A generalized examinee-centered method for setting standards on achievement tests. *Applied Measurement in Education*, 12, p. 343-366.
- Council of Europe (2001a): *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe (2001b): *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Paris: Didier.
- Council of Europe (2002): Seminar on relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR), Helsinki, 30 June – 2 July 2002: Report. DGIV/EDU/LANG (2002) 15. Strasbourg: Council of Europe.
- Council of Europe (2003): Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR), DGIV/EDU/LANG (2003) 5. Strasbourg: Council of Europe.
- Council of Europe (2009): Relating Language Examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR): a Manual. Strasbourg: Council of Europe.
- Davidson, F. and Lynch, B. (1993): Criterion-referenced language test development: a prolegomenon. In: Huhta, A., Sajavaara, K. & Takala, S. (eds.), *Language testing: new openings*. Jyväskylä, Finland: University of Jyväskylä, p. 73-89.
- Davidson, F. and Lynch, B. (2002): *Testcraft: A teacher's guide to writing and using language test specifications*. Yale University Press.
- Downing, S. M. and Haladyna, T. M. (eds.) (2006): *Handbook of test development*. Earlbaum.
- Ebel, R. L. and Frisbee, O. A. (1986): *Essentials of educational measurement (4th edition)*. Englewood Cliffs, N.J.: Prentice Hall.
- Feldt, L. S., Steffen, M. and Gupta, N. C. (1985): A comparison of five methods for estimating the standard error of measurement at specific score levels. *Applied Psychological Measurement*, 9, p. 351-361.
- Ferrara, S., Perie, M. and Johnson, E. (2002): *Matching the judgmental task with standard setting panelist expertise: the item-descriptor (ID) matching procedure*. Washington DC: American Institutes for Research.

- Fienberg, S. E. (1977): *The analysis of cross-classified categorical data*. Cambridge, Massachusetts: The MIT Press.
- Fienberg, S.E., Bishop, Y. M. M. and Holland, P. W. (1975): *Discrete multivariate analysis*. Cambridge (Massachusetts): The MIT Press.
- Glaboniat, M., Müller, M., Schmitz, H., Rusch, P., Wertenschlag, L. (2002/5): *Profile Deutsch*. Berlin: Langenscheidt, ISBN 3-468-49463-7.
- Hambleton, R.K. and Pitoniak, M.J. (2006): Setting performance standards. In Brennan, R.L. (ed.) *Educational Measurement* (4th edition). Westport, CT: American Council on Education/Praeger, p. 433-470.
- Instituto Cervantes (2007): *Niveles de referencia para el español, Plan Curricular del Instituto Cervantes*. Madrid: Biblioteca Nueva.
- Jaeger, R. M. (1991): Selection of judges for standard-setting. *Educational Measurement: Issues and Practice*, 10, p. 3-6.
- Kaftandjieva, F. (2007): Quantifying the quality of linkage between language examinations and the CEF. In Carlsen, C. and Moe, E. (eds.) *A human touch to language testing*. Oslo: Novus Press, p. 34-42.
- Keats, J. A. (1957): Estimation of error variances of test scores. *Psychometrika* 22, p. 29-41.
- Kingston, N. M., Kahl, S. R., Sweeny, K. P. and Bay, L. (2001): Setting performance standards using the body of work method. In Cizek G. J. (ed.), *Setting performance standards: Concepts, methods and perspectives*. Mahwah, NJ: Erlbaum, p. 219-248.
- Kolen, M. L. and Brennan, R-L. (2004): *Test equating, scaling and linking*. New York: Springer.
- Lepage, S. and North, B. (2005): Guide for the organisation of a seminar to calibrate examples of spoken performance in line with the scales of the Common European Framework of Reference for Languages. Strasbourg: Council of Europe DGIV/EDU/LANG (2005) 4.
- Linacre, J. M. (1989): *Multi-faceted measurement*. Chicago: MESA Press.
- Linacre, J. M. (2008): *A user's guide to FACETS. Rasch model computer program*. ISBN 0-941938-03-4. www.winsteps.com.
- Livingston, S. A. and Lewis, C. (1995): Estimating the consistency and accuracy of classification based on test scores. *Journal of Educational Measurement*, 32, p. 179-197.
- Lord, F. (1965): A strong true-score theory, with applications. *Psychometrika*, 30, p. 239-270.
- Lynch, B. and Davidson, F. (1994): Criterion-referenced language test development: linking curricula, teachers and tests. *TESOL Quarterly* 28:4, p. 727-743.

