



Septembre 2003

DGIV/EDU/LANG (2003) 5 rév. 1

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

**Manuel
Avant-projet**

**Division des Politiques Linguistiques
Strasbourg**

Table des matières

Liste des Figures	Page v
Liste des Tableaux	Page vii
Liste des Fiches	Page ix
Préface	Page xi
Chapitre 1 Introduction	Page 1
Chapitre 2 Le Cadre européen commun de référence	Page 13
Chapitre 3 Familiarisation	Page 28
Chapitre 4 Spécification	Page 32
Chapitre 5 Standardisation des évaluations	Page 67
Chapitre 6 Validation empirique	Page 102
Chapitre 7 Grandes lignes du rapport	Page 127
Annexe Rapport pour le grand public	Page 131

Références supplémentaires (à paraître séparément) :

Section A: Summary of the Linking Process

Section B: Standard-setting

Section C: Classical Test Theory

Section D: Qualitative Analysis Methods

Section E: Generalisability Theory

Section F: Factor Analysis

Section G: Item Response Theory

Section H: Test Equating

Liste des Figures

Figure 1.1	Représentation graphique des procédures permettant de relier les examens au CECR	Page 4
Figure 2.1	Approche arborescente des niveaux du CECR	Page 17
Figure 5.1	Le processus de standardisation	Page 72
Figure 6.1	Analyse graphique d'items	Page 107
Figure 6.2	Diagrammes de corrélation	Page 113
Figure 6.3	Diagrammes de corrélation pour des corrélations différentes	Page 114
Figure 6.4	Puissance de décision	Page 119
Figure 6.5	Boîtes à moustaches	Page 121
Figure 6.6	Référencer une banque d'items sur l'évaluation des enseignants	Page 124

Liste des Tableaux

Tableau 2.1	Les niveaux communs de référence	Page 19
Tableau 3.1	Gestion du temps pour les activités de Familiarisation	Page 31
Tableau 4.1	Fiches et échelles du CECR pour les activités de communication langagière	Page 36
Tableau 4.2	Echelles du CECR pour les aspects de la compétence communicative	Page 37
Tableau 4.3	Facteurs qualitatifs pertinents pour la réception	Page 56
Tableau 4.4	Paramètres qualitatifs pertinents pour l'interaction	Page 62
Tableau 4.5	Paramètres qualitatifs pertinents pour la production	Page 63
Tableau 5.1	Planification du processus de standardisation	Page 75
Tableau 5.2	Echantillons de performances pour la standardisation	Page 76
Tableau 5.3	Planification du temps pour l'évaluation d'échantillons vidéos de performances	Page 79
Tableau 5.4	Echelle de production orale générale	Page 81
Tableau 5.5	Grille des critères d'évaluation de l'oral (CECR Tableau 3)	Page 82
Tableau 5.6	Planification du temps pour l'évaluation d'échantillons de performances écrites	Page 84
Tableau 5.7	Documents et matériel à préparer pour l'évaluation de la production écrite	Page 84
Tableau 5.8	Grille des critères d'évaluation des productions écrites	Page 85
Tableau 5.9	Sources de références dans le CECR	Page 87
Tableau 5.10	Planification de la standardisation	Page 96
Tableau 5.11	Récapitulatif pour la formation à la standardisation	Page 101
Tableau 6.1	Table de décision	Page 114
Tableau 6.2	Seuils de l'échelle du CECR	Page 122
Tableau 6.3	Figure 6.6 sous forme de Table de décision	Page 125

Liste des Fiches

Fiche A1	Description générale de l'examen	Page 39
Fiche A2	Elaboration de l'examen	Page 40
Fiche A3	Corrélation	Page 41
Fiche A4	Notation	Page 42
Fiche A5	Délivrance des résultats	Page 42
Fiche A6	Analyse et révision de l'examen	Page 43
Fiche A7	Justification des décisions prises	Page 43
Fiche A8	Impression initiale du niveau globale	Page 44
Fiche A9	Compréhension de l'oral	Page 44
Fiche A10	Compréhension de l'écrit	Page 46
Fiche A11	Interaction orale	Page 47
Fiche A12	Interaction écrite	Page 49
Fiche A13	Production orale	Page 50
Fiche A14	Production écrite	Page 51
Fiche A15	Combinaisons de compétences intégrées	Page 52
Fiche A16	Compétences intégrées	Page 52
Fiche A17	Médiation orale	Page 53
Fiche A18	Médiation écrite	Page 54
Fiche A19	Aspects de la compétence langagière pour la réception	Page 57
Fiche A20	Aspects de la compétence langagière en interaction	Page 59
Fiche A21	Aspects de la compétence langagière en production	Page 60
Fiche A22	Aspects de la compétence langagière en médiation	Page 64
Fiche A23	Représentation graphique de la relation de l'examen aux niveaux du CECR	Page 66
Fiche B1	Standardisation : fiche de synthèse	Page 71
Fiche B2	Evaluation analytique	Page 83
Fiche B3	Fiche d'évaluation globale (DIALANG)	Page 83
Fiche B4	Fiche de synthèse de l'évaluation globale (DIALANG)	Page 84
Fiche B5	Fiche d'évaluation des questions/item (DIALANG)	Page 88
Fiche B6	Consignes d'évaluation (Finlande)	Page 98
Fiche B7	Fiche d'évaluation des items (Finlande)	Page 98
Annexe	Rapport pour le grand public	Page 131

Ces fiches sont également disponibles sur le site Internet www.coe.int/lang/fr

Préface de la version expérimentale

La Division des Politiques linguistiques du Conseil de l'Europe à Strasbourg publie cette version expérimentale afin d'aider les Etats membres et les producteurs d'examens au niveau national ou international à mettre en relation leurs certificats et leurs diplômes de langues avec le *Cadre européen commun de référence pour les langues*.

Le travail a été lancé au cours d'un séminaire gracieusement accueilli par les autorités finlandaises à Helsinki en juillet 2002 [DGIV/EDU/LANG (15) 2002]. Ce séminaire avait été organisé pour apporter une réponse à la question complexe que se posent les organismes de certification et les centres d'examen, question que le professeur J. Charles Anderson a brièvement résumé ainsi :

« Comment puis-je savoir que mon Niveau B1 est votre Niveau B1 ? »

Cette version expérimentale du Manuel « Relier les examens de langues au Cadre européen de référence pour les langues : apprendre, enseigner, évaluer (CECR) » devrait aider les Etats membres et les producteurs d'examens à trouver une réponse convenable à la question posée ci-dessus.

Ce Manuel s'inscrit dans la continuation des travaux de la Division des Politiques linguistiques du Conseil de l'Europe : élaborer des outils de planification présentant des éléments de référence et des objectifs communs comme fondements d'une structure cohérente et transparente pour un enseignement/apprentissage efficace et une évaluation pertinente prenant en compte les besoins des apprenants et de la société afin de favoriser la mobilité des individus. Ces travaux ont été largement diffusés et reconnus dans les années 70 avec la publication du « Threshold Level » et l'élaboration de versions dans différentes langues. Les années 90 ont été celles de la recherche puis de la conception du CECR, d'abord diffusé dans deux versions expérimentales avant sa publication définitive en 2001 (Année européenne des langues). Outre les deux versions officielles publiées par le Conseil de l'Europe en anglais et en français, on trouve maintenant ce document en allemand, basque, catalan, espagnol, finnois, galicien, géorgien, hongrois, italien, japonais, moldave, polonais, portugais, russe, serbe, tchèque et ukrainien. D'autres versions sont en préparation. L'impact initial le plus important du CECR a été la mise en place des « Niveaux communs de référence » (A1 - C2). Les concepteurs de programmes s'inspirent désormais du CECR pour définir une génération nouvelle d'ensembles d'objectifs développés à partir des descripteurs du CECR. Les prototypes en sont « Profile Deutsch » et « Un référentiel pour le français ». Le présent Manuel, qui met l'accent sur la relation des *évaluations* entre elles par le biais du CECR, est un complément logique à l'élaboration de ces *niveaux* et de ces *objectifs*.

Le Conseil de l'Europe est très reconnaissant à tous ceux qui, grâce à leurs contributions, ont rendu possible l'élaboration de cette édition expérimentale et notamment :

- Les autorités finlandaises qui ont permis la tenue du séminaire où fut lancé le projet ;
- le Groupe de consultants (Prof. Charles Alderson, Dr Gergely A. David, Dr John De Jong, Dr Felianka Kaftandjieva, Dr Michael Makosch, Dr Michael Milanovic, Dr Günther Nold, Professor Mats Oscarson, Prof. Günther Schneider, Dr Claude Springer et aussi Mr Josef Biro, Ms Erna van Hest, Mr Peter Lenz, Ms Jana Pernicová, Dr Vladimir Kondrat Shleg, Ms Christine Tagliante et Dr John Trim) dont la plupart avaient participé au séminaire d'Helsinki, pour le retour d'information détaillé qu'ils ont donné des cinq versions préliminaires ;
- le Groupe d'auteurs, sous la direction de M. Brian North :
 - Mme Neus Figueras - Departament d'Ensenyament, Généralité de Catalogne, Espagne
 - M. Brian North - Fondation Eurocentres, Suisse
 - M. Sauli Takala - Université de Jyväskylä, Finlande
 - M. Piet Van Avermaet - Centre for Language and Migration, Université catholique de Louvain, Belgique
 - Association des centres d'évaluation en langue en Europe (ALTE)
 - M. Norman Verhelst - CITO, Pays Bas

- Mme Jay Banerjee (Université de Lancaster) et Mme Felianka Kaftandjieva (Université de Sofia) pour leur aide au Groupe d'auteurs en produisant les Références complémentaires du Manuel. ;
- les Ecoles-clubs Migros et la Fondation des Eurocentres pour avoir rendu possible la production d'une cassette vidéo expérimentale présentant des performances étalonnées sur le CECR en vue de la formation à la standardisation ;
- les autorités finlandaises et le « University of Cambridge Local Examination Syndicate » pour les exemples d'items de Compréhension de l'écrit destinés à la formation ;
- tous les organismes qui ont proposé leur aide pour expérimenter cette version pilote du Manuel.

L'expérimentation préliminaire aura lieu pendant l'année scolaire 2003-2004. Elle débouchera sur une édition révisée à publier fin 2005 ou début 2006. Cette seconde édition présentera des exemples de (échantillons de performances et d'items de tests) représentatifs des principaux cycles d'enseignement, et cela, du moins nous l'espérons, dans un nombre relativement élevé de langues.

L'ampleur des matériaux de standardisation qui seront présentés dans la seconde édition expérimentale dépendra, dans une large mesure, de l'implication des Etats membres et des producteurs d'examens dans la présente version.

La Division des Politiques linguistiques et son Groupe d'auteurs sont donc reconnaissants pour toute offre de coopération à ce projet, ambitieux certes, mais qui peut contribuer de manière significative à la progression de la transparence et de la cohérence en ce qui concerne les niveaux d'apprentissage, les objectifs et l'évaluation en langue en Europe.

L'implication peut se situer à trois niveaux :

- Le **retour d'information** sur l'édition expérimentale. Des questionnaires et des formulaires standards sont disponibles sur le site Internet du Conseil de l'Europe : www.coe.int/lang
- L'**expérimentation** de tout ou partie du processus de mise en relation. Outre le retour d'information déjà mentionné, des suggestions concrètes de révisions issues de l'expérience seraient les bienvenues.
- Une **étude de cas** qui suivrait le processus de validation de la mise en relation des examens de langue et du CECR et apporterait des témoignages. Les études de cas fourniront des modèles standards pour les éditions successives du Manuel ainsi que des exemples de fiches complétées. Les études de cas seront soumises à un processus paritaire de révision et une sélection en sera publiée avec la seconde édition pilote dans une brochure complémentaire sous ISBN.

Toutes les fiches incluses dans la version papier sont disponibles sur le site Internet www.coe.int/lang/fr. Les Organismes désirant s'impliquer à un quelconque degré de coopération sont invités à consulter le document « Orientations pour les retours d'information, le pilotage et les études de cas » et à compléter le Formulaire électronique de participation à l'un ou plusieurs des trois niveaux décrits ci-dessus.

Pour des copies supplémentaires du "Manuel", veuillez contacter Johanna.Panthier@coe.int

Division des Politiques linguistiques

Direction de l'éducation scolaire, extrascolaire et de l'enseignement supérieur (DG IV)
F – 67075 STRASBOURG

<http://www.coe.int/lang/fr>

<http://www.coe.int/portfolio>

Brian North, Coordinateur, Groupe d'auteurs
Zürich

Le Groupe d'auteurs :

- Mme Neus Figueras
 - Departament d'Ensenyament, Généralité de Catalogne, Espagne
- M. Brian North
 - Fondation Eurocentres, Suisse
- M. Sauli Takala
 - Université de Jyväskylä, Finlande
- M. Piet Van Avermaet
 - Centre for Language and Migration, Université catholique de Louvain, Belgique
 - Association des centres d'évaluation en langue en Europe (ALTE)
- M. Norman Verhelst
 - CITO, Pays-Bas

Chapitre 1

Introduction

Ce Manuel a pour principal objectif d'aider les concepteurs d'examens à mettre en œuvre des démarches 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. L'approche développée ici propose des conseils pour :

- décrire ce que recouvre l'examen ainsi que les démarches de passation et d'analyse ;
- mettre en relation les résultats de l'examen et les « Niveaux communs de référence » présentés au Chapitre 3 du CECR ;
- rendre compte des procédures suivies pour ce faire en apportant des preuves à l'appui.

Toutefois, selon 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 encourager et à faciliter activement la coopération entre les organismes concernés et les spécialistes des pays membres. Le Manuel s'attache donc à :

- contribuer à construire une compétence dans le domaine de la relation des examens de langues avec le CECR ;
- encourager une plus grande transparence chez les concepteurs d'examen ;
- encourager la constitution de réseaux d'organismes et d'experts, officiels ou non, tant au plan national qu'international.

La mise en relation d'un examen ou d'un test avec le CECR est une entreprise complexe. L'existence d'une telle relation n'est pas un fait facilement observable mais relève d'une affirmation pour laquelle le producteur de l'examen doit apporter des preuves tant au plan théorique qu'empirique. On peut résumer les démarches qui permettent d'avancer ces preuves par l'expression « validation de l'affirmation ».

Il faut souligner que, si ce Manuel recouvre un large éventail d'activités, son ambition est cependant limitée :

- Il fournit un guide spécialement centré sur les procédures à mettre en œuvre pour permettre de valider l'affirmation qu'un examen ou un test donné est relié au CECR.
- Il *ne* fournit *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. La mise en relation des examens avec le CECR n'a de sens que si les examens sont de qualité.
- 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 et de l'utilisation des langues, il admet, dans son effort d'exhaustivité, que des examens différents puissent refléter des buts différents (concepts hypothétiques ou « constructs »). Avant de s'engager dans la mise en relation d'un examen avec le CECR il faut, au préalable, que les producteurs de cet examen en prouvent la validité en démontrant qu'il évalue bien les concepts hypothétiques ou « constructs » envisagés.

1.1. Approche retenue

Le manuel présente quatre séries de démarches étroitement liées. On conseille aux utilisateurs de suivre ces démarches une à une afin de concevoir leur plan de mise en relation en termes d'activités

indépendantes et réalisables. Les activités exécutées dans le cadre des quatre séries de démarches entrent dans le processus de validation.

Familiarisation : choix d'activités qui permettent de s'assurer que ceux qui participent à la mise en relation connaissent le CECR dans le détail. Cette étape de familiarisation doit se faire en amont des démarches de Spécification et de Standardisation.

En termes de validation, ces démarches constituent un point de départ indispensable. Un compte rendu des activités entreprises et des résultats obtenus est une composante essentielle préalable au rapport de validation.

Spécification : inventaire de ce que l'examen recouvre (contenu et types de tâches) par rapport aux catégories présentées dans le CECR aux Chapitres 4 et 5 : « L'utilisation de la langue et l'apprenant/utilisateur » et « Les compétences de l'apprenant/utilisateur ». Tout en faisant fonction de compte rendu, cet exercice sert aussi, dans une certaine mesure, à une prise de conscience qui pourra ultérieurement aider à améliorer la qualité de l'examen en question.

Ces démarches garantissent que l'on a soigneusement défini et élaboré l'examen selon les procédures recommandées.

Standardisation : démarches proposées pour permettre d'atteindre une compréhension commune des « Niveaux communs de référence » présentés au Chapitre 3 du CECR. Des modèles standards seront fournis en appui à la formation à la standardisation des évaluations.

Ces démarches garantissent que les jugements émis au moment de l'évaluation des performances reflètent les concepts hypothétiques décrits dans le CECR et que les décisions relatives à la difficulté des tâches et des activités sont régies par des principes fondés sur les résultats de pré-tests ainsi que sur des avis d'experts.

Validation empirique : recueil et analyse de données de tests et permettant d'apporter des preuves que l'examen lui-même et son lien au CECR sont valables. Des suggestions et des critères pour une validation adéquate et crédible adaptée à différents contextes sont fournis.

Ces démarches garantissent que l'on peut effectivement confirmer les hypothèses faites lors de la Spécification et de la Standardisation (« épreuves en cours de production ») par la passation réelle de l'examen (« épreuves mises en œuvre ») et par la mise à disposition des données sur le comportement des candidats appartenant à la population cible lors de la passation.

La mise en relation des examens avec le CECR peut réellement être considérée comme un processus « d'argumentation » basé sur un raisonnement théorique. Comme on l'a souligné ci-dessus, le concept de « validité » est au centre de ce processus.

De toute évidence, il faut d'abord garantir la **Familiarisation** avec le CECR (Chapitre 3) avant d'entreprendre une quelconque mise en relation.

Avant même qu'un examen puisse être relié à un cadre de référence externe tel que le CECR (validité externe), il faut prouver la validité du concept hypothétique (construct) ainsi que la fidélité et la constance de l'examen (« validité interne »). On peut combiner des méthodes quantitatives et qualitatives pour prouver la validité interne et externe. On peut considérer la **Spécification** (Chapitre 4) comme une méthode qualitative : apport de preuves à l'aide « d'arguments fondés sur le contenu ». Le résultat qui apparaît lorsqu'on remplit les fiches A1 et A3 à A7 du Chapitre 4 met l'accent sur la validité *interne* des examens. Les fiches A2 et A8 à A20 traitent essentiellement, de manière

qualitative, sur la validité externe. Il existe également des méthodes quantitatives de validation du contenu mais ce Manuel n'exige pas qu'on les utilise.

La **Standardisation** (Chapitre 5) entraîne des démarches à la fois qualitatives et quantitatives simples – par la formation et la comparaison avec des échantillons de tests et de performances étalonnés – pour prouver la validité externe. Même si ces activités sont d'abord orientées vers le qualitatif, il faut également apporter la preuve quantitative du niveau de réussite dans la standardisation des évaluations.

Enfin, la **Validation empirique** (Chapitre 6) utilise des démarches quantitatives fondées sur le recueil et l'analyse de données afin de prouver d'abord « la validité interne » et ensuite « la validité externe ». Le Chapitre 6 montre qu'une validation empirique correcte exige un savoir-faire important en psychométrie tout autant que l'exige l'élaboration d'un test. Si le producteur d'examen ne dispose pas de ce type d'expérience, on lui recommande d'acquérir une formation suffisante ou de s'adjoindre les services d'un psychométricien qualifié.

L'approche retenue pour ce processus est globale. Les démarches recommandées dans chacun des chapitres mentionnés ci-dessus encouragent l'alignement des examens sur le CECR avec des degrés différents de rigueur selon les différentes situations d'évaluation. Le Manuel vise à favoriser l'application des procédures correctes même dans les cas où les ressources et l'expertise disponibles sont limitées. Les premiers pas peuvent être modestes mais le but est d'aider les producteurs d'examens à travailler dans un cadre structuré de sorte que le travail ultérieur puisse s'appuyer sur ce qui a été fait précédemment ; en outre, une structure commune peut donner l'occasion à des organismes de conjuguer plus facilement leurs efforts dans certains domaines.

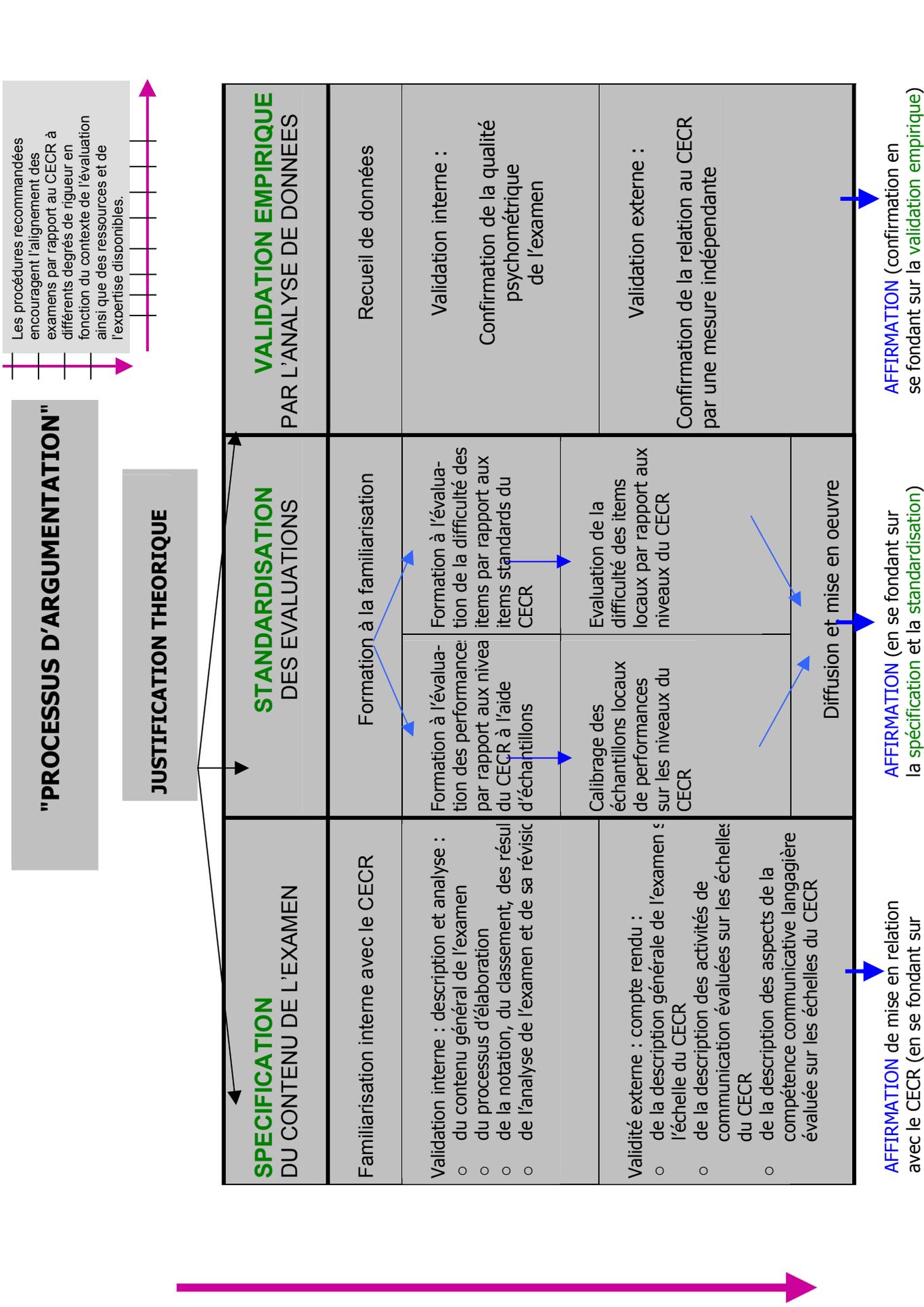
Les techniques préconisées s'organisent selon un ordre logique de sorte que tous les utilisateurs puissent être en mesure de suivre la même approche générale. On encourage les utilisateurs à commencer par la Familiarisation et on les aide à choisir des techniques pour chaque série de démarches, la Spécification, la Standardisation et la Validation empirique. Parmi les choix et les techniques proposés ici, ou que l'on peut trouver dans d'autres documents de référence, on demande aux utilisateurs d'identifier ceux qui sont les plus appropriés et réalisables dans leur contexte d'évaluation.

La Division des Politiques linguistiques a l'intention de publier, à la suite de ce Manuel pour les producteurs d'examen, une version simplifiée mais de même structure destinée à d'autres partenaires impliqués dans l'enseignement des langues, par exemple aux membres du réseau des Portfolios européens des langues, aux organisateurs de cours et aux enseignants isolés.

Les producteurs d'examens ne sont pas tous tenus d'entreprendre l'étude de tous les domaines mentionnés ci-dessus. Il se peut que certaines institutions, dans des situations où les « enjeux » sont faibles, décident de concentrer leur effort sur la spécification et la standardisation sans pouvoir conduire le processus jusqu'à sa conclusion logique de validation empirique à grande échelle telle que la décrivent les codes et les normes d'évaluation et de mesure reconnus internationalement. Toutefois, on recommande vivement, même aux producteurs d'examen les moins bien lotis, de sélectionner certaines techniques dans chacun des secteurs. La mise en relation d'une qualification en langue avec le CECR sera d'autant plus forte que les hypothèses fondées sur les spécifications des examens et leur contenu seront étayées à la fois par la standardisation des évaluations et par la validation empirique des données. Chaque producteur d'examen – même ceux dont les ressources sont limitées ou dont les pays ont des traditions de décentralisation – devrait être capable de prouver d'une manière ou d'une autre, par un choix de techniques, à la fois la qualité et la validité interne de son examen ainsi que sa validité externe, en d'autres termes, la validité de la relation revendiquée avec le CECR.

Les différents éléments du plan de mise en relation décrits ci-dessus sont présentés dans la Figure 1.1.

FIGURE 1.1: REPRESENTATION GRAPHIQUE DES PROCEDURES PERMETTANT DE RELIER LES EXAMENS AU CECR



1.2. Le besoin d'un Manuel

Depuis le Symposium de Rüslikon en 1991 où furent lancés le *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer (CECR)* et le *Portfolio européen des langues (PEL)*, la demande des Etats membres pour plus de cohérence et de transparence dans les certifications en langues n'a cessé de croître. Le CECR comme le PEL ont apporté une aide considérable dans ce sens. Le CECR a déjà été traduit en 16 langues et d'autres traductions sont en préparation. Presque tous les Etats membres ont conçu et mis en œuvre des versions du PEL. L'Union européenne (dans sa résolution sur Diversité linguistique et apprentissage des langues) a invité les Etats membres à mettre en place des systèmes de validation des compétences en langues basés sur les niveaux du CECR ; son projet DIALANG a pris le CECR comme point de départ. L'Association des centres d'évaluation en langues en Europe (ALTE) a également adopté l'échelle à six niveaux du Conseil de l'Europe.

Le Cadre européen commun de référence sert donc de plus en plus souvent de référence et de base pour la reconnaissance mutuelle des qualifications en langues en Europe. Il occupe déjà cette fonction dans le contexte du Portfolio européen des langues. Le portfolio est un outil éducatif *nouveau* qui a été développé dans le cadre d'une coopération internationale à la fois intensive et extensive. C'est pourquoi sa mise en œuvre s'est faite de façon relativement uniforme, même s'il a connu et s'il connaît encore des contraintes diverses. En revanche, la reconnaissance mutuelle des qualifications en langues décernées par tous les organismes compétents est une affaire beaucoup plus compliquée.

En Europe, les professionnels de l'évaluation en langue ont des traditions très différentes. A une extrémité, 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 niveau 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. Finalement, à l'autre extrémité, 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. Certains des examens en question reposent sur des traditions anciennes. 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 freiner 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 correctes ainsi que des systèmes convenables de formation, d'assurance qualité et de contrôle. Des ressources de ce type peuvent éventuellement être utilisées dans le cas des évaluations nationales de haut niveau qui déterminent l'éligibilité à des postes importants. 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 éducative centrée sur une norme et qui 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 donc pas étonnant qu'à la suite de la publication du CECR on ait souvent fait appel au Conseil de l'Europe pour qu'il joue un rôle plus actif auprès des producteurs d'examens et les aide à situer leurs examens par rapport au Cadre européen de référence et pour qu'il valide – d'une façon ou d'une autre – les examens qui se réclament d'une relation de ce type. Il y va de l'intérêt des producteurs d'examens de langue qu'ils surmontent les difficultés en établissant des liens valides et fiables entre les résultats de leurs systèmes et les niveaux du CECR afin de rendre ces liens transparents pour les utilisateurs de leurs examens de langue.

Ce manuel est la réponse du Conseil de l'Europe au besoin de conseil des producteurs d'examens afin de les aider à relier leurs examens aux CECR. Le premier pas de cette démarche a été fait à l'initiative des autorités finlandaises qui ont organisé un Séminaire à Helsinki du 30 juin au 2 juillet 2002 intitulé « Relier les examens de langue au Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer (CECR) ». C'est à la suite de cette rencontre dont le compte rendu est disponible sous la référence DG IV/EDU/LANG (15) 2002 un groupe de cinq auteurs a été chargé de produire une première ébauche du Manuel. Ils ont été soutenus par un Groupe de consultants ayant participé au Séminaire d'Helsinki. Les auteurs ont également consulté d'autres spécialistes et organismes dans le domaine.

1.3. L'utilisation du Manuel

Ce Manuel n'est pas une succession linéaire d'étapes à suivre mécaniquement. C'est un ensemble de démarches et de techniques régies par des principes que l'on propose pour accompagner les partenaires intéressés dans un processus que l'on sait techniquement compliqué et exigeant. A différentes étapes du processus, des experts doivent être consultés. Le producteur de l'examen en question est responsable de la conception et du processus de mise en relation adéquat et cohérent. Cette responsabilité entraîne :

- une réflexion sur les besoins, les ressources, les compétences et les priorités dans le contexte en question ;
- un choix des démarches appropriées parmi celles qui sont exposées – ou d'autres extraites d'ouvrages sur le sujet ;
- une planification réaliste du projet comportant une approche modulaire par degrés qui garantisse des résultats ;
- l'établissement d'un réseau de collaboration avec des partenaires d'autres secteurs et d'autres pays ;
- la coordination des participants au processus local de mise en relation ;
- l'application fidèle des démarches ;
- un enregistrement fiable des résultats ;
- un compte rendu exact des conclusions.

Ce Manuel est une contribution à l'effort collectif pour améliorer la transparence et la comparabilité des qualifications en langues en Europe. Il s'agit de l'une des nombreuses avancées faites dans ce sens à l'heure actuelle. Cette première édition est un texte expérimental. L'expérience de l'application du Manuel dans la pratique fournira un retour d'informations qui permettront de le perfectionner. Les études de cas documentés viendront en annexe de la seconde édition.

La suite de cette introduction donne un bref panorama des autres Chapitres.

1.4. Ebauche du plan de mise en relation

Les démarches préconisées pour relier les examens de langues au CECR se répartissent entre les domaines de la Spécification, de la Standardisation et de la Validation empirique. En amont de la Spécification et de la Standardisation se trouve la Familiarisation. Ces domaines sont présentés dans un ordre logique. Dans de nombreux cas, les techniques impliquées suivent la même logique. Des recommandations sont faites à ce sujet.

1.4.1. Spécification (Chapitre 4)

La Spécification implique un relevé de ce que recouvre l'examen en fonction des catégories et des niveaux du CECR. Le CECR fournit un ensemble très complet de niveaux et de catégories que l'on peut utiliser pour décrire le contenu de l'examen. Les listes de contrôle permettent de détailler les objectifs et le contenu de l'examen ainsi que les documents qui l'accompagnent. Des explications viennent les compléter.

Familiarisation avec le CECR

Avant de se lancer dans la description de l'examen et de tracer les grandes lignes de ce qu'il recouvre par rapport aux catégories et aux niveaux du CECR, l'équipe de travail doit se familiariser avec le CECR. Pour ce faire, les étapes suivantes (décrites en détail dans le chapitre 3) peuvent être utiles.

- a) Choix des encadrés qui apparaissent à la fin des parties appropriées du CECR.
- b) Examen de l'ensemble des niveaux du CECR.
- c) Auto-évaluation de son propre niveau de langue dans une langue étrangère.
- d) Tri, par niveau, des descripteurs d'une échelle du CECR.

Si, avant la mise en relation, on effectue une traduction des descripteurs du CECR, ou du CECR dans son ensemble, les techniques décrites plus haut peuvent être utilisées pour vérifier que la traduction reflète effectivement les descripteurs d'origine en termes de niveau.

Description du contenu

Le processus de Spécification exige deux formes différentes de description :

- D'abord la description de l'examen en question en tant que tel.
- Ensuite, une analyse du contenu de l'examen par rapport aux catégories et niveaux du CECR.

Description : elle doit préciser les objectifs de l'examen, les besoins de la population d'apprenants à qui il est destiné, la composition des épreuves et des différentes parties, les coefficients et leur justification, les types de textes, d'items et de tâches, les critères de notation y compris un barème détaillé avec des exemples de réponses possibles et la façon de rendre compte des résultats à l'apprenant. Cette description peut également annoncer comment les objectifs, le contenu et les coefficients sont périodiquement révisés.

Analyse du contenu : la référence principale de cette analyse est le CECR. Des éléments de référence complémentaires sont fournis par l'ensemble des spécifications de contenus liés au CECR et élaborés en association avec le Conseil de l'Europe ; pour l'anglais : A1 : Breakthrough¹ ; A2 : Waystage² ; B1 : Threshold Level³ ; B2 : Vantage Level⁴ ; pour l'allemand : Profile Deutsch⁵ ; pour le français : « Un référentiel pour le français »⁶. Outre une vue d'ensemble, l'analyse du contenu devrait aussi donner une description détaillée de la façon dont les tâches, pour chaque savoir faire, et les critères de notation, pour les savoir faire en production, sont reliés aux descripteurs du CECR.

¹ Breakthrough : à paraître.

² van EK, J.A. ; Trim, J.L.M., (2001), Waystage, Cambridge, CUP, ISBN 0-521-56707-6.

³ van EK, J.A. ; Trim, J.L.M., (2001), Threshold 1990, Cambridge, CUP, ISBN 0-521-56707-8.

⁴ van EK, J.A. ; Trim, J.L.M., (2001), Vantage, Cambridge, CUP, ISBN 0-521-56707-X.

⁵ Glaboniat, M. ; Müller, M. ; Schmitz, H. ; Rusch, P. ; Wertensclag, L., (2002), Profile Deutsch, Berlin, Langenscheidt, ISBN 3-468-49463-7.

⁶ A paraître en français, l'équivalent de Profile Deutsch.

Si l'examen aborde des domaines qui ne sont pas couverts par le CECR ou ne le sont que partiellement (la traduction, par exemple), il faudra le signaler.

Résultats :

Le résultat de la **Spécification** est une hypothèse de mise en relation de l'examen avec le CECR sur la base de son contenu et de ce qu'il recouvre. Le résultat présente :

- Un tableau visualisant ce que l'examen recouvre en termes de niveaux et de catégories du CECR.
- La partie du compte rendu à l'usage des praticiens en langues étrangères, comprenant une description ainsi que les formulaires pertinents complétés.

Si l'on a entrepris une traduction des descripteurs du CECR avant le projet de mise en relation, il faudrait alors, dans le compte rendu descriptif, apporter la preuve de l'équivalence de ces traductions avec les descripteurs d'origine. La preuve peut se faire par des activités de tri effectuées au cours d'ateliers (CECR Annexe A, Méthode n° 9 Tri de tâches) ou par une nouvelle traduction vers la langue d'origine.

1.4.2. Standardisation (Chapitre 5)

La Standardisation suppose que les membres de l'équipe aient une compréhension commune de la signification des niveaux du CECR. Il est indispensable de confronter cette compréhension à l'interprétation qu'en font les praticiens des langues étrangères dans d'autres secteurs, régions ou pays. Dans la pratique, cela suppose que les évaluations des experts impliqués dans le projet de mise en relation soient reconnues et validées.

Ce Manuel met l'accent sur les grandes lignes de la formation des correcteurs et sur les techniques d'harmonisation qui utilisent des modèles standards du CECR afin de :

- a) former les praticiens des langues étrangères à interpréter les niveaux du CECR de manière homogène ;
- b) élaborer localement des modèles à la fois pour les performances de l'apprenant et pour des types de tâches à des niveaux différents.

On peut utiliser, dans le contexte cible, à la fois les modèles standards du CECR et les modèles locaux mentionnés ci-dessus pour une formation ultérieure des examinateurs et des enseignants locaux à la standardisation.

Les lignes directrices détaillées au Chapitre 5 exposent les techniques de standardisation parmi lesquelles les utilisateurs peuvent choisir. Ces techniques sont accompagnées d'une indication sur le temps et les ressources exigées.

La première étape de la standardisation exige une définition de la structure du projet, des rôles de chacun, des objectifs, des procédures, des phases et des produits. Dans ce processus, il faudra distinguer entre (a) les démarches relatives aux savoir faire interactifs et productifs notés par les examinateurs et, (b) les démarches relatives à la réception ainsi que les aspects des compétences de l'apprenant qui sont évaluées sur des items.

Si ce processus est réalisable dans la situation donnée, il est recommandé aux producteurs d'examen de commencer par les savoir faire interactifs et de production puisque la performance dans les tâches d'interaction et de production peut être directement comparée aux échantillons de performance standards du CECR et notée directement en relation aux descripteurs du CECR. C'est une démarche logique puisqu'elle garantit que, très tôt dans le processus de mise en relation, la population

d'apprenants sera mesurée par rapport aux Niveaux communs de référence. Les personnes impliquées dans le processus peuvent alors confirmer leur interprétation des Niveaux communs de référence.

Le processus de Standardisation devrait alors suivre un déroulement logique, en partant de la **Familiarisation**, puis en passant par la **Formation**, jusqu'à l'**Etalonnage** des échantillons locaux de performance et la **Définition de standards** pour les examens locaux.

Familiarisation

Toutefois, que l'on travaille sur des échantillons de performances ou sur des items, il est impératif de commencer par les exercices de familiarisation ainsi que le recommande la Spécification ci-dessus :

- a) Choix des encadrés qui apparaissent à la fin des chapitres appropriés du CECR.
- b) Examen de l'ensemble des niveaux du CECR.
- c) Auto-évaluation de son propre niveau de langue dans une langue étrangère.
- d) Tri, par niveau, des descripteurs d'une échelle du CECR.

Formation à l'aide d'échantillons standards de performances pour les compétences de production :

- Formation à l'aide d'échantillons standards du CECR en Interaction orale et Production (vidéos⁷) et/ou en Interaction écrite et Production (textes en cours de production⁸).
- Examen du niveau des apprenants en relation avec la grille d'évaluation de la performance orale (CECR Tableau 3 / Tableau 5.5⁹) et de la performance écrite (Tableau 5.8).
- Notation individuelle suivie d'une analyse des notes proposées puis d'un débat comprenant des approximations successives jusqu'à ce qu'un consensus convenable soit atteint.

Formation à l'aide de tâches standards testant les capacités de réception et les compétences langagières :

- Formation à l'aide d'items types qui rendent opérationnel un descripteur du CECR et dont l'analyse de données a montré qu'ils étaient ciblés précisément au niveau en question¹⁰. Débat sur les caractéristiques d'items de ce type pour les niveaux adéquats du CECR.
- Analyse des niveaux des textes et des tâches dans ces items « prototypes » qui rendent opérationnels les descripteurs du CECR. Examen de l'étendue de la difficulté – à tous les niveaux – de ces prototypes et justification de cette difficulté.

Etalonnage des échantillons locaux de performance(s) :

- Notation individuelle d'échantillons locaux de performances, à l'aide des mêmes outils d'évaluation (Production orale : CECR Tableau 3 / Manuel, Tableau 5.5 ; Production écrite : Tableau 5.8), suivie d'un débat et du consensus du groupe/application.

⁷ Echantillons standardisés par rapport au CECR de production orale en continu et d'interaction. Une vidéo avec des performances d'apprenants suisses adultes en langue anglaise est disponible à la Division des Politiques linguistiques.

⁸ Echantillons standardisés par rapport au CECR de production écrite et d'interaction. Travail en cours en collaboration avec les producteurs d'examens et de cours.

⁹ Tableau 3 : Niveaux communs de référence – Aspects qualitatifs de l'utilisation de la langue parlée.

¹⁰ Le Gouvernement néerlandais lance en 2003 un projet international de construction d'une banque d'items pour une compétence donnée étalonnés sur le CECR. On pourrait alors utiliser ces items pour guider les évaluations dans la définition des normes et pour ancrer les tests à l'échelle du CECR.