Lynch, B. and Davidson, F. (1998): Criterion referencing. In: Clapham, C. & Dorson, D. (eds.) *Language testing and assessment*, Volume 7, *Encyclopedia of Language and Education*. Dordrecht: Kluwer Academic Publishers, p. 263-273.

Milanovic, M. (2002): *Language examining and test development*. Strasbourg: Language Policy Division, Council of Europe.

Mitzel, H. C., Lewis, D. M., Patz, R. J. & Green, D. R. (2001): The Bookmark Procedure: psychological perspectives. In Cizek G. J. (ed.) *Setting performance standards: concepts, methods and perspectives*, p. 249-281. Mahwah, NJ: Erlbaum.

Norcini, J., Lipner, R., Langdon, L., and Strecker, C. (1987): A comparison of three variations on a standard-setting method. *Journal of Educational Measurement*, 24, p. 56-64.

North, B (2002): Developing descriptor scales of language proficiency for the CEF common reference levels. In: Council of Europe (2002): *Common European Framework of Reference for Languages: learning, teaching, assessment: case studies*. Strasbourg: Council of Europe Publishing.

North, B. (2000a): The development of a common framework scale of language proficiency. New York: Peter Lang.

North, B. (2000b): Linking language assessments: an example in a low-stakes context. *System* 28, p. 555–577.

North, B. and Schneider, G. (1998): Scaling descriptors for language proficiency scales. *Language Testing* 15/2, p. 217-262.

OECD (2005): *Pisa 2003 technical report*. Paris: OECD.

Parizzi, F. and Spinelli, B. (forthcoming): *Profilo della lingua italiana*, Firenze: La Nuova Italia.

Plake, B. S. (2008): Standard setters: stand up and take a stand! *Educational Measurement: Issues and Practice* 27/1, p. 3-9.

Reckase, M. D. (2006a): A conceptual framework for a psychometric theory for standard setting with examples of its use for evaluating the functioning of two standard setting methods. *Educational Measurement: Issues and Practice*, 2006, 25(2), p. 4-18.

Reckase, M. D. (2006b): Rejoinder: evaluating standard setting methods using error models proposed by Schulz. *Educational Measurement: Issues and Practice*, 2006, p. 25 (3), p. 14-17.

Schneider, G. and North, B. (2000): Fremdsprachen können – Was heisst das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit. Chur/Zürich: Ruediger Verlag.

Siegel, S. and Castellan, N. J. (1988): *Non-parametric statistics for the behavioral sciences*. New York: McGraw-Hill.

- Subkoviak, M. J. (1988): A practitioner's guide to computation and interpretation of reliability for mastery tests. *Journal of Educational Measurement*, 13, p. 265-276.
- Thorndike, R.L. (ed.) (1971): *Educational Measurement* (2nd Edition), p. 508-600. Washington, D.C.: American Council on Education.
- Van der Schoot, F. (2001): *Standaarden voor Kerndoelen Basisonderwijs* [Standards for primary objectives in primary education]. PhD thesis. Arnhem: Cito.
- van Ek, Jan A. (1976): The threshold level in a European unit/credit System for modern language learning by adults. Strasbourg: Council of Europe.
- van Ek, J. A. and Trim, J. L. M., (2001a): *Waystage*. Cambridge: CUP, ISBN 0-521-56707-6
- van Ek, J. A. and Trim, J. L. M., (2001b): *Threshold 1990*. Cambridge: CUP, ISBN 0-521-56707-8
- van Ek, J. A. and Trim, J. L. M., (2001c): *Vantage*. Cambridge: CUP, ISBN 0-521-56705-X
- Verhelst, N. D. and Verstralen, H. H. F. M. (2008): Some considerations on the partial credit model. *Psicológica*, 29, 229–254.
- Weir, C. (1993): *Understanding and developing language tests*. Hemel Hempstead UK: Prentice Hall.