- Analyse de la fourchette de notation et débat jusqu'au consensus. Analyse statistique des notes avec, par exemple, le modèle multiple de Rasch si les données sont suffisantes. Ce modèle repère le comportement irrégulier d'un correcteur et rétablit les notes en termes de sévérité ou d'indulgence.

Etablissement de standards pour la réception et les compétences langagières :

- Identification des items locaux qui conviennent au niveau du CECR en question et peuvent ainsi servir de prototypes.
- Notation individuelle d'items suivie de débat.
- Analyse des notes et débat avec approximations successives jusqu'à ce qu'on arrive à un consensus.
- Hypothèse provisoire sur la difficulté des items.
- Retour d'information sur la difficulté réelle après expérimentation.
- Examen des caractéristiques des items mal classés (identification d'éléments ou de distracteurs mal ciblés).
- Examen des caractéristiques des items types.
- Confirmation des items types et de leur niveau de difficulté.

Résultats :

La **Standardisation** débouche sur un renforcement de l'hypothèse faite sur la base de la **Spécification** avec des exemples et des résultats documentés.

Cela constitue la partie Standardisation du rapport destiné aux professionnels des langues. Le rapport est complété par des annexes présentant des exemples et des modèles de tâches, des échantillons locaux de performances orale et écrite ainsi que des démarches de correction et de notation. Le rapport devait montrer avec ces exemples comment ont été exploités les échantillons de performances standards pour étalonner les échantillons locaux sur le CECR et mettre en place des systèmes d'harmonisation.

1.4.3. Validation empirique (Chapitre 6)

La validation empirique suppose que l'on recueille et analyse des données sur (a) la réalisation d'une tâche et les caractéristiques d'un item, (b) la qualité du test, (c) la performance de l'apprenant et les résultats du test, (d) le comportement des correcteurs et (e) l'adéquation des normes utilisées. Cet aspect peut être limité ou largement développé selon le type de contexte éducatif et d'évaluation et les ressources disponibles.

La Validation empirique sera vraisemblablement engagée à deux niveaux parfois de manière simultanée :

Validation interne

La reconnaissance de la qualité de l'examen en tant que tel est en fait un préalable à la mise en relation effective. Les items jouent-ils le rôle qu'on attend d'eux ? La notation est-elle appropriée ? Y a-t-il équivalence entre différentes formes du (ou des) test(s) ?

- a) Expérimentation d'items et d'examens pilotes afin de supprimer ou de modifier les items hors normes, d'améliorer les hypothèses sur la difficulté des items et le barème de notation et d'analyser les caractéristiques et la qualité de l'examen dans son ensemble.

- b) Confirmation que les items ont les caractéristiques d'items de bonne qualité et des indices de difficulté adéquats.
- c) Confirmation que l'examen évalue ce qu'il prétend évaluer.
- d) Calibrage des items et des tâches dans l'examen.

Validation externe

On peut parvenir à corréler les standards de l'examen essentiellement de deux façons : par l'utilisation d'un test d'ancrage ou en s'appuyant sur l'opinion d'enseignants bien au fait du CECR – en passant par les procédures de Familiarisation et de Standardisation. Les suggestions ci-dessous sont classées approximativement par ordre de complexité en termes de conception du projet et d'analyse des données.

- a) Etablir la corrélation des résultats au test à des examens qui ont déjà été calibrés sur le CECR de façon valide et fiable et utiliser l'analyse de régression pour relier les résultats aux niveaux du CECR.
- b) Etablir la corrélation comme ci-dessus et utiliser ensuite l'examen critérié comme test d'ancrage afin d'étalonner l'examen étudié, et établir la même corrélation pour les tâches et les items qu'il propose avec l'échelle utilisée par l'examen critérié – et ainsi avec les niveaux du CECR.
- c) Etablir la corrélation entre la notation des enseignants et les descripteurs des « capacités de faire (Can-Do) » déjà calibrés sur le CECR de façon valide et fiable puis utiliser l'analyse de régression pour relier les résultats à l'examen aux niveaux du CECR.
- d) Etablir la corrélation comme ci-dessus et utiliser ensuite les valeurs étalonnées des descripteurs du CECR comme ancrage pour calibrer l'examen en question directement sur les échelles qui sous-tendent les niveaux du CECR.

1.4.3.4. Résultat

La **Validation empirique** débouche sur la confirmation à des degrés divers de l'hypothèse avancée après **Spécification** et **Standardisation** en se fondant sur l'analyse des données de l'examen. Le degré de confirmation dépendra de l'adéquation, de la portée et de la qualité du travail effectué.

Les résultats se présentent sous la forme :

- De représentations graphiques de la nature et de la solidité de la relation empirique de l'examen avec le CECR. Ces représentations sont normalisées.
- D'un compte rendu descriptif qui apporte des informations sur des aspects ayant trait à :
 - la forme des outils
 - la conception du recueil de données
 - le nombre de sujets
 - les méthodes d'analyse
 - les problèmes rencontrés et les solutions retenues
 - les résultats obtenus
 - l'interprétation des résultats

1.5. Guide pour la rédaction des rapports (Chapitre 7)

La mise en relation d'un examen avec le CECR ne saurait être dissociée de celle du descriptif du contenu et du type de l'examen. De même, on ne peut dissocier le compte rendu de la mise en relation elle-même de celui du processus suivi pour la réaliser. Cette partie comprend donc des consignes détaillées sur la façon de rédiger les deux rapports :

- **Rapport destiné au grand public :**

Il est principalement fondé sur la Spécification mais il présente des informations appropriées sur l'étendue et les conclusions de la Standardisation et de la Validation empirique. C'est un rapport très court conforme à la présentation standard fournie en Annexe. On recommande aux titulaires d'un Portfolio européen des langues de joindre à leur Dossier un exemplaire de ce rapport pour les examens qu'ils ont passés. On espère pouvoir, à terme, rendre ces rapports disponibles sur le site de la Division des Politiques linguistiques du Conseil de l'Europe ; ils témoigneront de l'authenticité de la mise en relation.

- **Rapport destiné aux professionnels :**

Il s'agit du résumé des résultats des procédures mises en œuvre pour la Spécification, la Standardisation et la Validation empirique.

Le rapport commence par une représentation graphique :

- Du niveau du CECR auquel l'examen se situe ;
- Du contenu testé schématisé, présentant la relation avec les catégories et les niveaux du CECR. Quelles sont les catégories du CECR testées ? Exige-t-on le même niveau du CECR pour toutes les catégories ou bien la norme est-elle fixée plus haut pour certaines catégories que pour d'autres ? Certains examens visent à tester un niveau de compétence équivalent dans les différentes capacités savoir faire. D'autres cherchent à encourager un niveau plus élevé dans certaines catégories (par exemple en réception) ou autorisent un niveau plus faible dans d'autres (par exemple, en production écrite). Il faut que des caractéristiques de ce type apparaissent clairement sur le schéma graphique ;
- De l'ensemble des procédures suivies en utilisant une liste de contrôle informant sur le niveau d'achèvement du processus (niveau élevé / couverture convenable / perfectionnement nécessaire) ;
- De la solidité de la corrélation avec le critère externe utilisé pour corroborer la relation au CECR ;
- De la solidité de la classification en niveaux du CECR (*Pouvoir décisionnel*).

Il est préférable de fournir ces deux derniers graphiques pour chaque épreuve de l'examen.

Le corps du rapport détaille les démarches entreprises ainsi qu'on l'a souligné sous dans « Résultats » (grandes lignes du plan de mise en relation présenté ci-dessus). Les rapports sur la Spécification et la Standardisation doivent inclure les fiches standards des Chapitres 4 et 5. On les complètera d'un historique rédigé dans un style accessible au plus grand nombre : professionnels de l'enseignement des langues et examinateurs. Le rapport de la Validation empirique devrait apporter des informations et des références suffisantes pour permettre une analyse critique des démarches et des méthodologies mises en œuvre. Dans ce cas également, nous espérons pouvoir rendre ces rapports accessibles en téléchargement.

Chapitre 2

Le Cadre européen commun de référence

Le « Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer » (CECR) fournit 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 savoir-faire qu'ils doivent acquérir afin d'avoir un comportement langagier efficace.
- Il définit les niveaux de compétence qui permettent de mesurer le progrès de l'apprenant à chaque étape de l'apprentissage et à tout moment de la vie.

Le CECR encourage une « approche actionnelle » de l'apprentissage des langues. Les descriptions de compétence des niveaux du CECR sont formulées en termes de ce que les apprenants peuvent faire avec la langue à un niveau donné. Cette centration sur la réalisation relève d'un choix philosophique fondamental : seules des formulations positives de niveaux de compétence sont utilisées pour définir les objectifs. On a en effet constaté que les formulations négatives aux niveaux inférieurs entraînaient des interprétations peu cohérentes.

Le CECR a été élaboré de 1993 à 1996 par un Groupe de travail international et un Groupe d'auteurs¹¹ plus réduit suivant la recommandation du Symposium intergouvernemental de Rüşchlikon (Suisse, novembre 1991) sur le thème « Transparence et cohérence dans l'apprentissage des langues en Europe : objectifs, évaluation, certification ». Le Symposium a adopté les conclusions suivantes :

« Il est souhaitable d'élaborer un Cadre européen commun de référence pour l'apprentissage des langues à tous les niveaux, dans le but :

- de promouvoir et de faciliter la coopération entre les établissements d'enseignement de différents pays ;
- d'asseoir sur une bonne base la reconnaissance réciproque des qualifications en langues ;
- d'aider les apprenants, les enseignants, les concepteurs de cours, les organismes de certification et les administrateurs de l'enseignement à situer et à coordonner leurs efforts ; ... »

L'échelle des niveaux définis dans le CECR (« Niveaux communs de référence ») a été élaborée de 1993 à 1996 dans le cadre d'un projet du Fonds national suisse de recherche scientifique mis en place à la suite du Symposium de Rüşchlikon¹².

Un sondage à grande échelle a montré que les premières versions du CECR avaient reçu un accueil extrêmement favorable. La version révisée qui tenait compte d'un retour d'information à la fois formel et informel a été publiée en 2001 pour l'anglais (Cambridge University Press), le français (Didier), l'allemand (Langenscheidt) et le portugais (ASA) puis en 2002 pour le tchèque (Univerzita Palachého v Olomuci), le hongrois (Pedagogus-tvabbképzési Modszernati és Információs Központ), l'italien (La Nuova Italia – Oxford) et l'espagnol (Ministerio de Educacion, Cultura y Deporte et grupo ANAYA).

¹¹ John Trim, Daniel Coste, Brian North, Joseph Sheils.

¹² Schneider et North 1999 : „In anderen Sprachen kann ich” ...Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit. Berne, Project Report, Programme national de recherche 33, Centre suisse de coordination pour la recherche en éducation. / Schneider, G. et North, B. 2000 : „Dans d'autres langues je suis capable de ...” Echelles pour la description, l'évaluation et l'auto-évaluation des compétences en langues étrangères. Rapport de valorisation, Berne / Aarau : Programme national de recherche 33 (PNR 33), Centre suisse de coordination pour la recherche en éducation (CSRE).

Les représentants de tous les Etats membres du Conseil de l'Europe ont officiellement approuvé le CECR en avril 1997 lors de la conférence finale du projet à moyen terme sur « Apprentissage des langues et citoyenneté européenne ». L'Union européenne, (Résolution du Conseil européen sur la promotion de la diversité linguistique et de l'apprentissage des langues), invite les Etats membres « à mettre en place des systèmes de compétence en langues fondés sur le Cadre européen commun de référence pour les langues conçu par le Conseil de l'Europe en tenant suffisamment compte des savoir-faire acquis par un apprentissage non institutionnel ».

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*.

Avec les catégories du *Schéma descriptif* et les *Niveaux communs de référence*, le CECR est conçu de façon à prendre en compte les diversités locales et à permettre aux producteurs d'examens de construire leurs épreuves. Il existe des échelles pour de nombreux paramètres de la compétence langagière. Ils sont détaillés dans les Chapitres 4 et 5. Ces échelles permettent de définir des profils différents pour des apprenants ou des groupes d'apprenants particuliers.

2.1. Qu'est-ce que le Schéma descriptif ?

Dans la Partie 2.1 du CECR on le résume comme suit :

L'usage d'une langue, y compris son apprentissage, comprend les actions accomplies par des gens qui, comme individus et comme acteurs sociaux, développent un ensemble de **compétences générales** et, notamment une **compétence à communiquer langagièrement**. Ils mettent en oeuvre les compétences dont ils disposent dans des **contextes** et des **conditions** variés et en se pliant à différentes **contraintes** afin de réaliser des **activités langagières** permettant de traiter (en réception et en production) des **textes** portant sur des **thèmes** à l'intérieur de **domaines** particuliers, en mobilisant les **stratégies** qui paraissent le mieux convenir à l'accomplissement des **tâches** à effectuer. Le contrôle de ces activités par les interlocuteurs conduit au renforcement ou à la modification des compétences.

En ce qui concerne la conception et/ou la description des examens et évaluations en langues, les points clés du descriptif ci-dessus sont les différentes composantes de la *compétence communicative langagière*, les *activités langagières* et les *domaines*.

2.1.1. Compétence communicative langagière

(D'après la Partie 2.1.2 du CECR)

La compétence communicative présente plusieurs composantes : une *composante linguistique*, une *composante sociolinguistique*, une *composante pragmatique*. Chacune de ces composantes est constituée notamment de savoirs, de capacités et de savoir-faire.

La *compétence linguistique* est celle qui a trait aux savoirs et savoir-faire relatifs au lexique, à la phonétique, à la syntaxe et aux autres dimensions du système d'une langue, pris en tant que tel, indépendamment de la valeur sociolinguistique de ses variations et des fonctions pragmatiques de ses réalisations. Cette composante, considérée ici sous l'angle de la compétence communicative d'un individu donné, se rapporte non seulement à l'étendue et à la qualité des connaissances (par exemple en termes de distinctions phonétiques établies ou d'étendue et de précision du lexique), mais aussi à l'organisation cognitive et au mode de stockage mémoriel de ces connaissances (par exemple, les réseaux associatifs divers dans lesquels un élément lexical peut se trouver inclus pour ce locuteur) ainsi qu'à la façon dont cet individu peut mobiliser ses connaissances (activation, rappel et

disponibilité). Leur organisation et leur accessibilité varient d'un individu à l'autre et, pour un même individu, connaissent aussi des variations.

La **compétence sociolinguistique** renvoie aux paramètres socioculturels de l'utilisation de la langue. Sensible aux normes sociales (règles d'adresse et de politesse, régulation des rapports entre générations, sexes, statuts, groupes sociaux, codification par le langage de nombre de rituels fondamentaux dans le fonctionnement d'une communauté), la composante sociolinguistique affecte fortement toute communication langagière entre représentants de cultures différentes, même si c'est souvent à l'insu des participants eux-mêmes.

La **compétence pragmatique** recouvre l'utilisation fonctionnelle des ressources de la langue (réalisation de fonctions langagières, d'actes de parole) en s'appuyant sur des scénarios ou des scripts d'échanges interactionnels. Elle renvoie également à la maîtrise du discours, à sa cohésion et à sa cohérence, au repérage des types et genres textuels, des effets d'ironie, de parodie. Plus encore pour cette composante que pour la composante linguistique, il n'est guère besoin d'insister sur les incidences fortes des interactions et des environnements culturels dans lesquels s'inscrit la construction de telles capacités.

2.1.2. Activités langagières

(D'après la Partie 2.1.3 du CECR)

La compétence communicative du sujet apprenant et communiquant est mise en oeuvre dans la réalisation d'**activités langagières** variées pouvant relever de la **réception**, de la **production**, de l'**interaction** et de la **médiation** (notamment les activités de traduction et d'interprétation), chacun de ces modes d'activité étant susceptible de s'accomplir soit à l'oral, soit à l'écrit, soit à l'oral et à l'écrit.

Les activités langagières de **réception** (orale et/ou écrite) ou de **production** (orale et/ou écrite) sont évidemment essentielles. Les premières (activités de réception) sont en relation avec la lecture, la fréquentation des médias et tiennent également une grande place dans bien des formes d'apprentissage. Les secondes (activités de production) ont une fonction importante dans nombre de secteurs éducatifs et professionnels (par exemple, présentations et exposés oraux, études et rapports écrits).

Dans l'**interaction**, au moins deux acteurs participent à un échange oral et/ou écrit et alternent des moments de production et de réception qui peuvent même se chevaucher. Même lorsque les tours de parole sont strictement respectés, l'auditeur est généralement en train d'anticiper sur la suite du message et de préparer une réponse. Ainsi, apprendre à interagir implique plus d'éléments qu'apprendre à recevoir et à produire des énoncés.

La **médiation** rend la communication possible entre des acteurs qui, pour une raison ou une autre sont incapables de communiquer directement. La traduction ou l'interprétariat, la paraphrase, le résumé ou le compte rendu fournissent, à l'intention d'un tiers, la (re)formulation accessible d'un texte premier auquel ce tiers n'a pas un accès direct.

2.1.3. Domaines

(D'après la Partie 2.1.4 du CECR)

Les activités langagières s'inscrivent à l'intérieur de **domaines**. Ils peuvent être très divers mais, pour des raisons essentiellement pratiques, en ce qui concerne l'apprentissage des langues, on les classe en quatre secteurs majeurs : le *domaine public*, le *domaine professionnel*, le *domaine éducationnel* et le *domaine personnel*.

Le *domaine public*, se réfère à tout ce qui relève des échanges sociaux ordinaires (relations commerciales et civiles, services publics, activités culturelles et de loisir dans des lieux publics, relations aux médias, etc.). De façon complémentaire, le *domaine personnel* est caractérisé par les relations familiales et par les pratiques sociales individuelles.

Le *domaine professionnel* recouvre tout ce qui concerne les interventions et les relations des individus dans l'exercice de leur activité professionnelle. Le *domaine éducationnel* est celui où l'acteur se trouve dans un contexte (le plus souvent institutionnel) de formation et est censé y acquérir des connaissances ou des savoir-faire définis.

La répartition des activités communicatives exposée dans la partie précédente permet l'élaboration et l'organisation d'échelles de descripteurs pertinents pour les types d'activités des domaines *professionnel* et *éducatif*. Les décideurs en matière de politique linguistique manifestent de plus en plus d'intérêt pour ces deux derniers domaines.

2.2. En quoi consistent les Niveaux communs de référence ?

Les Niveaux communs de référence proposent un ensemble de 6 niveaux de critères définis pour fournir des standards communs (A1, A2, B1, B2, C1, C2). Ces standards communs ont pour but d'aider les producteurs de cours et d'examens à confronter leurs produits à un système de référence commun et de les mettre ainsi, indirectement, en relation les uns avec les autres.

Concernant les Niveaux communs de référence, deux remarques essentielles doivent être faites. La première c'est qu'ils sont fondés sur la recherche¹³ et non sur des hypothèses. La seconde c'est que ce ne sont pas les seuls niveaux possibles mais seulement les plus clairement reconnaissables. Il existe une grande variété d'usages de la langue et de types de communication au-delà de C2 ainsi qu'une utilisation limitée de la langue et de la communication en deçà de A1 et il est possible de définir des niveaux entre A2 et B1, B1 et B2 et B2 et C1.

2.2.1. Elaboration

Les Niveaux communs de référence et les descripteurs qui les définissent ont été produits au cours d'un projet de recherche qui a duré trois ans dans un contexte multisectoriel et multilingue. On trouve une description simple mais détaillée du processus (North 2002) dans la brochure Etudes de cas du CECR (Conseil de l'Europe 2002) ainsi qu'un très bref résumé dans l'Annexe B du CECR. Un court compte rendu technique en est donné par North et Schneider (1998), une documentation complète sur la méthodologie par North (2000) et un rapport complet sur le projet (essentiellement en allemand) par Schneider et North (2000). Dans le projet, des descripteurs soigneusement sélectionnés ont été échelonnés sur la base de l'interprétation que des professeurs en ont donnée quand ils les ont utilisés pour évaluer la performance d'apprenants réels. Au cours du processus analytique, on a écarté les descripteurs qui donnaient lieu à une interprétation peu cohérente. La façon dont des descripteurs différents qui décrivaient un contenu similaire ont été étalonnés sur l'échelle a manifesté un niveau élevé de cohérence.

La méthodologie s'est organisée en cinq étapes :

- **L'étape intuitive** : analyse d'échelles de compétence en langue empruntées au domaine public et classement de leurs éléments selon les catégories du Schéma descriptif des Chapitres 4 et 5 du CECR.
- **L'étape qualitative** : vérification que le métalangage était bien celui employé par les enseignants et que l'interprétation des descripteurs correspondait bien à la catégorie et au niveau voulus ; environ 32 ateliers de travail avec des enseignants ont eu lieu pour ce faire.

¹³ North, B. 2000 : *The development of a common framework scale of language proficiency*. New York, Peter Lang.

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

Schneider, G. et North, B. 2000 : *Fremdsprachen können – was heisst das ? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit*. Chur/Zürich.

- **L'étape quantitative** : évaluation, par les enseignants, d'élèves types à la fin d'une année scolaire, en utilisant des questionnaires comprenant les meilleurs descripteurs ; évaluation, par tous les enseignants participant aux ateliers, de vidéos de certains des apprenants du sondage afin de repérer la subjectivité et d'y remédier ; placement des descripteurs et des apprenants sur la même échelle arithmétique en utilisant le modèle d'analyse de Rasch.
- **L'étape interprétative** : analyse de **bandes de niveau cohérentes** sur l'échelle des descripteurs afin de produire les Niveaux communs de référence ; résumé de ces niveaux sur une échelle globale (CECR Tableau 1), une grille d'auto-évaluation où sont décrites les activités langagières (CECR Tableau 2) et une grille d'évaluation de la performance décrivant des aspects **différents** de la compétence communicative langagière (CECR Tableau 5.5). Intégration des échelles pour différents aspects du Schéma descriptif dans les Chapitres 4 et 5.
- **L'étape de reproduction** : reproduction du projet, cette fois pour le français et l'allemand avec des indices d'étalonnage ayant une **corrélation** de 0.99 avec le projet original ; validation dans le projet DIALANG (corrélation 0.90), les projets de l'Université de Bâle (corrélation 0.90), des Eurocentres (corrélation 0.88) et de ALTE (corrélation 0.97). On trouvera tous les détails dans la brochure Etudes de cas du CECR (Conseil de l'Europe 2002).

2.2.2. Points communs de référence

Le choix de 6 Niveaux communs de référence plutôt que de 9 ou de 12 relève en partie d'un arbitraire et en partie d'une réflexion sur la façon extrêmement cohérente selon laquelle le contenu (certaines tâches, situations et capacités) se regroupait sur l'échelle. L'intention était de produire une échelle de niveaux basée sur le cadre de référence fourni par les spécifications du Conseil de l'Europe déjà existantes pour le Waystage et le Niveau seuil ainsi que Wilkins l'avait suggéré (1998). Le Niveau seuil et les spécifications annexes ont eu une importance considérable sur l'élaboration des manuels et des examens partout dans le monde. Le Niveau seuil a eu également une grande influence sur les auteurs d'échelles de compétence en langue et il n'était donc pas surprenant que l'essentiel du contenu des descripteurs empruntés aux échelles de compétence en langue existantes soit en adéquation avec le contenu des spécifications du Waystage et du Niveau seuil. Les Niveaux communs de référence fournissent donc un échelonnage consensuel et objectif de l'interprétation, par des enseignants, de formulations en rapport avec les spécifications produites par le Conseil de l'Europe dans les années 70 et que reflètent largement depuis les manuels et les examens. Le niveau élevé de confirmation de l'ordre de l'étalonnage atteint par les différentes reproductions entreprises dans des contextes différents et mentionnées à la fin de la partie précédente permet de penser que ce « cadre commun de référence » est partagé largement par les professionnels des langues vivantes en Europe. Il est important de se rendre compte que si les 6 niveaux sont significatifs en tant qu'éléments de référence, ce ne sont pas les seuls niveaux possibles. En fait, l'analyse qui les sous-tend a débouché sur une échelle de 9 seuils de taille à peu près égale comme on le verra dans le Tableau 2.1 qui suit.

Il peut être opportun de subdiviser les niveaux afin de planifier des modules et de montrer aux apprenants les progrès qu'ils font. Le Chapitre 3 du CECR suggère que les institutions conçoivent des « approches arborescentes ». Le premier exemple donné dans la Partie 3.4 du CECR concerne l'école primaire ou le premier cycle du secondaire. La figure 2.1. montre comment il est possible de relier un ensemble de 6 mini-niveaux aux Niveaux communs de référence :

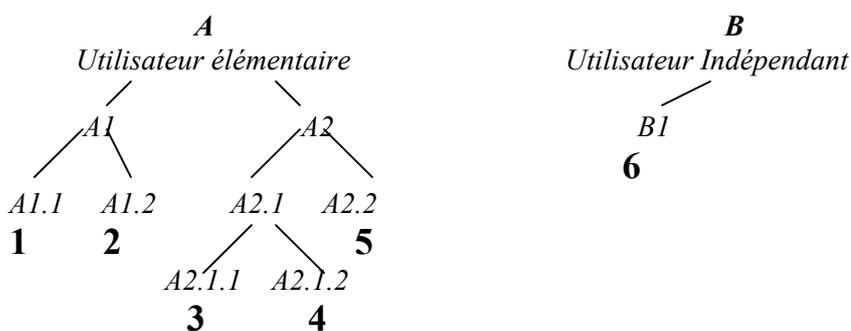


Figure 2.1 : Approche arborescente des niveaux du CECR

En ce qui concerne la conception de programmes d'enseignement il est possible, par exemple, d'utiliser les six mini-niveaux pour planifier un programme sur trois ans avec un palier par semestre.

En ce qui concerne l'évaluation, la Figure 2.1 peut montrer comment les degrés d'un examen scolaire visant le niveau A2 sont en relation avec les Niveaux communs de référence. Quelques apprenants particulièrement doués peuvent se trouver déjà en B1 (degré 6). D'autres n'atteindront pas l'ensemble des standards (degrés 1 et 2). Parmi ceux qui les atteignent, il serait souhaitable de faire la distinction entre une performance acceptable mais limite (degré 3), une bonne performance (degré 4) et une performance nettement au-dessus du niveau attendu (degré 5).

Il n'y a aucune raison pour qu'un producteur d'examen se réfère à 6 niveaux plutôt qu'à 4 ou 3. Certains établissements d'enseignement ou d'évaluation souhaiteront diviser les niveaux de manière distincte dans des situations différentes et, en conséquence, il est logique de limiter les points communs de référence à six niveaux plus larges et plus facilement reconnaissables.

2.2.3. Caractéristiques des Niveaux communs de référence

Lorsqu'on examine les niveaux du CECR, il est intéressant de remarquer que la Partie 3.6. du CECR « Cohérence du contenu des Niveaux communs de référence » se révèle en fait plus utile que le Tableau 1 « Niveaux communs de compétences – Echelle globale » parce qu'elle donne des détails très utiles pour « se faire une idée » des niveaux.

C'est pour cette raison qu'on a juxtaposé un montage de la Partie 3.5. à la formulation du Tableau 1 du CECR (tableau 2.1). Ce commentaire montre à quel point les Niveaux communs de référence donnent une image cohérente du développement de la compétence langagière.. La mise en évidence de l'existence de seuils de compétence entre les Niveaux communs de référence, au-delà et en deçà est voulue dans le contexte actuel. L'existence de seuils de ce type peut être pertinente pour la mise en relation de données de tests avec les six niveaux.

TABLEAU 2.1 : LES NIVEAUX COMMUNS DE REFERENCE

Niveau	Echelle globale (Tableau 1)	Caractéristiques principales (CECR, Partie 3.5 simplifiée)
Utilisateur Expérimenté		<p>On ne saurait trop insister sur le fait que le Niveau C2 n'a pas l'ambition d'égaliser la compétence du locuteur natif ou quasi-natif. La recherche initiale ainsi qu'un projet ayant utilisé les descripteurs du CECR pour évaluer la langue maternelle et la compétence en langue étrangère (North REF) ont démontré l'existence de locuteurs bilingues bien au-dessus du niveau le plus élevé défini (C2). Wilkins avait repéré un septième niveau de « Compétence bilingue » dans sa proposition, en 1978, d'une Echelle européenne pour des plans d'unité/crédit (REF)</p>
	<p>C 2 Peut comprendre sans effort pratiquement tout ce qu'il/elle lit ou entend. Peut résumer faits et arguments de diverses sources écrites ou orales en les résumant de façon cohérente. Peut s'exprimer spontanément, très couramment et de façon précise et peut rendre distinctes de fines nuances de sens en rapport avec des sujets complexes.</p>	<p>Le Niveau C2 intitulé « Maîtrise » a pour but de caractériser le degré de précision, d'adéquation et d'aisance de la langue que l'on trouve dans les 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 relative exactitude, une gamme étendue de modalités</i> ; <i>avoir une bonne maîtrise des expressions idiomatiques et familières accompagnées de la conscience des connotations</i> ; <i>revenir en arrière et reformuler une difficulté sans heurts de sorte que l'interlocuteur s'en aperçoive à peine.</i></p>
	<p>C1 Peut comprendre une grande gamme de textes longs et exigeants, ainsi que saisir des significations implicites. Peut s'exprimer spontanément et couramment sans trop apparemment devoir chercher ses mots. Peut utiliser la langue de façon efficace et souple dans sa vie sociale, professionnelle ou académique. Peut s'exprimer sur des sujets complexes de façon claire et bien structurée et manifester son contrôle des outils d'organisation, d'articulation et de cohésion du discours</p>	<p>Le Niveau C1 est intitulé Niveau autonome. Ce niveau 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 et de la garder, et de gagner du temps pendant qu'il/elle réfléchit ; produit un discours bien construit et sans hésitation qui montre l'utilisation bien maîtrisée des structures, des connecteurs et des articulateurs.</i></p>
Utilisateur indépendant	<p>B2+</p>	<p>Le Niveau B2+ - correspond au degré supérieur du Niveau avancé ou utilisateur indépendant (Vantage) – l'accent est mis sur l'argumentation, sur un discours social efficace et sur la conscience de la langue, apparue en B2 et poursuivie 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 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.</p>
	<p>B2 Peut comprendre le contenu essentiel de sujets concrets ou abstraits dans un texte complexe, y compris une discussion technique dans sa spécialité. Peut communiquer avec un degré de spontanéité et d'aisance tel qu'une conversation avec un locuteur natif ne comportant de tension ni pour l'un ni pour l'autre. Peut s'exprimer de façon claire et détaillée sur une grande gamme de sujets, émettre un avis sur un sujet d'actualité et exposer les avantages et les inconvénients de différentes possibilités.</p>	<p>Les descripteurs étalonnés au Niveau B2 marquent une coupure importante avec ce qui précède. 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 sur un sujet 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 ; expose un problème en signifiant clairement que le partenaire de la négociation doit faire des concessions ; s'interroge sur les causes, les conséquences, les situations hypothétiques ; prend une part active dans une discussion informelle dans un contexte familier, 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</p>

Utilisateur élémentaire		<p>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 ce 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>.</p> <p>Ce Niveau B1+ correspond à un degré élevé du Niveau seuil. On y retrouve les deux mêmes traits caractéristiques auxquels s'ajoutent 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 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 fait parfois 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>.</p> <p>Le Niveau B1 correspond aux spécifications du Niveau seuil. Deux traits le caractérisent particulièrement. Le premier est la capacité à 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 ; fait passer de manière compréhensible l'opinion principale qu'il veut transmettre ; puise avec souplesse dans un large éventail de formes simples pour dire l'essentiel de ce qu'il veut ; peut poursuivre une conversation ou une discussion même s'elle est quelquefois difficile à comprendre lorsqu'elle essaie de dire exactement ce qu'il/elle souhaite ; reste compréhensible même si la recherche des mots et des formes grammaticales ainsi que la remédiation sont évidentes, notamment au cours de longs énoncés. 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 lors de l'organisation d'un voyage chez un voyageur ou au cours du voyage ; intervenir sans préparation dans des conversations sur des sujets familiers</i>.</i></p> <p>Le Niveau A2+ correspond à une capacité nettement supérieure au niveau Niveau intermédiaire ou de survie. On remarquera ici une participation plus active, encore que limitée et accompagnée d'aide, par exemple : <i>est capable de lancer, poursuivre et clore une conversation simple à condition qu'elle soit en face à face ; 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 ; arrive à communiquer sur des sujets élémentaires à condition de pouvoir demander de l'aide pour exprimer ce qu'il/elle veut ; se débrouille dans les situations quotidiennes dont le contenu est prévisible bien qu'en devant adapter le message et chercher ses mots ; et, de manière plus significative, une meilleure capacité à poursuivre un monologue, par exemple : <i>exprime ses impressions en termes simples ; fait une longue description des données quotidiennes de son environnement comme les gens, les lieux, une expérience professionnelle ; 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>.</i></p>
B1+		
B1	Peut comprendre les points essentiels quand une langue claire et standard est utilisée et s'il s'agit de choses familières dans le travail, à l'école, dans les loisirs, etc. Peut se débrouiller dans la plupart des situations rencontrées en voyage dans une région où la langue cible est parlée. Peut produire un discours simple et cohérent sur des sujets familiers et dans ses domaines d'intérêt. Peut raconter un événement, une expérience ou un rêve, décrire un espoir ou un but et exposer brièvement des raisons ou des explications pour un projet ou une idée.	
A2+		

	<p>Peut comprendre des phrases isolées et des expressions fréquemment utilisées en relation avec des domaines immédiats de priorité (par exemple, informations personnelles et familiales simples, achats, environnement proche, travail). Peut communiquer lors de tâches simples et habituelles ne demandant qu'un échange d'informations simple et direct sur des sujets familiers et habituels. Peut décrire avec des moyens simples sa formation, son environnement immédiat et évoquer des sujets qui correspondent à des besoins immédiats.</p>	<p>Le Niveau A2 semble correspondre à la spécification du niveau intermédiaire ou de survie. 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 ; mène à bien un échange très court ; répond à des questions sur ce qu'il fait professionnellement et pour ses loisirs et en pose de semblables ; 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 seul 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>.</p>
<p>A1</p>	<p>Peut comprendre et utiliser des expressions familières ainsi que des énoncés très simples qui visent à satisfaire des besoins concrets. Peut se présenter ou présenter quelqu'un et poser à une personne des questions la concernant – par exemple, sur son lieu d'habitation, ses relations, ce qui lui appartient, etc. – et peut répondre au même type de questions. Peut communiquer de façon simple si l'interlocuteur parle lentement et distinctement et se montre coopératif.</p>	<p>Le Niveau A1 (introductif ou de découverte – Breakthrough) est le niveau le plus élémentaire d'utilisation de la langue à titre personnel – celui où l'apprenant 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, en ne se contentant pas de répéter des expressions toutes faites et préorganisées.</p>

2.2.4. Caractérisation

Le Tableau 1 du CECR donne un aperçu de la compétence langagière considérée dans sa globalité. La compétence globale est un raccourci utile pour répondre à la question « Quel est mon niveau ? » ou « Quel est le niveau de ce manuel ? ». Toutefois, comme cela a été signalé dans la présentation du CECR, le concept de ce que l'on appelle les « compétences partielles » est particulièrement approprié sur un continent multilingue et l'on peut obtenir de nombreuses informations en traçant le profil particulier de chacune d'entre elles. Il existe même un ordre implicite de difficulté dans la série : Réception, Interaction, Production, Médiation, que l'on peut exploiter dans les politiques linguistiques. Pour la première langue étrangère on peut évaluer (et décrire) ces quatre activités langagières. Lorsqu'on arrive à la troisième langue étrangère, on peut se contenter de n'encourager que la réception de l'oral et de l'écrit. Une approche par activité langagière peut ainsi devenir un encouragement au plurilinguisme.

Quoiqu'il en soit, de nombreux examens ou démarches d'évaluation se contentent de mesurer une partie du concept hypothétique (ou construct) « compétence communicative en langues étrangères » à un certain niveau de compétence. Même lorsqu'un examen essaie de couvrir tous les aspects de la « compétence communicative en langues étrangères », l'importance de chacun d'entre eux est accentuée différemment ce qui affecte les résultats.

On peut utiliser le *Schéma descriptif* des Chapitres 4 et 5 du Cadre de référence pour établir ce que l'on évalue et les *Niveaux communs de référence* (Chapitre 3) pour spécifier le niveau de compétence attendu dans chacun des aspects. Le Tableau 2 du CECR présente une grille qui trouve son origine dans les descripteurs du Chapitre 4 (Opérations de communication langagière) élaborés pour le Portfolio européen des langues. Le Tableau 3 du CECR offre l'exemple d'une grille d'analyse issue des descripteurs du Chapitre 5 (Compétences communicatives langagières) élaborés pour un séminaire de formation de correcteurs.

2.3. Comment peut-on exploiter le CECR pour relier les examens aux Niveaux communs de référence ?

La question essentielle qui rend confuses les tentatives de mise en relation des examens entre eux est que des examens différents ont tendance à évaluer des concepts hypothétiques différents ; en effet, ils mettent en œuvre des ensembles différents de spécifications. Toute évaluation suppose un échantillonnage ; un échantillonnage suppose que l'on choisisse des priorités et les priorités varient selon les contextes. En outre, comme cela a été signalé dans la partie précédente, un grand nombre d'examens et d'évaluations limitent volontairement leur champ à certains domaines.

Un cadre commun de référence peut faciliter une vision commune du construct que l'on évalue. On peut ainsi relier les unes aux autres plusieurs évaluations différentes à l'aide de l'échelle commune fournie par les Niveaux communs de référence sans toutefois pouvoir prétendre que deux examens sont exactement équivalents. Chaque examen aura tendance à se focaliser sur des aspects différents – mais on peut analyser ce que chacun d'entre eux recouvre. Si l'on procède à cette analyse à un degré raisonnable de détail, on peut s'attendre à ce que, de même que deux apprenants de Niveau B2 ne seront pas à ce niveau B2 pour les mêmes raisons, on ne trouvera pas deux examens de Niveau B2 conçus de façon identique. Ceci n'enlève rien au fait que les deux examens pourraient néanmoins être décrits avec exactitude comme appartenant au Niveau B2.

2.3.1. Quel rôle le CECR peut-il jouer dans le processus de mise en relation ?

(D'après la Partie 9.2.3. du CECR)

On distingue cinq façons classiques de relier des évaluations distinctes entre elles : (1) l'appariement ;

(2) le calibrage ; (3) l'harmonisation statistique ; (4) le repérage et (5) l'harmonisation sociale¹⁴.

Les trois premières méthodes correspondent à : (1) la production de versions parallèles du même test (appariement) ; (2) la mise en relation des résultats de tests différents et d'une échelle commune (calibrage) et, (3) la correction de la difficulté de certaines épreuves ou de la sévérité d'un examinateur (harmonisation statistique).

Les deux suivantes supposent que l'on parvienne à un consensus par la discussion (harmonisation sociale) et que l'on compare des échantillons de travail à des définitions standards et à des exemples (repérage). Soutenir la démarche qui aboutit à un consensus est l'un des buts du Cadre de référence. C'est pour cette raison que les échelles de descripteurs à utiliser dans ce but ont été standardisées selon une méthodologie rigoureuse (North 2000). Dans le domaine de l'éducation, on décrit de plus en plus cette approche comme une évaluation centrée sur les standards ; mais on reconnaît généralement que la mise en place de cette approche prend du temps parce c'est la démarche d'exemplification et l'échange d'opinions qui permet aux différents partenaires de se mettre d'accord sur la signification des standards.

On peut objecter que cette approche est, potentiellement, la méthode de mise en relation la plus solide parce qu'elle implique l'élaboration et la validation d'un point de vue commun sur le concept hypothétique (ou construct). La principale raison qui rend la mise en relation d'évaluations en langues difficile, et ce en dépit de la magie statistique des techniques traditionnelles, c'est que les évaluations, en règle générale, testent des choses radicalement différentes même quand elles prétendent couvrir le même domaine. Ceci est dû, d'une part, (a) à une conceptualisation et à une opérationnalisation insuffisantes du construct et, d'autre part, (b) à l'interférence de la méthode d'évaluation.

Le Cadre européen de référence propose une méthode guidée pour trouver une solution au premier problème sous-jacent, relatif à l'apprentissage des langues vivantes en contexte européen : conceptualisation et opérationnalisation insuffisantes du construct. Toutefois, le cadre de référence commun proposé par le CECR ne saurait être suffisant pour relier les examens.