Glossaire concernant les processus de mise en relation (points essentiels du manuel)

Accréditation: La reconnaissance officielle de la réussite à un examen, fournie généralement par une entité officielle, un gouvernement, un centre d'examen, etc.

Alignement: Un processus qui consiste à relier le contenu, la performance standard à l'évaluation, aux consignes et à l'apprentissage en classe. La stratégie type d'alignement est de progressivement concevoir (a) le contenu standard, (b) la performance standard, (c) les évaluations et (d) les consignes pour l'apprentissage en classe.

Approche compensatoire: Une approche qui permet qu'une compétence de haut niveau à une des composantes d'une évaluation compense un bas niveau dans d'autres composantes.

Approche conjonctive: Une approche qui suppose qu'un niveau minimal prédéfini de compétence soit atteint dans chacune des composantes afin que le résultat final puisse être considéré comme étant acceptable.

Biais: Un test ou un item peuvent être considérés comme biaisés si un de leurs attributs se révèle non pertinent par rapport à ce qu'ils sont censés tester et qu'ils avantagent ou désavantagent une partie des candidats. Le biais est principalement lié au sexe, à l'âge, à la culture, etc.

Calibrage: Une description détaillée et validée d'un niveau particulier de performance attendu de candidats d'âge et de niveau donnés. Les performances à calibrer sont souvent des échantillons de performances d'étudiants.

Combiner: Etablir un score total à partir de 2 ou plusieurs scores.

Concept hypothétique/construit (construct): L'attribut hypothétique des individus ou opération mentale qui ne peut être directement ni observée ni mesurée (par exemple en évaluation des langues, la capacité de réception orale).

Contenus standards: Ce que l'on attend généralement de candidats en termes de connaissance et de savoir faire dans des situations et à des niveaux spécifiques.

Définition des scores de césure sur plusieurs langues: Une méthode qui a pour objectif de vérifier que des examens de langues différentes sont reliés aux standards communs en utilisant des procédures comparables.

Descripteur de niveau de performance: Descriptions des standards que des candidats devraient avoir atteints. Les descriptions des niveaux dans le CECR sont des exemples de ces standards.

- Echantillons représentatifs:** Exemples de performance qui, après validation, sont représentatifs d'un niveau de performance.
- Equation d'ajustement:** Le processus consistant à comparer la difficulté de deux ou plusieurs types de tests afin d'en établir leurs équivalences.
- Estimation de difficulté de l'item:** Dans la théorie classique du test, la difficulté d'un item est la proportion de candidats qui y répondent correctement. Dans la théorie de réponse à l'item, c'est l'estimation de la difficulté d'un item calculée indépendamment de la population.
- Evaluer:** Le processus consistant à attribuer un score à une performance dans un test en utilisant sa capacité d'évaluer.
- Evaluateur:** Une personne qui évalue la performance d'un candidat selon des critères précis.
- Evaluation à faible enjeu:** Une évaluation dont les enjeux sont peu importants pour le candidat qui s'y soumet.
- Evaluation à enjeu critique:** Une évaluation dont les conséquences sont importantes pour le candidat qui s'y soumet.
- Evaluation holistique:** Une évaluation d'un candidat dont le score dépend d'un jugement global de sa performance plutôt que de critères indépendants spécifiques.
- Familiarisation:** L'organisation de tâches permettant aux personnes impliquées dans une démarche de mise en relation d'un examen avec les niveaux du CECR d'avoir une excellente connaissance du processus.
- Fiabilité/fidélité inter correcteurs:** Le degré d'accord sur des évaluations de performances effectuées par des évaluateurs différents.
- Fiabilité/fidélité intra correcteur:** Le degré de similitude entre les évaluations d'une même performance faites par le même évaluateur à des moments différents.
- Grille d'évaluation:** Un ensemble de critères d'évaluation présentés sous forme de grille.
- KR20:** Une mesure de la consistance interne conçue par Kuder et Richardson et mise en œuvre pour estimer la fiabilité d'un test.
- Maîtrise:** Le terme indiquant que le candidat a répondu à un ensemble de critères, définis en termes de capacités et de connaissances.
- (Le) Manuel:** Le document produit par le Conseil de l'Europe pour aider à relier les tests et les examens au CECR.
- Méthode centrée sur le candidat:** Une méthode de définition des points de césure selon laquelle une personne qui connaît bien les candidats fournit une évaluation holistique de leur compétence langagière, en attribuant par exemple un niveau du CECR.