Deux des trois méthodes statistiques mentionnées ci-dessus (appariement ; harmonisation statistique) portent sur la **validation** de la stabilité des standards dans des versions différentes du même examen. Cette stabilité est un préalable à une mise en relation significative. Si des versions différentes du même examen ne peuvent être mises en relation, comment pourrait-on les relier à quelque chose d'autre ? La troisième méthode statistique de mise en relation (le calibrage) porte sur la mesure de la façon dont les deux échelles distinctes sont en relation l'une avec l'autre. Les méthodologies fondées sur cette approche proposent une façon de corroborer de manière indépendante les conclusions auxquelles on est arrivé sur la base de l'approche descriptive adoptée par le CECR.

En ce qui concerne l'élaboration et la description de standards, on peut utiliser le CECR essentiellement de deux façons :

1. Pour la **spécification** du contenu des tests et examens : *ce qu'on évalue*
2. Pour la formulation des **standards** qui déterminent qu'un objectif d'apprentissage est atteint : *comment on interprète la performance*

2.3.2. Quelles sont les parties les plus utiles du CECR ?

En règle générale, les parties du CECR les plus pertinentes pour relier les examens aux Niveaux communs de référence sont le Chapitre 3 « Les Niveaux communs de référence », le Chapitre 4

¹⁴ Mislevy, R.J., 1992. Linking Educational Assessments : Concepts, Issues, Methods and Prospects. Princeton N.J. Educational Testing Service.

Linn, R.L. Linking Results of Distinct Assessments. Applied Measurement in Education 6 (1), 83-102.

North, B., 2000. Linking Language Assessments : an example in a low stakes context. System 28 555-577.

« L'utilisation de la langue et l'apprenant utilisateur » pour examiner les spécifications relatives au contenu et la Partie 5.2 « Compétences communicatives langagières » du Chapitre 5 « Les compétences de l'utilisateur/apprenant ».

Toutes les échelles des Chapitres 4 et 5 du CECR sont pertinentes, notamment pour les producteurs d'examens qui souhaitent concevoir des épreuves qui évaluent des compétences partielles. Dans une perspective plus générale, néanmoins, les utilisateurs pourront trouver utiles les documents suivants :

Vue d'ensemble des Niveaux communs de référence

- Chapitre 3, Tableau 1 « Niveaux communs de compétences Echelle globale » (Version française, p. 25)
- Partie 3.6 « Cohérence du contenu des Niveaux communs de référence » (pp. 32-33)
- Document B3 « Cohérence du contenu des Niveaux communs de référence » (p. 159)
- « Niveaux de compétence dans le Cadre de référence de ALTE » (p. 176)

Vue d'ensemble des activités de communication

- Tableau 2, « Niveaux communs de compétence – Grille pour l'auto-évaluation » (pp. 26-27)
- DIALANG Document C3 « Echelles de démonstration détaillées ... » (pp. 170-171)
- ALTE Document D1 « Résumé des niveaux de savoir-faire de ALTE » (p. 178)
- Compréhension générale de l'oral : échelle (p. 55)
- Compréhension générale de l'écrit : échelle (p. 57)
- Interaction orale générale : échelle (p. 61)
- Interaction écrite générale : échelle (p. 68)
- Production orale générale : échelle (p. 49)
- Production écrite générale : échelle (p. 51)

Activités communicatives particulièrement pertinentes dans les domaines éducatif et professionnel

- Comprendre en tant qu'auditeur (p. 56)
- Prendre des notes (conférences, séminaires) (p. 77)
- Lire pour s'orienter (p. 58)
- Lire pour s'informer et discuter (p. 58)
- Lire des instructions (p. 59)
- Traiter un texte (p. 77)
- Echange d'information (p. 67)
- Coopération à visée fonctionnelle (p. 65)
- Discussions et réunions formelles (p. 64)
- Comprendre une interaction entre locuteurs natifs (p. 55)
- Monologue suivi : argumenter (p. 50)
- S'adresser à un auditoire (p. 50)
- Essais et rapports (p. 52)

Vue d'ensemble des aspects de la compétence communicative langagière

- Tableau 3 « Aspects qualitatifs de l'utilisation de la langue parlée » (p. 28)
- Etendue linguistique générale (p. 88)
- Correction grammaticale : échelle (p. 90)
- Correction sociolinguistique (p. 95)
- Aisance à l'oral (p. 100)

2.3.3. Par où commencer ?

Les Niveaux communs de référence sont une échelle de compétence langagière. L'expérience que l'on a, depuis 1950, des échelles de compétence langagière permet de penser qu'il est beaucoup plus facile

de comparer les descripteurs d'une échelle aux performances du type qu'ils sont sensés décrire que d'essayer de commencer par un débat sur le niveau auquel se situent les items d'un test.

Il est logique de commencer (a) à travailler à partir des définitions des standards (descripteurs du CECR) puis, (b) à travailler avec des échantillons de performances écrites et orales que l'on peut directement appairer avec certains de ces descripteurs et enfin, (c) à travailler avec les items . Il est beaucoup plus facile de débattre du niveau de performances enregistrées (orales ou écrites) que l'on peut directement comparer aux critères (les descripteurs) que de se prononcer sur la difficulté d'items . L'évaluation directe se prête naturellement à des stratégies de jugement : l'évaluateur juge la qualité de la performance de la personne en fonction des critères.

Les performances obtenues à des tests de compréhension orale ou écrite ou à des tests sur les structures de la langue ne sont pas directement observables. On ne peut pas s'y référer ni en discuter de façon significative. Ce à quoi on peut se référer ou sur quoi on peut discuter, ce sont les manifestations de la performance en termes de réponses. C'est pourquoi on parle quelquefois de « tests indirects ». Les tests indirects permettent une stratégie de comptage : à combien de questions le candidat a-t-il bien répondu ? On n'est alors pas en mesure d'indiquer directement à quel niveau se trouve le candidat mais on peut donner le nombre de questions auxquelles il a apporté une réponse acceptable. Il faut trouver une méthode pour dire à quoi ces résultats correspondent en termes de niveaux. Si le test visait le Niveau B2, combien de réponses correctes le candidat doit-il avoir obtenues pour être placé au Niveau B2 ? (Où se trouve le seuil fonctionnel de B2 ?). Combien, parmi les items du test considérés comme « B2 » étaient effectivement au niveau B2 plutôt qu'à un niveau supérieur ou inférieur ?

L'évaluation de la difficulté des items d'un test et la décision concernant les « seuils fonctionnels » pour différents niveaux et degrés du test (« définition de standards ») est une procédure complexe – même quand les items ont été rédigés avec des spécifications contraignantes. Même quand un groupe de producteurs de tests parvient à un consensus définitif sur son jugement initial du niveau de difficulté d'un item. Lorsqu'ils disposent des données empiriques du pré-test (montrant le degré réel de difficulté de l'item), les producteurs peuvent être amenés à amender voire à modifier considérablement la formulation des items. Le problème se pose ainsi : « Un item très difficile portant sur un sujet élémentaire est-il un item intermédiaire ou n'est-il qu'un item de niveau élémentaire mal formulé qu'il faudrait rédiger de nouveau pour éviter les pièges ? ». L'histoire de l'évaluation critériée montre que ces questions posent aux enseignants, aux producteurs de tests et aux examinateurs de gros problèmes pour évaluer judicieusement la difficulté des items de tests. On recommande vivement une formation qui inclue le retour d'information donné par l'analyse des données et l'étalonnage afin d'aider à guider le jugement et à produire des items bien formulés et bien ciblés.

On décrit comme « centrée sur le test » une démarche de définition des standards comme celle que l'on vient d'exposer. La procédure suivie permet d'établir la difficulté des items de tests en se fondant sur le jugement d'experts qui s'appuient sur des données. Mais même après avoir mené à bien cette démarche, la question suivante demeure : peut-on être sûr de la relation entre les résultats produits par l'ensemble du test et les niveaux du CECR ? On conseillera une forme d'étalonnage sur les niveaux du CECR. Mais sur quoi peut-on calibrer ? Lorsqu'on en dispose, les données sur les mêmes élèves dans des évaluations directes seront une aide. On peut alors utiliser ce que l'on appelle une approche « centrée sur la personne » pour tracer point par point les résultats de chaque personne aux évaluations subjectives et les résultats de chaque personne aux tests. Une telle démarche est utile pour corroborer indépendamment les ensembles de standards. Des évaluations de ce type utilisant les descripteurs du CECR pourraient être :

- Evaluations d'échantillons de performances écrites et orales notées par des correcteurs formés aux niveaux du CECR.
- Evaluation des mêmes capacités que celles évaluées par les tests, notées par des correcteurs qui connaissent bien les élèves et sont formés aux niveaux du CECR.
- Auto-évaluation des candidats sur les capacités évaluées par ces tests.

Il n'y a pas obligatoirement de relation étroite entre le résultat d'un individu à un test et une notation subjective de la performance de cet individu. Mais pour des groupes assez grands on peut néanmoins arriver à une corrélation entre les deux mesures suffisamment significative pour ce type d'analyse.

2.3.4. Que faut-il éviter ?

Dans le processus que l'on vient de décrire il y a deux raccourcis à éviter si l'on veut éviter de confirmer des conclusions toutes faites :

1. **Repérage prématuré** : il n'est que trop facile d'adopter comme repère un standard local prétendument déjà lié au CECR, pour relier le système local au CECR. Transférer un standard reconnu vers les Niveaux communs de référence est *l'aboutissement* de la démarche de mise en relation et non le point de départ. A moins de disposer d'informations sur *la façon* dont on est parvenu à cette hypothèse, et à moins de pouvoir montrer que le processus en question est voisin des approches préconisées dans ce Manuel, il faudra traiter l'hypothèse avec prudence. Il se peut qu'elle soit juste – mais elle doit être validée. Ce Manuel propose des démarches pour le faire. L'adoption prématurée comme repère d'un standard local que l'on prétend être un niveau du CECR risque d'être contre productive car les individus impliqués dans le projet auront tendance à ne plus se référer au CECR. Ils auront tendance à se reposer sur les standards qu'ils connaissent. Toute erreur au début de la mise en relation de ce standard au CECR aura des répercussions ultérieures amplifiées.

La meilleure chose à faire est donc d'identifier les standards locaux qui peuvent accélérer la mise en relation du système local avec le CECR. Il sera utile de recueillir des échantillons de performances dont on pense qu'elles illustrent ces standards. Ces échantillons s'avéreront très utiles lors de l'étape de la Standardisation quand une comparaison de performances sur des échantillons standards du CECR et sur des échantillons apparentés à aux standards locaux aidera à valider la relation présumée. Il peut alors être possible, lors de la Validation empirique, d'utiliser les notations faites selon les standards locaux dans une étude de validation centrée sur la personne telle que celle que l'on a présentée plus haut. Ces deux mesures fourniront des preuves à l'appui du lien présumé – et éviteront éventuellement de prolonger un malentendu.

2. **Ancrage prématuré** : une façon de relier un examen existant au CECR est de l'étalonner sur l'échelle du CECR en utilisant comme « ancrage » un test déjà étalonné sur le CECR. Les apprenants passent alors à la fois le test d'ancrage et l'examen qui fait l'objet de l'étude. On peut confronter les notes obtenues et établir les équivalences avec le CECR. Mais, de nouveau, il est important d'évaluer d'abord l'hypothèse que le test choisi comme ancrage est lui-même étalonné sur le CECR. Qu'est-ce qui justifie l'hypothèse ? Toute erreur dans l'étalonnage original de ce test d'ancrage causera des distorsions dans la suite du travail.

La meilleure chose à faire consiste à étudier les bases justificatives des hypothèses faites pour le test. Si ces justifications ne sont pas disponibles ou s'il y a un doute, le mieux est d'établir d'abord la relation de l'examen que l'on étudie et du CECR en suivant les démarches préconisées par ce manuel. On peut alors exploiter le test d'ancrage potentiel dans un projet de validation indépendante à la fin du processus. Si l'ancrage vient confirmer les conclusions tirées du processus de mise en relation, c'est une validation forte – et qui, incidemment, justifie l'hypothèse initiale faite au sujet du test d'ancrage. Si les résultats de l'étude d'ancrage sont peu concluants, voire contradictoires, ce sera décevant – mais on n'aura pas biaisé le processus de mise en relation.

Il est donc bon de connaître les standards et les tests locaux potentiellement utiles qui pourraient aider à simplifier le compte rendu des équivalences, une fois que leur relation au CECR a été clairement établie.

Mais la façon la plus sûre de procéder est de commencer à construire une compréhension commune de ce à quoi correspondent les niveaux du CECR. On peut concevoir un projet de mise en relation de telle sorte qu'on valide une hypothèse existante sur un examen ou un standard local. Mais alors, l'hypothèse définitive qu'il y a une relation entre cet examen ou ce standard et le CECR est le *résultat* du processus de mise en relation – pas le point de départ. En conséquence, les techniques proposées par ce Manuel sont présentées dans l'ordre logique suivant :

- 1) **Familiarisation** (Formation au CECR)
- 2) **Spécification** (Centration sur le contenu)
- 3) **Standardisation** (Formation à l'interprétation des Niveaux)
- 4) **Validation empirique** (Analyse des relations avec les standards locaux, les tests d'ancrage).

Chapitre 3

Familiarisation

Avant d'entreprendre les activités de Spécification et de Standardisation, les personnes impliquées dans une démarche de mise en relation de leurs examens avec les niveaux du Cadre doivent avoir une excellente connaissance du CECR.

Les activités de familiarisation avec le CECR peuvent être divisées en deux groupes :

Activités introductives (de *a* à *c*)

- a) Examen d'une sélection des encadrés présents dans le CECR à la fin des parties nécessaires à cette étude : Chapitre 3 (Les Niveaux communs de référence), Chapitre 4 (L'utilisation de la langue et l'apprenant/utilisateur) et Chapitre 5 (Les compétences de l'apprenant/utilisateur).
- b) Débat sur les niveaux du CECR dans leur ensemble à l'aide de l'échelle globale (Tableau 1. Niveaux communs de compétences – Echelle globale) ainsi que de la Partie 3.6 (Cohérence du contenu) éventuellement complétée par l'énoncé des descripteurs du CECR du même niveau et qui sont présentés dans les listes de contrôle du PEL suisse.
- c) Auto-évaluation de son propre niveau de langue dans une langue étrangère – à l'aide du Tableau 2 (Grille du PEL et autres échelles) – suivie d'une discussion.

Analyse qualitative des échelles du CECR (de *d* à *f*)

- d) Tri des descripteurs d'une échelle du CECR (par exemple Aisance) ou de deux ou trois échelles apparentées (par exemple Conversation, Discussion informelle, Tours de parole) selon leurs niveaux ou leur rang. On découpe les descripteurs qui composent l'échelle pour réaliser cette activité.
- e) Reconstitution de chaque case de la grille du Tableau 2 du CECR (Grille des PEL pour l'auto-évaluation) à partir des descripteurs découpés.
- f) Tri par niveau des descripteurs de DIALANG pour l'auto-évaluation, dans une compétence donnée (CECR, annexe C1. Il y a plusieurs descripteurs par niveau).

Ces techniques sont expliquées plus en détail dans la suite du chapitre. Il est vivement conseillé aux utilisateurs de sélectionner deux activités dans chaque groupe avant de démarrer les processus de Spécification et de Standardisation.

3.1. Activités introductives

a : Examen d'une sélection des encadrés du CECR

Cette activité sera plus intéressante si les participants connaissent un tant soit peu les niveaux du CECR (pour avoir, par exemple, assisté à une présentation). L'objectif de l'exercice est de leur faire prendre conscience des nombreux aspects à considérer lors de l'élaboration et de l'analyse d'activités d'évaluation ainsi que de l'étendue de ce que le CECR recouvre.

On peut organiser cette activité suivant différentes modalités :

- photocopier une liste de contrôle d'un encadré (par exemple, ci-dessous, celle centrée sur l'expression orale) et organiser une discussion en petits groupes ou en plénière.
- sélectionner certains encadrés du CECR particulièrement appropriés et faire élaborer une nouvelle liste de contrôle.

- sélectionner un certain nombre d'autres encadrés appropriés et les distribuer à de petits groupes. Désigner un rapporteur. Le compte rendu sera suivi une discussion en plénière.

Cette dernière variante constitue une manière efficace de traiter la question assez vite et de rafraîchir la mémoire des participants au sujet du CECR.

<p><i>Les personnes qui utilisent le Cadre pour analyser et évaluer les performances d'expression orale envisageront et expliciteront selon le cas :</i></p> <ul style="list-style-type: none"> - <i>comment les conditions matérielles dans lesquelles l'apprenant sera amené à communiquer affecteront ce qu'il doit faire ;</i> - <i>comment le nombre et la nature des interlocuteurs affecteront ce que l'apprenant doit faire ;</i> - <i>avec quelles contraintes de temps l'apprenant devra effectuer sa performance ;</i> - <i>dans quelle mesure les apprenants devront s'adapter au contexte mental de leur interlocuteur ;</i> - <i>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.</i> 	<p>Approprié ? Pourquoi ?</p>
--	-----------------------------------

b : Débat sur les niveaux du CECR dans leur ensemble

Cette deuxième activité est recommandée si le CECR n'est pas du tout connu des participants. On peut l'organiser suivant différentes modalités :

- demander aux participants de lire individuellement le Tableau 1 du CECR puis de voir ensuite, en tandems ou en petits groupes, s'ils connaissent des apprenants qui pourraient correspondre à certains niveaux. A l'issue de cette discussion en groupe, on peut distribuer des copies de la Partie 3.6 du CECR pour que les participants se rendent mieux compte des traits caractéristiques de chacun des niveaux du CECR.
- Si les participants connaissent déjà le CECR, le coordinateur peut photocopier et utiliser le Tableau 2.1 de ce Manuel qui combine le Tableau 1 du CECR et la Partie 3.6.
- Après l'examen du Tableau 1 du CECR (ou du Tableau 2.1 de ce Manuel), demander aux participants quel niveau leur semble se rapprocher le mieux de leur situation. On les regroupe alors selon ces niveaux et on leur communique la liste de contrôle des descripteurs du CECR (prototype suisse du PEL) pour le niveau qu'ils traitent.

c : Auto-évaluation à l'aide du Tableau 2 (Grille du PEL)

Cette activité est un bon point de départ pour des groupes de participants qui connaissent déjà le Portfolio.

- On demande aux participants d'auto évaluer leur compétence dans deux langues étrangères à l'aide de la grille du PEL (Tableau 2 du CECR). Ils en débattent ensuite avec leurs voisins directs. L'importance de cette discussion ne doit pas être sous-estimée.
- On peut avantageusement compléter cette auto-évaluation (conseillée aux utilisateurs du PEL) en consultant les listes de contrôle des descripteurs du CECR correspondants aux niveaux en question (prototype suisse du PEL).
- on peut également demander aux participants d'auto évaluer la qualité de leur propre langue : dans quelle mesure font-ils bien ce qu'ils savent faire ? Pour ce faire, on peut utiliser :
 - a) Ou bien le Tableau 3 du CECR qui définit chaque niveau pour l'Etendue (linguistique générale), la Correction (grammaticale), l'Aisance, la Cohérence et l'Interaction.
 - b) Ou bien l'Aisance à l'oral (CECR p.100) et la Correction grammaticale (CECR p.90)

3.2. Analyse qualitative des échelles du CECR

Une fois achevées les activités introductives, 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. Le coordinateur doit choisir au moins l'une des trois options (de *d* à *f*).

d : *Tri des descripteurs d'une échelle du CECR*

Le tri de descripteurs a été largement expérimenté dans différents contextes : projet suisse d'élaboration des descripteurs pour le PEL, plusieurs projets finlandais. Cette tâche a le mérite d'obliger les participants à examiner les descripteurs indépendamment les uns des autres comme des critères autonomes.

Elle exige toutefois qu'on la prépare afin d'en faire une activité relativement simple.

- Le coordinateur prépare 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 comme des bandes de téléscripneur. Si l'on mélange des échelles apparentées (par exemple, Conversation, Discussion informelle, Tours de parole), 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 ... de couper droit !)
- 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, comme ils préfèrent.
- Ils en discutent ensuite avec les autres participants afin d'arriver à un consensus.
- Puis ils comparent à 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.

e : *Reconstitution du Tableau 2 du CECR*

Cette activité est une variante de la précédente mais elle utilise la grille du PEL – elle-même élaborée à partir des descripteurs du CECR – plutôt que les échelles du CECR. On peut utiliser l'ensemble des échelles (6 activités langagières x 6 niveaux = 36 descripteurs) ou une version plus simple (une seule colonne = 6 descripteurs).