Méthode centrée sur le test: Un ensemble de méthodes où des évaluateurs estiment par exemple quel niveau un candidat doit avoir acquis pour répondre correctement à un ensemble d'items.

Panel: Un ensemble d'évaluateurs.

Paneliste: Un membre du panel.

Performance limite: Un niveau de connaissance et de capacités permettant d'être juste au niveau de la performance attendue. Par exemple le niveau B2.

Performance standard: La définition explicite de ce que des étudiants doivent faire pour faire preuve de leur compétence en relation à des contenus standards à un niveau défini.

Phase pilote: Une étude préliminaire menée par des concepteurs de tests qui soumettent des tâches à un nombre limité de sujets pour repérer les problèmes avant de faire faire une expérimentation à grande échelle.

Pre-test: L'étape de la conception des tests au cours de laquelle on expérimente les items sur des échantillons représentatifs de la population cible afin de déterminer leur degré de difficulté. Suivant une analyse statistique, les items considérés comme satisfaisants pourront être utilisés dans des tests réels.

Probabilité de réponse: Dans la définition des points de césure, c'est un critère essentiel. Dans beaucoup de tests, il est fixé aux 2/3 du score maximum, tandis que quelques experts préfèrent le fixer à 50% et d'autres à 80%.

Régression logistique: Une technique statistique fournissant une formule qui traduit une ou plusieurs informations (par exemple les scores d'une personne à un test) en une estimation de probabilité d'un événement spécifique (par exemple l'évaluation du travail d'un étudiant au niveau d'utilisateur expérimenté).

Réponses construites: Une réponse construite à un item. Cette réponse implique une production active et non un simple choix parmi différentes propositions.

Scores de césure: Le plus petit score qu'un candidat doit avoir pour qu'on puisse lui attribuer un niveau ou un classement dans un test ou un examen.

Spécification: Une étape dans le processus de mise en relation qui concerne l'analyse du contenu d'un examen ou d'un test afin de le relier au CECR.

Spécifications du test: Une description des caractéristiques d'un examen comprenant ce qui est testé (contenu, niveau, tâche), comment le test se déroule, la population cible, l'usage du test avec des détails sur le nombre et la longueur des épreuves, le type d'items utilisé.

Test direct: Un test qui mesure la capacité de production orale ou écrite et où la performance est directement mesurée.

Test indirect: Un test ou une tâche à réaliser qui tente de mesurer les capacités sous-jacentes à une aptitude langagière plutôt que de tester la performance directement. On teste par exemple la capacité de production écrite en demandant au candidat de noter dans un texte les structures incorrectes.

Théorie classique des tests: La TCT se réfère à un ensemble de modèles statistiques de données de test. La notion de base de la TCT est que le score X observé obtenu par une personne P passant une version V du test X est la somme du score vrai et de l'erreur de mesurer. Voir aussi la théorie de réponse à l'item. (TRI).

Théorie de réponse à l'item: La TRI est utilisée pour mettre en relation les compétences d'un candidat aux données d'un test. L'analyse concerne l'item contrairement à la théorie classique des tests (TCT) qui se concentre sur les scores du test.

Transparence: Le terme implique ouverture, communication et responsabilité. C'est une extension du terme utilisé en physique (on voit à travers un objet transparent).