- On distribue la grille vierge (agrandissement au format A3 de la grille du PEL dont les cases ont été vidées de leur contenu) ainsi que, dans une enveloppe, les descripteurs découpés. 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 différents savoir-faire. 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.

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. Le coordinateur contrôle le travail de groupe et aide à clarifier les doutes sur l'interprétation des différents descripteurs.

- Le coordinateur distribue une copie du Tableau 2 achevé et « complet » pour que les participants vérifient leur exercice de reconstitution et pour faciliter la discussion.
- On demande aux participants d'auto évaluer leur propre connaissance des langues étrangères (d'abord individuellement) puis d'en discuter avec le groupe en se référant aux niveaux et compétences du Tableau 2 du CECR.

f: Tri par niveau des descripteurs de DIALANG pour l'auto-évaluation dans une compétence donnée

Les descripteurs de DIALANG (Annexe C du CECR) reformulent les descripteurs du CECR de façon à permettre une auto-évaluation sur des échelles abrégées pour la Compréhension de l'écrit, la Production écrite et la Compréhension de l'oral. On effectue l'exercice exactement comme dans (d) *Tri des descripteurs du CECR* ci-dessus.

Tableau 3.1 : Gestion du temps pour les activités de Familiarisation

Gestion du temps pour les activités de Familiarisation

- *Elles peuvent être organisées indépendamment de toute autre activité de formation. On peut les utiliser au début des activités de Spécification et de Standardisation.*
- *Elles durent environ 3 heures :*
 - *Brève présentation du CECR par le coordinateur (30 minutes)*
 - *Activités introductives (a à c) et discussion (45 minutes)*
 - *Activités qualitatives (d à f) y compris le travail de groupe (60 minutes)*
 - *Discussion (30 minutes)*
 - *Conclusion (15 minutes)*

Documents à préparer

- *Listes de questions basées sur les encadrés du CECR les mieux adaptés à la situation (recommandé uniquement pour des participants chevronnés)*
- *Copies des Tableaux 1 (ou du Tableau 2.1) et 2 du CECR pour tous les participants (et dans tous les contextes)*
- *Découpages des énoncés du Tableau 2 du CECR en vue du travail de groupe (tous contextes, un ensemble par enveloppe, une enveloppe pour chaque sous-groupe de travail)*
- *Découpages des descripteurs des échelles appropriées du CECR pour l'évaluation choisie (par exemple, pour la Production orale : Interaction orale générale, Aisance à l'oral, Etendue linguistique générale ; pour la Compréhension de l'oral : Compréhension générale de l'oral, Comprendre une interaction entre locuteurs natifs, Comprendre des émissions de radio et des enregistrements). Pour entrer dans le détail d'une compétence particulière : une échelle découpée par enveloppe, une enveloppe pour chaque sous-groupe de travail.*
- *Copies de listes de contrôle de descripteurs pour un ou deux niveaux particuliers, choisies sur l'ensemble des échelles du CECR (pour entrer dans le détail d'un niveau donné).*

Chapitre 4

Spécification

Ce chapitre traite de la description et de l'analyse du contenu de ce que recouvrent les épreuves de l'examen ou du test étudié pour une mise en relation avec le CECR. Il se présente comme un processus continu allant du général au particulier.

Les démarches présentées sont destinées à être suivies par l'équipe qualifiée responsable de l'examen. Elles peuvent donner lieu soit à des débats, soit à des réalisations individuelles suivies de débats. Le produit final doit faire l'objet d'une déclaration officielle de la part de l'organisme concerné.

Ce chapitre a deux principaux objectifs :

- **La prise de conscience** : son objectif est de faire prendre conscience, aux concepteurs des examens, des éléments qui entrent dans la composition d'un examen de qualité :
 - l'importance d'une bonne analyse de contenu de ces examens ;
 - les possibilités d'utilisation du CECR pour la planification et la description des examens de langues ;
 - l'importance de la mise en relation des examens avec un cadre de référence international tel que le CECR.
- **La définition de standards** : il s'agit de définir des standards minimaux en ce qui concerne à la fois la qualité de la spécification du contenu des examens de langues et le processus de mise en relation des examens avec le CECR.

Les démarches concrètes que ce chapitre expose doivent aider les utilisateurs à se familiariser avec le CECR, à prendre conscience de l'importance d'une mise en relation correcte des examens avec les échelles du CECR et, en dernier lieu, à faciliter cette analyse de contenu et ce processus de mise en relation. Le compte rendu basé sur le résultat de ces démarches a pour objectif de rendre les examens plus transparents aussi bien pour les utilisateurs des résultats que pour les candidats eux-mêmes.

L'analyse de contenu prend le CECR comme principal élément de référence. Des éléments de référence complémentaires sont fournis par la série de spécifications de contenus liés au CECR, élaborées en association avec le Conseil de l'Europe. Pour l'anglais, la liste des spécifications est la suivante : Breakthrough (A1) ; Waystage (A2) ; The Threshold Level (B1) ; Vantage Level (B2). Pour l'allemand, Profile Deutsch vient d'être publié et un outil semblable (Un référentiel pour le français) est en préparation pour le français.

En plus d'un panorama général, il est nécessaire de faire une description détaillée de la façon dont les tâches, pour chaque savoir-faire, et les critères de notation, pour les activités de production sont reliées aux descripteurs du CECR.

Ce chapitre s'ouvre sur une introduction qui donne des **informations** sur les objectifs, les démarches et la façon de les appliquer, les personnes à qui il s'adresse et ce qui le justifie. Il propose ensuite un ensemble de **démarches** pour l'analyse du contenu d'une épreuve d'examen ou de test de langue, en liaison avec le CECR. Il se présente comme un ensemble de tableaux et de fiches.

4.1. Introduction

Outre les deux buts principaux mentionnés ci-dessus, les objectifs plus concrets du chapitre sur la Spécification sont :

- d'analyser le contenu des épreuves de l'examen ou du test par rapport aux *catégories* pertinentes du CECR, après s'être familiarisé avec celui-ci (Chapitre 3) ;
- d'établir le profil des épreuves de l'examen ou du test par rapport aux *échelles de descripteurs* du CECR, en se fondant sur cette analyse de contenu ;
- d'émettre l'hypothèse qu'une épreuve d'examen ou de test est reliée, sur la base de ce profil, à un *niveau* donné du CECR.

Ce chapitre propose un ensemble de démarches pour l'analyse de contenu. L'application de ces démarches suppose trois types d'activité :

- Se familiariser avec le CECR (activités décrites dans le Chapitre 3).
- Compléter de façon détaillée un certain nombre de listes de contrôle portant sur le contenu de l'examen de langue.
- Utiliser les descripteurs du CECR pour relier l'examen de langue aux niveaux et aux catégories du CECR.

Le processus de Spécification donne aux producteurs d'examen l'occasion :

- D'approfondir leur réflexion sur tout ce que l'élaboration d'examens de langues de qualité suppose ;
- D'accroître leur prise de conscience de l'importance d'une bonne analyse du contenu des examens ;
- De bien connaître le CECR et de l'utiliser pour la planification et la description des examens ; de se convaincre de l'importance de relier les examens de langue à un cadre de référence local ou international (le CECR) ;
- De décrire et d'analyser de façon détaillée le contenu des épreuves de l'examen ou du test à l'étude ;
- De fournir la preuve de la qualité de l'examen ;
- De fournir la preuve de la relation entre les examens et le CECR ;
- D'apporter des conseils aux rédacteurs d'items ;
- D'accroître la transparence du contenu et de la qualité de l'examen ou du test ainsi que de la relation de l'examen au CECR pour les enseignants, les examinateurs professionnels, les utilisateurs des examens et les candidats.

La plupart des fiches à remplir doivent être considérées comme des outils qui accompagnent le processus décrit. Elles constituent également un recueil de preuves pour les enseignants et les examinateurs professionnels et permettent d'appuyer l'hypothèse avancée. Pour les candidats et le grand public, on rédigera un bref compte rendu adapté (voir Chapitre 7) sur la base des fiches qui ont été remplies.

Le processus s'organise autour de deux phases :

- Une **description générale** des épreuves de l'examen ou du test en question.
- Une **description détaillée** des épreuves de l'examen ou du test en question.

Phase 1 : Description générale de l'examen

Cette phase comprend deux étapes :

- En premier lieu, la fiche A1, qui fournit une analyse globale des épreuves de l'examen ou du test en question, doit être complétée.
- Dans un second temps, on complètera les fiches A2 à A7 afin de recueillir plus de détails sur différents aspects de l'examen. Cela suppose une description de la démarche d'élaboration,

une information sur la correction, la notation et le niveau des examens ainsi que sur la façon de rendre compte des résultats et sur les procédures d'analyse et de révision.

Phase 2 : Description détaillée du contenu par rapport au CECR

Cette phase suppose que l'on procède à une analyse plus détaillée des épreuves de l'examen ou du test étudié par rapport au CECR. Cette analyse se fera en deux étapes :

- En premier lieu on donnera une information générale sur la relation de l'examen avec l'échelle globale du CECR (Fiche A8).
- Dans un second temps, on complètera les Fiches A9 à A22 pour recueillir plus de détails sur les différentes Activités de communication langagière (CECR Chapitre 4) et sur les Aspects de la compétence communicative langagière que couvre l'examen par rapport aux sous échelles adéquates du CECR.

Ce processus débouche sur la présentation graphique de ce que recouvre l'examen par rapport aux catégories et aux niveaux les plus pertinents du CECR (Fiche A23).

Qu'il s'agisse d'examens de langue générale ou d'examens sur objectifs spécifiques, la démarche est identique. En effet, bien que les objectifs détaillés mentionnés ci-dessus portent sur l'apprentissage général d'une langue, le CECR prend en compte différents domaines (public, personnel, *éducatif*, *professionnel*). La raison principale du regroupement des activités de communication langagière dans les catégories de Réception, Interaction, Production et Médiation plutôt que sous les quatre savoir-faire traditionnels est précisément que ce regroupement inclut efficacement les domaines éducatifs et professionnels (voir la liste dans la Partie 2.4.2).

Dans le cas où un domaine évalué n'apparaîtrait pas dans le CECR, on demandera à l'utilisateur de le décrire.

Les démarches proposées dans ce chapitre ne sont pas les seules qui existent. Elles ont été conçues pour le projet en cours. Les utilisateurs peuvent naturellement consulter d'autres publications proposant d'autres démarches d'analyse descriptive¹⁵ ou autre permettant la mise en relation d'un examen avec un cadre de référence¹⁶.

Comment suivre les procédures

Ce chapitre fournit un choix important de tableaux tirés des échelles de descripteurs du CECR, accompagnés de fiches à remplir. Le CECR étant conçu pour être le plus complet possible, le nombre de fiches de ce chapitre est considérable.

¹⁵ Liste sélective d'ouvrages sur les procédures d'analyse descriptive :

Alderson, Clapham et Wall. 1995. *Language Test Construction and Evaluation*. Cambridge University Press, Chapitre 2.

Davidson F. et B. Lynch. 1993. *Criterion-referenced language test development: a prolegomenon*. In Huhta, A., K. Sajavaara et S. Takala (sous la direction de). *Language Testing: New Openings*. Jyväskylä, Finlande : Université de Jyväskylä, pp. 73-89.

Davidson F. et B. Lynch. 2002. *Testcraft: A Teacher's Guide to Writing and Using Test Specifications*. Yale University Press.

Hambleton

Lynch B. et F. Davidson. 1994. *Criterion-referenced language test development: linking curricula, teachers and tests*. *TESOL Quarterly* 28:4, pp.727-743.

Lynch B. et F. Davidson. 1998. *Criterion referencing*. In : C. Clapham et D. Dorson (sous la direction de). *Language Testing and Assessment, Volume 7, Encyclopedia of Language and Education*. Dordrecht: Kluwer Academic Publishers, pp. 263-273.

¹⁶ Les listes de contrôle de la Charte de ALTE et les outils d'analyse du contenu des examens de ALTE – disponible sur infoalte.org

Tous les utilisateurs doivent remplir les fiches se rapportant à la description générale de la Phase 1 (Fiches A1 à A7). Toutefois, on n'attend pas d'un producteur d'examen, d'un coordinateur ou d'un enseignant qu'il remplisse toutes les fiches de la description détaillée de la Phase 2 (Fiches A9 à A22). Il faut choisir celles de ces fiches qui sont adéquates pour l'analyse de l'épreuve de l'examen ou du test en question. Donnons deux exemples. Si un examen ne comprend que des épreuves de vocabulaire, on ne remplira que les fiches correspondantes et on examinera uniquement l'échelle concernant l'étendue du vocabulaire. Si un examen mesure plusieurs compétences langagières pour des savoir-faire différents, il faudra remplir un plus grand nombre de fiches et se reporter à un plus grand nombre d'échelles. On recommande de suivre les étapes dans l'ordre ci-dessous qui est l'ordre logique dans lequel mener à bien le processus de Spécification.

Description générale

1. Il est impératif de se familiariser avec le CECR lui-même. Les Activités de Familiarisation du Chapitre 3 constitueront donc le point de départ.
2. La Fiche A1 de « Description générale de l'examen » doit être complétée en analysant l'examen en question.
3. Les Fiches A2 à A7 décrivent ensuite les étapes les plus importantes du déroulement de l'examen. Ces fiches exigent que l'on ait des informations sur les objectifs de l'examen, les besoins de la population d'apprenants à qui il est destiné, l'élaboration de l'examen, les parties et les différentes épreuves, les coefficients et leur justification, les types de textes, de tâches et d'items, les critères de correction – y compris un barème détaillé contenant des exemples de réponses possibles et la façon dont on rendra compte des résultats auprès des apprenants. La description doit également préciser comment les objectifs, le contenu et les coefficients sont périodiquement révisés.

FICHE	SUJET
A2	Elaboration du test
A3	Correction
A4	Notation
A5	Communication des résultats
A6	Analyse des données
A7	Justification logique

Remplir ces fiches est une exigence minimale. Elles donnent une image de la qualité et de la validité de l'épreuve de l'examen ou du test en question. L'information recueillie augmente la transparence de l'examen ou de l'épreuve pour l'enseignant et le candidat.

Description détaillée

Le nombre important de fiches et de tableaux à compléter doit permettre de donner une analyse plus détaillée des épreuves de l'examen ou du test traité et de relier ces descriptions détaillées aux sous échelles du CECR. Dans la plupart des fiches, une brève description, une référence et/ou une justification sont demandées.

4. A ce niveau de la description, il existe des fiches et des tableaux pour chacune des Activités de communication langagière (CECR Chapitre 4) ainsi que pour les Aspects des compétences communicatives langagières (CECR Chapitre 5). Seules les fiches correspondant au contenu de l'examen seront complétées.
5. Nous proposons tout d'abord les fiches sur les Activités de communication langagière. Le Tableau 4.1 ci-dessous en donne une vue d'ensemble. A la fin de la plupart des fiches, une

comparaison de l'épreuve en question avec la sous échelle correspondante du CECR est demandée. Ces échelles sont présentées sous forme de tableau après chaque fiche.

TABLEAU 4.1 : FICHES ET ECHELLES DU CECR POUR LES ACTIVITES DE COMMUNICATION LANGAGIERE			
FICHE	ACTIVITES DE COMMUNICATION LANGAGIERE	FICHE	ECHELLE
A8	Impression générale (évaluation intuitive) du niveau global de l'examen	➡➡	(CECR Tableau 1)
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	➡➡	

- Viennent ensuite les fiches pour les Aspects de la compétence communicative langagière. Le Tableau 4.2 ci-dessous en donne une vue d'ensemble. Cette partie est organisée différemment. On fournit d'abord les descripteurs du CECR pour les « Paramètres qualitatifs pertinents pour la Réception » (Tableau 4.3). On demande ensuite aux utilisateurs de remplir la fiche sur les « Aspects de la compétence langagière pour la Réception » (Fiche A19) sur la base d'une analyse des épreuves de l'examen ou du test à l'étude. A la fin de chaque fiche, les utilisateurs compareront 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.

**TABLEAU 4.2 : ECHELLES DU CECR POUR LES ASPECTS DE LA
COMPÉTENCE COMMUNICATIVE**

	RECEPTION		INTERACTION		PRODUCTION		MEDIATION	
	Compréhension de l'oral	Compréhension de l'écrit	Interaction Orale	Interaction Ecrite	Production Orale	Production Ecrite	Médiation Orale	Médiation Ecrite
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			➡➡		➡➡		➡➡	
Précision	➡➡	➡➡			➡➡	➡➡	➡➡	➡➡
Compétence stratégique								
Reconnaître des indices et faire des déductions	➡➡	➡➡					➡➡	➡➡
Tours de parole (reprise)			➡➡					
Coopérer			➡➡	➡➡				
Faire clarifier			➡➡	➡➡				
Planifier					➡➡	➡➡		➡➡
Compenser			➡➡	➡➡	➡➡	➡➡	➡➡	➡➡
Contrôler et corriger			➡➡	➡➡	➡➡	➡➡	➡➡	➡➡

Si cela s'avère nécessaire pour l'examen en question, il est possible de suivre la même démarche pour l'« Interaction » en utilisant le Tableau 4.4 et la Fiche A20 ; pour la « Production », le Tableau 4.5 et la Fiche A21 et pour la « Médiation », la Fiche A22. Pour la Médiation, aucune échelle du CECR n'est donnée. Nous renvoyons les utilisateurs aux descripteurs de la Réception et de la Production.

Conclusion

- Après avoir complété toutes les fiches de la Phase de Description générale et de la Phase de Description détaillée qui sont pertinentes pour les épreuves de l'examen ou du test en question, on devra présenter sous une forme graphique le détail de la mise en relation. Cette présentation schématise la relation de l'examen analysé avec les sous échelles du CECR.

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 et de mise en relation. On recommande

vivement de réexaminer chaque hypothèse avancée au cours du processus et de corriger, le cas échéant, l'indication donnée dans la Fiche A8. Autrement dit, si le processus d'analyse détaillée aboutit à une conclusion différente de celle annoncée dans la Fiche A8, l'utilisateur doit reconsidérer l'analyse et réviser son jugement quant à l'exactitude de l'hypothèse initiale.

On trouvera, à la fin du Chapitre, une représentation graphique d'un résultat (Fiche A23). Il s'agit de l'affirmation que l'examen étudié est en relation avec différentes échelles du CECR. Cette affirmation se base sur une étude minutieuse ainsi que sur une analyse approfondie du contenu de l'examen.

Les chapitres suivants (Chapitres 5 et 6) fournissent des outils pour renforcer cette affirmation. Une recherche supplémentaire et une analyse encore plus approfondie lors d'étapes ultérieures peuvent provoquer une révision de l'hypothèse avancée. Il faut donc mettre l'accent sur un processus d'amélioration continue de l'exactitude de l'hypothèse. On recommande vivement aux producteurs d'examens d'impliquer leurs collègues dans des débats et des échanges au cours du processus.

Il est très difficile de se prononcer sur le temps nécessaire pour mener le processus à bien. Cela sera différent d'un examen à un autre et dépendra de la complexité des épreuves de l'examen ou du test étudié. La deuxième version du Manuel, qui fera suite à l'expérimentation, proposera, en annexe, des conseils supplémentaires, des exemples de procédures correctes, ainsi que des fiches complétées.

Standards minimaux

Le standard minimal est que les fiches suivantes soient complétées :

- Les fiches de la Phase 1 (Description générale : A1 à A7)
- La Fiche A8 (Impression générale sur le niveau global de l'examen)
- Certaines fiches de la Phase 2 (Description détaillée) appropriées aux épreuves de l'examen ou du test en question
- La Fiche A23 (Représentation graphique de la relation de l'examen avec les niveaux du CECR)
- Le bref compte rendu pour le grand public (voir Chapitre 7).

4.2. Procédures

Les fiches de ce chapitre Sont disponibles sur le site Internet www.coe.int/lang/fr

Ces fiches doivent être complétées selon la procédure décrite ci-dessus. Dans chacune d'entre elles, on peut cocher autant de cases que nécessaire. Par exemple, si l'examen convient à la fois pour des scolaires et pour des adultes, on peut le signaler en cochant les deux cases.

4.2.1. Description générale

Fiche A1 : Description générale de l'examen

DESCRIPTION GENERALE DE L'EXAMEN		
Informations générales		
Intitulé de l'examen		
Langue évaluée		
Organisme examinateur		
Date de cette version de l'examen		
Type d'examen <input type="checkbox"/> international <input type="checkbox"/> national <input type="checkbox"/> régional <input type="checkbox"/> propre à l'institution		
Objectifs		
Population cible <input type="checkbox"/> 1 ^{er} cycle du sec. <input type="checkbox"/> 2 ^{ème} cycle du sec. <input type="checkbox"/> enseignement supérieur <input type="checkbox"/> adultes		
Nombre de candidats par an		
Quel est l'objectif général ?		
Quels sont les objectifs plus particuliers ? Si possible, décrivez les besoins des utilisateurs potentiels à qui cet examen est destiné.		
Quel/s est/sont les principaux domaines traités ?		<input type="checkbox"/> public <input type="checkbox"/> personnel <input type="checkbox"/> professionnel <input type="checkbox"/> éducationnel
Quelles sont les activités communicatives testées ?	<input type="checkbox"/> 1 compréhension de l'oral <input type="checkbox"/> 2 compréhension de l'écrit <input type="checkbox"/> 3 interaction orale <input type="checkbox"/> 4 interaction écrite <input type="checkbox"/> 5 production orale <input type="checkbox"/> 6 production écrite <input type="checkbox"/> 7 compétences intégrées ?? <input type="checkbox"/> 9 médiation orale <input type="checkbox"/> 10 médiation écrite <input type="checkbox"/> 11 langue (par ex. grammaire, vocabulaire, cohésion) <input type="checkbox"/> 12 autre (préciser) :.....	Intitulé de l'épreuve
Intitulé et durée des épreuves (voir colonne de droite ci-dessus)	Titre de l'épreuve 1 2 3 4 5 6 7 8 9	Durée de l'épreuve
Quels sont les types de tâches proposés ?	<input type="checkbox"/> choix multiple <input type="checkbox"/> vrai/faux <input type="checkbox"/> appariement <input type="checkbox"/> classement <input type="checkbox"/> phrase <input type="checkbox"/> phrase à compléter <input type="checkbox"/> texte lacunaire fermé / test de closure fermé <input type="checkbox"/> texte lacunaire ouvert / test de closure ouvert <input type="checkbox"/> question ouverte à réponse courte <input type="checkbox"/> réponse développée (texte, monologue)	Epreuves (indiquer ci-dessous)

	<input type="checkbox"/> interaction avec l'examineur <input type="checkbox"/> interaction avec des pairs	
Quels sont les renseignements fournis aux candidats et aux enseignants ?	<input type="checkbox"/> objectif général <input type="checkbox"/> domaine(s) principaux <input type="checkbox"/> épreuves <input type="checkbox"/> activités <input type="checkbox"/> exemples d'épreuves <input type="checkbox"/> vidéo de l'oral	<input type="checkbox"/> exemples de feuilles de réponses <input type="checkbox"/> critères de correction <input type="checkbox"/> barème de notation <input type="checkbox"/> échantillons de performances standards du niveau de réussite <input type="checkbox"/> fac-similé de diplôme
Sous quelle forme les résultats sont-ils délivrés ?	<input type="checkbox"/> note globale <input type="checkbox"/> note par épreuve	<input type="checkbox"/> note globale plus profil graphique <input type="checkbox"/> profil par épreuve

Fiche A1 : Description générale de l'examen

Les Fiches A2 à A7 apportent de brefs détails complémentaires sur différents aspects :

- La démarche d'élaboration de l'examen (Fiche A2)
- Les critères de correction et les barèmes de notation pour les différentes épreuves (Fiche A3)
- La notation et les procédures de définition des standards pour les différentes épreuves (Fiche A4)
- La délivrance des résultats (Fiche A5)
- Les procédures d'analyse et de révision (Fiche A6)
- La justification des décisions (Fiche A7)

Fiche A2 : Elaboration de l'examen

Elaboration de l'examen	Brève description et/ou références
Quel organisme a décidé de la nécessité de cet examen ?	<input type="checkbox"/> L'institution <input type="checkbox"/> Un institut culturel <input type="checkbox"/> Le ministère de l'Éducation <input type="checkbox"/> Le ministère de la Justice <input type="checkbox"/> Autre (préciser) :
Si un organisme extérieur est impliqué, quelle est son influence sur la conception et l'élaboration ?	<input type="checkbox"/> définir les objectifs généraux <input type="checkbox"/> fixer le niveau de compétence en langue <input type="checkbox"/> fixer le domaine et le contenu de l'examen <input type="checkbox"/> fixer la forme de l'examen et le type d'épreuves <input type="checkbox"/> autre (préciser) :
S'il n'y avait pas d'implication d'un organisme extérieur, quels sont les autres paramètres qui ont influencé la conception et l'élaboration de l'examen ?	<input type="checkbox"/> une analyse de besoins <input type="checkbox"/> une description interne des objectifs de l'examen <input type="checkbox"/> une description interne du niveau de langue <input type="checkbox"/> un référentiel ou un programme <input type="checkbox"/> le profil des candidats
Lors de l'élaboration des épreuves a-t-on tenu compte des différents profils des candidats ?	<input type="checkbox"/> origine linguistique (L1) <input type="checkbox"/> acquis linguistiques antérieurs <input type="checkbox"/> âge <input type="checkbox"/> niveau d'instruction <input type="checkbox"/> milieu socio-économique <input type="checkbox"/> facteurs socioculturels <input type="checkbox"/> origine ethnique <input type="checkbox"/> sexe
Qui rédige les items ou élabore les épreuves du test ?	
Les rédacteurs d'épreuves bénéficient-ils de conseils pour en garantir la qualité ?	<input type="checkbox"/> formation <input type="checkbox"/> lignes directrices <input type="checkbox"/> listes de contrôle <input type="checkbox"/> exemples d'activités valides, fiables et appropriées <input type="checkbox"/> descriptions calibrées sur les niveaux du cadre <input type="checkbox"/> descriptions calibrées sur d'autres niveaux
Donne-t-on une formation aux rédacteurs ?	<input type="checkbox"/> oui <input type="checkbox"/> non
Les épreuves font-elles l'objet d'une discussion avant leur utilisation ?	<input type="checkbox"/> oui <input type="checkbox"/> non

Si oui, qui y participe ?	<input type="checkbox"/> les collègues, individuellement <input type="checkbox"/> un groupe interne de discussion <input type="checkbox"/> une commission externe d'examen <input type="checkbox"/> des personnes impliquées, en interne <input type="checkbox"/> des personnes impliquées, en externe
Les épreuves sont-elles prétestées ?	<input type="checkbox"/> oui <input type="checkbox"/> non
Si oui, comment ?	
Sinon, pourquoi ?	
La fiabilité du test est-elle évaluée ?	<input type="checkbox"/> oui <input type="checkbox"/> non
Si oui, comment ?	<input type="checkbox"/> recueil de données et mesures psychométriques <input type="checkbox"/> autre (préciser) :
Les différents aspects de la validité sont-ils évalués ?	<input type="checkbox"/> validité apparente <input type="checkbox"/> validité de contenu <input type="checkbox"/> validité convergente <input type="checkbox"/> validité prédictive <input type="checkbox"/> validité de construct
Si oui, décrivez de quelle façon	

Fiche A2 : Elaboration de l'examen

Fiche A3 : Correction

Correction : épreuve de _____	Remplir un exemplaire de cette fiche pour chaque épreuve. Brève description et/ou référence
Comment les épreuves sont-elles corrigées ?	Epreuves de réception : <input type="checkbox"/> lecteur optique <input type="checkbox"/> examinateur Epreuves de production ou épreuves intégrées : <input type="checkbox"/> examinateur formé <input type="checkbox"/> enseignants
Où sont corrigées les épreuves ?	<input type="checkbox"/> par un organisme central <input type="checkbox"/> localement : <input type="checkbox"/> par des équipes locales <input type="checkbox"/> par des correcteurs individuels
Quels sont les critères de sélection des correcteurs ?	
Comment l'exactitude de la notation est-elle recherchée ?	<input type="checkbox"/> contrôles réguliers du coordinateur <input type="checkbox"/> formation des correcteurs <input type="checkbox"/> sessions de formation à la standardisation des évaluations <input type="checkbox"/> utilisation d'exemples standards d'épreuves : <input type="checkbox"/> calibrées par rapport au CECR <input type="checkbox"/> calibrées par rapport à une autre description de niveaux <input type="checkbox"/> non calibrées par rapport au CECR ou à toute autre description
Décrire les spécifications des critères de notation des épreuves de production ou des épreuves intégrées.	<input type="checkbox"/> note globale pour chaque épreuve <input type="checkbox"/> notes pour différents aspects de chaque épreuve <input type="checkbox"/> échelle de notation pour la performance globale <input type="checkbox"/> grille de notation pour différents aspects de la performance <input type="checkbox"/> échelle de notation pour chaque épreuve <input type="checkbox"/> grille de notation pour différents aspects de chaque épreuve

	<input type="checkbox"/> échelle de notation par niveau, sans lien avec le CECR <input type="checkbox"/> échelle de notation par niveaux en liaison avec le CECR
Les épreuves intégrées ou de production font-elles ou non l'objet d'une double correction ?	<input type="checkbox"/> correction simple <input type="checkbox"/> deux correcteurs simultanément <input type="checkbox"/> double correction des copies d'écrits / des enregistrements des productions orales <input type="checkbox"/> autre (préciser) :
S'il y a double correction, que fait-on en cas de désaccord entre les correcteurs ?	<input type="checkbox"/> appel à un troisième correcteur dont la note sera celle qui sera gardée <input type="checkbox"/> appel à un troisième correcteur et choix des deux notes les plus proches <input type="checkbox"/> moyenne des deux notes <input type="checkbox"/> consensus entre deux correcteurs après discussion <input type="checkbox"/> autre (préciser) :
L'accord entre les correcteurs est-il mesuré ?	<input type="checkbox"/> oui <input type="checkbox"/> non

Fiche A3 : Correction

Fiche A4 : Notation

Notation : Epreuve de _____	Remplir un exemplaire de cette fiche pour chaque épreuve. Brève description et/ou référence
Y a-t-il une note d'admissibilité ou donne-t-on des mentions ?	<input type="checkbox"/> admissibilité <input type="checkbox"/> mentions
Décrivez la démarche suivie pour définir les notes d'admissibilité, les mentions et les seuils fonctionnels	
S'il y a des mentions, comment définit-on leur seuil ?	
Si l'on ne rend compte que de l'admissibilité, comment en définit-on le seuil ?	
Comment assure-t-on la fidélité de ces standards ?	

Fiche 5 : Délivrance des résultats

Résultats	Brève description et/ou référence
Sous quelle forme se présentent les résultats délivrés aux candidats ?	<input type="checkbox"/> note globale ou échec/réussite <input type="checkbox"/> note ou échec/réussite pour chaque épreuve <input type="checkbox"/> note globale plus profil graphique <input type="checkbox"/> profil pour chaque épreuve
Sous quelle forme rend-on compte des résultats ?	<input type="checkbox"/> note brute <input type="checkbox"/> notes indéterminées (par exemple « C ») <input type="checkbox"/> niveau sur une échelle donnée <input type="checkbox"/> profils diagnostiques
Sur quel type de document fait-on connaître les résultats ?	<input type="checkbox"/> lettre ou courriel <input type="checkbox"/> rapport <input type="checkbox"/> certificat/diplôme
Donne-t-on des renseignements aux candidats pour les aider à interpréter les résultats ? Détaillez	
Les candidats ont-ils le droit de voir leurs copies corrigées et notées ?	
Les candidats ont-ils le droit de demander une nouvelle correction ?	

Fiche 6 : Analyse et révision de l'examen

Analyse de l'examen et révision après passation	Brève description et/ou référence
Recueille-t-on des retours d'information sur l'examen ?	<input type="checkbox"/> oui <input type="checkbox"/> non
Si oui, par qui ?	<input type="checkbox"/> experts internes (collègues) <input type="checkbox"/> experts extérieurs <input type="checkbox"/> organismes locaux d'évaluation <input type="checkbox"/> personnel administrant l'examen <input type="checkbox"/> enseignants <input type="checkbox"/> candidats
Tient-on compte du retour d'information pour les versions révisées de l'examen ?	<input type="checkbox"/> oui <input type="checkbox"/> non
Recueille-t-on des données pour procéder à des analyses sur les examens ?	<input type="checkbox"/> sur tous les examens <input type="checkbox"/> sur un échantillon de candidats : combien : _____ ; combien de fois : _____ <input type="checkbox"/> non
Si oui, dites comment on recueille les données.	<input type="checkbox"/> pendant les prétests <input type="checkbox"/> pendant la passation de l'examen <input type="checkbox"/> après la passation
Pour quelles caractéristiques fait-on une analyse des données recueillies ?	<input type="checkbox"/> la difficulté <input type="checkbox"/> la discrimination <input type="checkbox"/> la fiabilité <input type="checkbox"/> la validité
Dites quelles méthodes analytiques ont été mises en œuvre (par exemple, en termes de procédures psychométriques)	
Analyse-t-on les performances de candidats appartenant à des groupes différents ? Si oui, dites comment.	
Décrivez les moyens mis en œuvre pour garantir la confidentialité des données.	
Les concepts de mesure appropriés sont-ils expliqués aux utilisateurs du test ? Si oui, dites comment.	

Fiche A7 : Justification des décisions

Justification des décisions prises	Brève description et/ou référence
Donnez la justification des décisions prises relatives à l'examen ou aux épreuves en question.	

4.2.2. Description détaillée

Cette seconde phase entre dans le détail de la description du contenu de l'examen. Il est préférable de traiter cette partie en collaboration avec des collègues utilisant des sources d'information et d'expérience à la fois formelles et informelles. Vous devez également citer (si vous en disposez) les références de la documentation existante.

Certains examens mesurent la compétence des candidats de manière intégrée (plusieurs savoir-faire évalués dans une même épreuve) ; d'autres examens disposent d'épreuves différentes pour évaluer des différents types d'activités communicatives. Veuillez suivre les lignes directrices proposées dans la Partie 4.1 pour compléter, parmi les fiches et les tableaux ci-dessous, ceux qui sont pertinents par rapport à l'examen que vous étudiez.

NIVEAU GLOBAL

Après avoir consulté le panorama des Niveaux du CECR donné dans le Tableau 2.1 du Chapitre 2 de ce Manuel, on remplira la Fiche A8 pour indiquer à quel(s) niveau(x) se situe l'examen ou les épreuves que l'on étudie. On se fondera sur la description générale de l'examen, les besoins et les objectifs de la population cible ainsi que sur l'opinion avisée de collègues. Si on en dispose, les références de la documentation existante doivent être citées.

Impression initiale du niveau global de l'examen par rapport au 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"/>
Brève justification, référence pour de la documentation		

Fiche A8 : Impression initiale du niveau global

Les fiches concernant les Activités communicatives langagières sont présentées dans l'ordre suivant :

- Compréhension de l'oral
- Compréhension de l'écrit
- Interaction orale
- Interaction écrite
- Production orale
- Production écrite
- Combinaisons de compétences intégrées
- Compétences intégrées
- Médiation orale
- Médiation écrite

Seules les fiches pertinentes pour les Activités de communication langagière évaluées dans l'examen à l'étude seront complétées. Chaque fiche doit faire état de la comparaison des épreuves concernées avec les sous échelles les plus pertinentes du CECR. Ces échelles sont présentées ici à la suite de chaque fiche.

RECEPTION

Fiche A9 : Compréhension de l'oral

Compréhension de l'oral	Brève description et/ou référence
Dans quelles situations, catégories de contenu et domaines attend-on des candidats qu'ils prouvent leur compétence ? ➤ Le Tableau 5 dans le CECR 4.1 peut servir de référence	
Quels sont les thèmes de communication que les candidats doivent être capables de traiter ? ➤ Les listes du CECR 4.2 peuvent servir de référence	
Quelles activités communicatives les candidats doivent-ils être capables d'effectuer ? ➤ Les listes du CECR 4.3 peuvent servir de référence	
Quels types d'activités communicatives et quelles stratégies les candidats doivent-ils être capables de mettre en œuvre ? ➤ Les listes du CECR 4.4.2.1 peuvent servir de référence	
Quels types de textes et de quelle longueur attend-on que les candidats soient capables de traiter ? ➤ Les listes du CECR 4.6.2 et 4.6.3 peuvent servir de référence	

Quelles sortes de tâches les candidats doivent-ils être capables d'exécuter ? ➤ La description du CECR 7.1, 7.2 et 7.3 peut servir de référence	
Après avoir pris connaissance de l'échelle de Compréhension générale de l'oral reproduite ci-dessous, dites et justifiez à quel(s) niveau(x) de l'échelle le sous test devrait se situer. ➤ Les sous échelles de compréhension de l'oral du CECR 4.4.2.1 ainsi que celles qui s'y rapportent peuvent servir de référence	Niveau : Justification (y compris références documentaires)

COMPRÉHENSION GÉNÉRALE DE L'ORAL	
C2	Peut comprendre toute forme de langue orale qu'elle soit en direct ou à la radio et quel qu'en soit le débit.
C1	Peut suivre une intervention d'une certaine longueur sur des sujets abstraits ou complexes même hors de son domaine mais peut avoir besoin de faire confirmer quelques détails, notamment si l'accent n'est pas familier. Peut reconnaître une gamme étendue d'expressions idiomatiques et de tournures courantes en relevant les changements de registre. Peut suivre une intervention d'une certaine longueur même si elle n'est pas clairement structurée et même si les relations entre les idées sont seulement implicites et non explicitement indiquées.
B2	Peut comprendre une langue orale standard en direct ou à la radio sur des sujets familiers et non familiers se rencontrant normalement dans la vie personnelle, sociale, universitaire ou professionnelle. Seul un très fort bruit de fond, une structure inadaptée du discours ou l'utilisation d'expressions idiomatiques peuvent influencer la capacité à comprendre.
	Peut comprendre les idées principales d'interventions complexes du point de vue du fond et de la forme, sur un sujet concret ou abstrait et dans une langue standard, y compris des discussions techniques dans son domaine de spécialisation. Peut suivre une intervention d'une certaine longueur et une argumentation complexe à condition que le sujet soit assez familier et que le plan général de l'exposé soit indiqué par des marqueurs explicites.
B1	Peut comprendre une information factuelle directe sur des sujets de la vie quotidienne ou relatifs au travail en reconnaissant les messages généraux et les points de détail, à condition que l'articulation soit claire et l'accent courant.
	Peut comprendre les points principaux d'une intervention sur des sujets familiers rencontrés régulièrement au travail, à l'école, pendant les loisirs, y compris des récits courts.
A2	Peut comprendre assez pour pouvoir répondre à des besoins concrets à condition que la diction soit claire et le débit lent.
	Peut comprendre des expressions et des mots porteurs de sens relatifs à des domaines de priorité immédiate (par exemple, information personnelle et familiale de base, achats, géographie local, emploi).
A1	Peut comprendre une intervention si elle est lente et soigneusement articulée et comprend de longues pauses qui permettent d'en assimiler le sens.