Validité de contenu: On dit d'un test qu'il a une validité de contenu si les items ou les tâches dont il est composé sont représentatifs des items ou des tâches dans la capacité ou la connaissance à évaluer.

Validation croisée: L'application d'un système de notation d'un échantillon appliqué à un autre échantillon venant d'une population identique.

Validité de décisions: Le degré de similitude à l'issue d'évaluations similaires répétées des décisions de classement d'un même candidat.

Validation externe: La collecte de preuves venant de sources indépendantes qui confirment les résultats des procédures utilisées.

Validation interne: Le processus consistant à s'assurer de la précision et de la consistance des évaluations à partir des évaluations dans le test.

Validation procédurale: Collecte de preuves démontrant que les procédures adéquates ont été mise en œuvre aux différentes étapes de la définition des points de césure.

Sales agents for publications of the Council of Europe Agents de vente des publications du Conseil de l'Europe

BELGIUM/BELGIQUE

La Librairie Européenne -
The European Bookshop
Rue de l'Orme, 1
BE-1040 BRUXELLES
Tel.: +32 (0)2 231 04 35
Fax: +32 (0)2 735 08 60
E-mail: info@libeurop.eu
http://www.libeurop.be

Jean De Lannoy/DL Services
Avenue du Roi 202 Koningslaan
BE-1190 BRUXELLES
Tel.: +32 (0)2 538 43 08
Fax: +32 (0)2 538 08 41
E-mail: jean.de.lannoy@dl-servi.com
http://www.jean-de-lannoy.be

BOSNIA AND HERZEGOVINA/ BOSNIE-HERZÉGOVINE

Robert's Plus d.o.o.
Marka Manuĉa 2/V
BA-71000, SARAJEVO
Tel.: + 387 33 640 818
Fax: + 387 33 640 818
E-mail: robertsplus@bih.net.ba

CANADA

Renouf Publishing Co. Ltd.
1-5369 Canotek Road
CA-OTTAWA, Ontario K1J 9J3
Tel.: +1 613 745 2665
Fax: +1 613 745 7660
Toll-Free Tel.: (866) 767-6766
E-mail: order.dept@renoufbooks.com
http://www.renoufbooks.com

CROATIA/CROATIE

Robert's Plus d.o.o.
Marasoviĉeva 67
HR-21000, SPLIT
Tel.: + 385 21 315 800, 801, 802, 803
Fax: + 385 21 315 804
E-mail: robertsplus@robertsplus.hr

CZECH REPUBLIC/ RÉPUBLIQUE TCHÈQUE

Suweco CZ, s.r.o.
Klecakova 347
CZ-180 21 PRAHA 9
Tel.: +420 2 424 59 204
Fax: +420 2 848 21 646
E-mail: import@suweco.cz
http://www.suweco.cz

DENMARK/DANEMARK

GAD
Vimmelskaffet 32
DK-1161 KØBENHAVN K
Tel.: +45 77 66 60 00
Fax: +45 77 66 60 01
E-mail: gad@gad.dk
http://www.gad.dk

FINLAND/FINLANDE

Akateeminen Kirjakauppa
PO Box 128
Keskuskatu 1
FI-00100 HELSINKI
Tel.: +358 (0)9 121 4430
Fax: +358 (0)9 121 4242
E-mail: akatilaus@akateeminen.com
http://www.akateeminen.com

FRANCE

La Documentation française
(diffusion/distribution France entière)
124, rue Henri Barbusse
FR-93308 AUBERVILLIERS CEDEX
Tél.: +33 (0)1 40 15 70 00
Fax: +33 (0)1 40 15 68 00
E-mail: commande@ladocumentationfrancaise.fr
http://www.ladocumentationfrancaise.fr

Librairie Kléber
1 rue des Fracs Bourgeois
FR-67000 STRASBOURG
Tel.: +33 (0)3 88 15 78 88
Fax: +33 (0)3 88 15 78 80
E-mail: librairie-kleber@coe.int
http://www.librairie-kleber.com

GERMANY/ALLEMAGNE

AUSTRIA/AUTRICHE
UNO Verlag GmbH
August-Bebel-Allee 6
DE-53175 BONN
Tel.: +49 (0)228 94 90 20
Fax: +49 (0)228 94 90 222
E-mail: bestellung@uno-verlag.de
http://www.uno-verlag.de

GREECE/GRÈCE

Librairie Kauffmann s.a.
Stadiou 28
GR-105 64 ATHINAI
Tel.: +30 210 32 55 321
Fax: +30 210 32 30 320
E-mail: ord@otenet.gr
http://www.kauffmann.gr

HUNGARY/HONGRIE

Euro Info Service
Pannónia u. 58.
PF. 1039
HU-1136 BUDAPEST
Tel.: +36 1 329 2170
Fax: +36 1 349 2053
E-mail: euroinfo@euroinfo.hu
http://www.euroinfo.hu

ITALY/ITALIE

Licosa SpA
Via Duca di Calabria, 1/1
IT-50125 FIRENZE
Tel.: +39 0556 483215
Fax: +39 0556 41257
E-mail: licosa@licosa.com
http://www.licosa.com

NORWAY/NORVÈGE

Akademika
Postboks 84 Blindern
NO-0314 OSLO
Tel.: +47 2 218 8100
Fax: +47 2 218 8103
E-mail: support@akademika.no
http://www.akademika.no

POLAND/POLOGNE

Ars Polona JSC
25 Obroncow Street
PL-03-933 WARSZAWA
Tel.: +48 (0)22 509 86 00
Fax: +48 (0)22 509 86 10
E-mail: arspolona@arspolona.com.pl
http://www.arspolona.com.pl

PORTUGAL

Livraria Portugal
(Dias & Andrade, Lda.)
Rua do Carmo, 70
PT-1200-094 LISBOA
Tel.: +351 21 347 42 82 / 85
Fax: +351 21 347 02 64
E-mail: info@livrariaportugal.pt
http://www.livrariaportugal.pt

RUSSIAN FEDERATION/ FÉDÉRATION DE RUSSIE

Ves Mir
17b, Butlerova ul.
RU-101000 MOSCOW
Tel.: +7 495 739 0971
Fax: +7 495 739 0971
E-mail: orders@vesmirbooks.ru
http://www.vesmirbooks.ru

SPAIN/ESPAGNE

Diaz de Santos Barcelona
C/ Balmes, 417-419
ES-08022 BARCELONA
Tel.: +34 93 212 86 47
Fax: +34 93 211 49 91
E-mail: david@diazdesantos.es
http://www.diazdesantos.es

Diaz de Santos Madrid

C/Albasanz, 2
ES-28037 MADRID
Tel.: +34 91 743 48 90
Fax: +34 91 743 40 23
E-mail: jpinilla@diazdesantos.es
http://www.diazdesantos.es

SWITZERLAND/SUISSE

Planetis Sàrl
16 chemin des Pins
CH-1273 ARZIER
Tel.: +41 22 366 51 77
Fax: +41 22 366 51 78
E-mail: info@planetis.ch

UNITED KINGDOM/ROYAUME-UNI

The Stationery Office Ltd
PO Box 29
GB-NORWICH NR3 1GN
Tel.: +44 (0)870 600 5522
Fax: +44 (0)870 600 5533
E-mail: book.enquiries@tso.co.uk
http://www.tsoshop.co.uk

UNITED STATES and CANADA/ ÉTATS-UNIS et CANADA

Manhattan Publishing Co
2036 Albany Post Road
USA-10520 CROTON ON HUDSON, NY
Tel.: +1 914 271 5194
Fax: +1 914 271 5886
E-mail: coe@manhattanpublishing.com
http://www.manhattanpublishing.com

Council of Europe Publishing/Éditions du Conseil de l'Europe

FR-67075 STRASBOURG Cedex

Tel.: +33 (0)3 88 41 25 81 – Fax: +33 (0)3 88 41 39 10 – E-mail: publishing@coe.int – Website: http://book.coe.int