Sous échelles relatives à la Compréhension de l'oral

Version française

- | | |
|--|-------|
| ➤ Comprendre une interaction entre locuteurs natifs | p. 55 |
| ➤ Comprendre en tant qu'auditeur | p. 56 |
| ➤ Comprendre des annonces et instructions orales | p. 56 |
| ➤ Comprendre des émissions de radio et des enregistrements | p. 56 |
| ➤ Comprendre des émissions de télévision et des films | p. 59 |
| ➤ Reconnaître des indices et faire des déductions | p. 60 |

Fiche A10 : Compréhension de l'écrit

Compréhension de l'écrit	Brève description et/ou référence
Dans quelles situations, catégories de contenu et domaines attend-on des candidats qu'ils prouvent leur compétence ? ➤ Le Tableau 5 dans le CECR 4.1 peut servir de référence	
Quels sont les thèmes de communication que les candidats doivent être capables de traiter ? ➤ Les listes du CECR 4.2 peuvent servir de référence	
Quelles activités communicatives les candidats doivent-ils être capables d'effectuer ? ➤ Les listes du CECR 4.3 peuvent servir de référence	
Quels types d'activités communicatives et quelles stratégies les candidats doivent-ils être capables de mettre en œuvre ? ➤ Les listes du CECR 4.4.2.2 peuvent servir de référence	
Quels types de textes et quelle longueur de texte attend-on que les candidats soient capables de traiter ? ➤ Les listes du CECR 4.6.2 et 4.6.3 peuvent servir de référence	
Quelles sortes de tâches les candidats doivent-ils être capables d'exécuter ? ➤ La description du CECR 7.1, 7.2 et 7.3 peut servir de référence	
Après avoir pris connaissance de l'échelle de Compréhension générale de l'écrit reproduite ci-dessous, dites et justifiez à quel(s) niveau(x) de l'échelle le sous test devrait se situer. ➤ Les sous échelles de compréhension de l'écrit du CECR 4.4.2.2 ainsi que celles qui s'y rapportent peuvent servir de référence	Niveau : Justification (y compris références documentaires)

COMPRÉHENSION GÉNÉRALE DE L'ÉCRIT	
C2	Peut comprendre et interpréter de façon critique presque toute forme d'écrit, y compris des textes (littéraires ou non) abstraits et structurellement complexes ou très riches en expressions familières Peut comprendre une gamme étendue de textes longs et complexes en appréciant de subtiles distinctions de style et le sens implicite autant qu'explicite.
C1	Peut comprendre dans le détail des textes longs et complexes, qu'ils se rapportent ou non à son domaine, à condition de pouvoir relire les parties difficiles.
B2	Peut lire avec un grand degré d'autonomie en adaptant le mode et la rapidité de lecture à différents textes et objectifs et en utilisant les références convenables de manière sélective. Possède un vocabulaire de lecture large et actif mais pourra avoir des difficultés avec des expressions peu fréquentes
B1	Peut lire des textes factuels directs sur des sujets relatifs à son domaine et à ses intérêts avec un niveau satisfaisant de compréhension.
A2	Peut comprendre de courts textes simples sur des sujets concrets courants avec une fréquence élevée de langue quotidienne ou relative au travail
	Peut comprendre des textes courts et simples contenant un vocabulaire extrêmement fréquent, y compris un vocabulaire internationalement partagé
A1	Peut comprendre des textes très courts et très simples, phrase par phrase, en relevant des noms, des mots familiers et des expressions très élémentaires et en relisant si nécessaire.

Sous échelles relatives à la Compréhension de l'écrit

Version française

- Comprendre la correspondance p. 58
- Lire pour s'orienter p. 58
- Lire pour s'informer et discuter p. 58
- Lire des instructions p. 59
- Reconnaître des indices et faire des déductions p. 60

INTERACTION

Fiche A11 : Interaction orale

Interaction orale	Brève description et/ou référence
<p>Dans quelles situations, catégories de contenu et domaines attend-on des candidats qu'ils prouvent leur compétence ?</p> <ul style="list-style-type: none"> ➤ Le Tableau 5 dans le CECR 4.1 peut servir de référence 	
<p>Quels sont les thèmes de communication que les candidats doivent être capables de traiter ?</p> <ul style="list-style-type: none"> ➤ Les listes du CECR 4.2 peuvent servir de référence 	
<p>Quelles activités communicatives les candidats doivent-ils être capables d'effectuer ?</p> <ul style="list-style-type: none"> ➤ Les listes du CECR 4.3 peuvent servir de référence 	
<p>Quels types d'activités communicatives et quelles stratégies d'interaction les candidats doivent-ils être capables de mettre en œuvre ?</p> <ul style="list-style-type: none"> ➤ Les listes du CECR 4.4.3.1 et 4.4.3.5 peuvent servir de référence 	
<p>Quels textes et types de textes attend-on que les candidats soient capables de traiter ?</p> <ul style="list-style-type: none"> ➤ Les listes du CECR 4.6.2 et 4.6.3 peuvent servir de référence 	
<p>Quelles sortes de tâches les candidats doivent-ils être capables d'exécuter ?</p> <ul style="list-style-type: none"> ➤ La description du CECR 7.1, 7.2 et 7.3 peut servir de référence 	
<p>Après avoir pris connaissance de l'échelle de l'Interaction orale générale reproduite ci-dessous, dites et justifiez à quel(s) niveau(x) de l'échelle le sous test devrait se situer.</p> <ul style="list-style-type: none"> ➤ Les sous échelles d'interaction orale du CECR 4.4.3.1 ainsi que celles qui s'y rapportent peuvent servir de référence 	<p>Niveau :</p> <p>Justification (y compris références documentaires)</p>

INTERACTION ORALE GÉNÉRALE	
C2	Possède une bonne maîtrise d'expressions idiomatiques et de tournures courantes, avec une conscience du sens connotatif. Peut exprimer avec précision des nuances fines de signification, en utilisant assez correctement une gamme étendue de modalités. Peut revenir sur une difficulté et la restructurer de manière si habile que l'interlocuteur s'en rende à peine compte.
C1	Peut s'exprimer avec aisance et spontanéité, presque sans effort.. Possède une bonne maîtrise d'un vaste répertoire lexical lui permettant de surmonter facilement des lacunes par des périphrases avec apparemment peu de recherche d'expressions ou de stratégies d'évitement. Seul un sujet conceptuellement difficile est susceptible de gêner le flot naturel et fluide du discours.
B2	Peut utiliser la langue avec aisance, correction et efficacité dans une gamme étendue de sujets d'ordre général, éducationnel, professionnel et concernant les loisirs, en indiquant clairement les relations entre les idées. Peut communiquer spontanément avec un bon contrôle grammatical sans donner l'impression d'avoir à restreindre ce qu'il/elle souhaite dire et avec le degré de formalisme adapté à la circonstance
	Peut communiquer avec un niveau d'aisance et de spontanéité tel qu'une interaction soutenue avec des locuteurs natifs sera tout à fait possible sans entraîner de tension d'une part ni de l'autre. Peut mettre en valeur la signification personnelle de faits et d'expériences, exposer ses opinions et les défendre avec pertinence en fournissant explications et arguments.
B1	Peut communiquer avec une certaine assurance sur des sujets familiers habituels ou non en relation avec ses intérêts et son domaine professionnel. Peut échanger, vérifier et confirmer des informations, faire face à des situations moins courantes et expliquer pourquoi il y a une difficulté. Peut exprimer sa pensée sur un sujet abstrait ou culturel comme un film, des livres, de la musique, etc.
	Peut exploiter avec souplesse une gamme étendue de langue simple pour faire face à la plupart des situations susceptibles de se produire au cours d'un voyage. Peut aborder sans préparation une conversation sur un sujet familier, exprimer des opinions personnelles et échanger de l'information sur des sujets familiers, d'intérêt personnel ou pertinents pour la vie quotidienne (par exemple, la famille, les loisirs, le travail, les voyages et les faits divers)
A2	Peut interagir avec une aisance raisonnable dans des situations bien structurées et de courtes conversations à condition que l'interlocuteur apporte de l'aide le cas échéant.. Peut faire face à des échanges courants simples sans effort excessif ; peut poser des questions, répondre à des questions et échanger des idées et des renseignements sur des sujets familiers dans des situations familières prévisibles de la vie quotidienne.
	Peut communiquer dans le cadre d'une tâche simple et courante ne demandant qu'un échange d'information simple et direct sur des sujets familiers relatifs au travail et aux loisirs. Peut gérer des échanges de type social très courts mais est rarement capable de comprendre suffisamment pour alimenter volontairement la conversation..
A1	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 répondre à des questions simples et en poser, réagir à des affirmations simples et en émettre dans le domaine des besoins immédiats ou sur des sujets très familiers.

Sous échelles relatives à l'Interaction orale

Version française

- Comprendre un locuteur natif p. 62
- Conversation p. 62
- Discussion informelle p. 63
- Discussions et réunions formelles p. 64
- Coopération à visée fonctionnelle p. 65
- Obtenir des biens et des services p. 66
- Echange d'information p. 67
- Interviewer et être interviewé p. 68

Fiche A12 : Interaction écrite

Interaction écrite	Brève description et/ou référence
Dans quelles situations, catégories de contenu et domaines, attend-on des candidats qu'ils prouvent leur compétence ? ➤ Le Tableau 5 dans le CECR 4.1 peut servir de référence	
Quels sont les thèmes de communication que les candidats doivent être capables de traiter ? ➤ Les listes du CECR 4.2 peuvent servir de référence	
Quelles activités communicatives les candidats doivent-ils être capables d'effectuer ? ➤ Les listes du CECR 4.3 peuvent servir de référence	
Quels types d'activités communicatives et quelles stratégies les candidats doivent-ils être capables de mettre en œuvre ? ➤ Les listes du CECR 4.4.3.1 peuvent servir de référence	
Quels textes et types de textes attend-on que les candidats soient capables de traiter ? ➤ Les listes du CECR 4.6.2 et 4.6.3 peuvent servir de référence	
Quelles sortes de tâches les candidats doivent-ils être capables d'exécuter ? ➤ La description du CECR 7.1, 7.2 et 7.3 peut servir de référence	
Après avoir pris connaissance de l'échelle de l'Interaction écrite générale reproduite ci-dessous, dites et justifiez à quel(s) niveau(x) de l'échelle le sous test devrait se situer. ➤ Les sous échelles d'interaction écrite du CECR 4.4.3.4 énumérées à la suite de l'échelle peuvent servir de référence	Niveau :
	Justification (y compris références documentaires)

INTERACTION ÉCRITE GÉNÉRALE	
C2	<i>Comme C1</i>
C1	Peut s'exprimer avec clarté et précision, en s'adaptant au destinataire avec souplesse et efficacité
B2	Peut relater des informations et exprimer des points de vue par écrit et s'adapter à ceux des autres.
B1	Peut apporter de l'information sur des sujets abstraits et concrets, contrôler l'information, poser des questions sur un problème ou l'exposer assez précisément.
	Peut écrire des notes et lettres personnelles pour demander ou transmettre des informations d'intérêt immédiat et faire comprendre les points qu'il/elle considère importants.
A2	Peut écrire de brèves notes simples en rapport avec des besoins immédiats
A1	Peut demander ou transmettre par écrit des renseignements personnels détaillés

Sous échelles relatives à l'Interaction écrite

Version française

- Correspondance
- Notes, messages et formulaires

p. 83
p. 83

PRODUCTION

Fiche A13 : Production orale

Production orale	Brève description et/ou référence
Dans quelles situations, catégories de contenu et domaines attend-on des candidats qu'ils prouvent leur compétence ? ➤ Le Tableau 5 dans le CECR 4.1 peut servir de référence	
Quels sont les thèmes de communication que les candidats doivent être capables de traiter ? ➤ Les listes du CECR 4.2 peuvent servir de référence	
Quelles activités communicatives les candidats doivent-ils être capables d'effectuer ? ➤ Les listes du CECR 4.3 peuvent servir de référence	
Quels types d'activités communicatives et quelles stratégies les candidats doivent-ils être capables de mettre en œuvre ? ➤ Les listes du CECR 4.4.1.1 peuvent servir de référence	
Quels textes et types de textes attend-on que les candidats soient capables de traiter ? ➤ Les listes du CECR 4.6.2 et 4.6.3 peuvent servir de référence	
Quelles sortes de tâches les candidats doivent-ils être capables d'exécuter ? ➤ La description du CECR 7.1, 7.2 et 7.3 peut servir de référence	
Après avoir pris connaissance de l'échelle de Production orale générale reproduite ci-dessous, dites et justifiez à quel(s) niveau(x) de l'échelle le sous test devrait se situer. ➤ Les sous échelles de production orale du CECR 4.4.1.1 énumérées à la suite de l'échelle peuvent servir de référence	Niveau : Justification (y compris références documentaires)

PRODUCTION ORALE GÉNÉRALE	
C2	Peut produire un discours élaboré, limpide et fluide, avec une structure logique efficace qui aide le destinataire à remarquer les points importants et à s'en souvenir.
C1	Peut faire une présentation ou une description d'un sujet complexe en intégrant des arguments secondaires et en développant des points particuliers pour parvenir à une conclusion appropriée.
B2	Peut méthodiquement développer une présentation ou une description soulignant les points importants et les détails pertinents.
	Peut faire une description et une présentation détaillées sur une gamme étendue de sujets relatifs à son domaine d'intérêt en développant et justifiant les idées par des points secondaires et des exemples pertinents.
B1	Peut assez aisément mener à bien une description directe et non compliquée de sujets variés dans son domaine en la présentant comme une succession linéaire de points.
A2	Peut décrire ou présenter simplement des gens, des conditions de vie, des activités quotidiennes, ce qu'on aime ou pas, par de courtes séries d'expressions ou de phrases non articulées.
A1	Peut produire des expressions simples isolées sur les gens et les choses.

Sous échelles relatives à la Production orale

- Monologue suivi : décrire l'expérience
- Monologue suivi : argumenter (par exemple lors d'un débat)
- Annonces publiques
- S'adresser à un auditoire

Version française

- p. 49
- p. 50
- p. 50
- p. 50

Fiche A14 : Production écrite

Production écrite	Brève description et/ou référence
Dans quelles situations, catégories de contenu et domaines attend-on des candidats qu'ils prouvent leur compétence ? ➤ Le Tableau 5 dans le CECR 4.1 peut servir de référence	
Quels sont les thèmes de communication que les candidats doivent être capables de traiter ? ➤ Les listes du CECR 4.2 peuvent servir de référence	
Quelles activités communicatives les candidats doivent-ils être capables d'effectuer ? ➤ Les listes du CECR 4.3 peuvent servir de référence	
Quels types d'activités communicatives et quelles stratégies les candidats doivent-ils être capables de mettre en œuvre ? ➤ Les listes du CECR 4.4.1.2 peuvent servir de textes de référence	
Quels types de textes et de quelle longueur attend-on que les candidats soient capables de traiter ? ➤ Les listes du CECR 4.6.2 et 4.6.3 peuvent servir de référence	
Quelles sortes de tâches les candidats doivent-ils être capables d'exécuter ? ➤ La description du CECR 7.1, 7.2 et 7.3 peut servir de référence	
Après avoir pris connaissance de l'échelle de Production écrite générale reproduite ci-dessous, dites et justifiez à quel(s) niveau(x) de l'échelle le sous test devrait se situer. ➤ Les sous échelles de production écrite du CECR 4.4.1.2 énumérées à la suite de l'échelle peuvent servir de référence	Niveau :
	Justification (y compris références documentaires)

PRODUCTION ÉCRITE GÉNÉRALE	
C2	Peut écrire des textes élaborés, limpides et fluides, dans un style approprié et efficace, avec une structure logique qui aide le destinataire à remarquer les points importants.
C1	Peut écrire des textes bien structurés sur des sujets complexes, en soulignant les points pertinents les plus saillants et en confirmant un point de vue de manière élaborée par l'intégration d'arguments secondaires, de justifications et d'exemples pertinents pour parvenir à une conclusion appropriée.
B2	Peut écrire des textes clairs et détaillés sur une gamme étendue de sujets relatifs à son domaine d'intérêt en faisant la synthèse et l'évaluation d'informations et d'arguments empruntés à des sources diverses.
B1	Peut écrire des textes articulés simplement sur une gamme de sujets variés dans son domaine en liant une série d'éléments discrets en une séquence linéaire.
A2	Peut écrire une série d'expressions et de phrases simples reliées par des connecteurs simples tels que "et", "mais" et "parce que".
A1	Peut écrire des expressions et phrases simples isolées.

Sous échelles relatives à la Production écrite

Version française

- Ecriture créative
- Rapports et essais

p. 52
p. 52

AUTRES COMPETENCES INTEGREES

Quelles sont les combinaisons de compétences qui apparaissent dans les épreuves d'examen ? Indiquez-les dans la Fiche A14 puis remplissez un exemplaire de la Fiche A15 pour chaque combinaison.

Fiche A15 : Combinaison de compétences intégrées

Combinaisons d'aptitudes intégrées	Epreuve dans laquelle elles apparaissent
Compréhension de l'oral et prise de notes <input type="checkbox"/>	
Compréhension de l'oral et production orale <input type="checkbox"/>	
Compréhension de l'oral et production écrite <input type="checkbox"/>	
Compréhension de l'écrit et prise de notes <input type="checkbox"/>	
Compréhension de l'écrit et production orale <input type="checkbox"/>	
Compréhension de l'écrit et production écrite <input type="checkbox"/>	
Compréhension de l'oral et de l'écrit et prise de notes <input type="checkbox"/>	
Compréhension de l'oral et de l'écrit et production orale <input type="checkbox"/>	
Compréhension de l'oral et de l'écrit et production écrite <input type="checkbox"/>	

Fiche A16 : Compétences intégrées

	Répondez pour chacune des combinaisons citées ci-dessus
Compétences intégrées	Brève description et/ou référence
Quelles sont les combinaisons qui apparaissent ? ➤ Reportez-vous aux réponses données dans la Fiche A13	
Quelles sont les activités de texte à texte ? ➤ Le Tableau 6 dans le CECR 4.6.4 peut servir de référence	
Dans quelles situations, catégories de contenu et domaines attend-on des candidats qu'ils prouvent leur compétence ? ➤ Le Tableau 5 dans le CECR 4.1 peut servir de référence	
Quels sont les thèmes de communication que les candidats doivent être capables de traiter ? ➤ Les listes du CECR 4.2 peuvent servir de référence	
Quelles activités communicatives les candidats doivent-ils être capables d'effectuer ? ➤ Les listes du CECR 4.3 peuvent servir de textes de référence	
Quels types de textes attend-on que les candidats soient capables de traiter ? ➤ Les listes du CECR 4.6.2 et 4.6.3 peuvent servir de référence	
Après avoir pris connaissance de l'échelle Traiter un texte reproduite ci-dessous ainsi que des échelles Compréhension de l'oral/écrit et de Production écrite déjà données, dites et justifiez à quel(s) niveau(x) de l'échelle le sous test devrait se situer. ➤ La sous échelle Prendre des notes du CECR 4.6.2 peut servir de référence	Niveau :
	Justification (y compris références documentaires)

TRAITER UN TEXTE	
C2	Peut faire le résumé d'informations de sources diverses en recomposant les arguments et les comptes rendus dans une présentation cohérente du résultat général
C1	Peut résumer de longs textes difficiles
B2	Peut résumer un large éventail de textes factuels et de fiction en commentant et en critiquant les points de vue opposés et les thèmes principaux Peut résumer des extraits de nouvelles (information), d'entretiens ou de documentaires traduisant des opinions, les discuter et les critiquer Peut résumer l'intrigue et la suite des événements d'un film ou d'une pièce
B1	Peut collationner des éléments d'information issus de sources diverses et les résumer pour quelqu'un d'autre
	Peut paraphraser simplement de courts passages écrits en utilisant les mots et le plan du texte
A2	Peut prélever et reproduire des mots et des phrases ou de courts énoncés dans un texte court qui reste dans le cadre de sa compétence et de son expérience limitées
	Peut copier des textes courts en script ou en écriture lisible
A1	Peut copier des mots isolés et des textes courts imprimés normalement

MEDIATION

Fiche A17 : Médiation orale

Médiation orale	Brève description et/ou référence
Quelles sont les activités de texte à texte ? ➤ Le Tableau 6 dans le CECR 4.6.4 peut servir de référence	
Quelles sont les activités de médiation évaluées ? ➤ La liste du CECR 4.4.4.1 peut servir de référence	
Dans quelles situations, catégories de contenu et domaines attend-on des candidats qu'ils prouvent leur compétence ? ➤ Le Tableau 5 dans le CECR 4.1 peut servir de référence	
Quels sont les thèmes de communication que les candidats doivent être capables de traiter ? ➤ Les listes du CECR 4.2 peuvent servir de textes de référence	
Quelles activités communicatives les candidats doivent-ils être capables d'effectuer ? ➤ Les listes du CECR 4.3 peuvent servir de référence	
Quels types de textes attend-on que les candidats soient capables de traiter ? ➤ Les listes du CECR 4.6.2 et 4.6.3 peuvent servir de référence	
Le CECR ne présente pas d'échelle pour la Traduction. En généralisant à partir des échelles de Compréhension de l'oral, Traiter un texte et Production orale, dites et justifiez à quel(s) niveau(x) le sous test devrait se situer	Niveau :
	Justification (y compris références documentaires)

Fiche A18 : Médiation écrite

Médiation écrite	Brève description et/ou référence
Quelles sont les activités de texte à texte ? ➤ Le Tableau 6 dans le CECR 4.6.4 peut servir de référence	
Quelles sont les activités de médiation évaluées ? ➤ La liste du CECR 4.4.4.2 peut servir de référence	
Dans quelles situations, catégories de contenu et domaines attend-on des candidats qu'ils prouvent leur compétence ? ➤ Le Tableau 5 dans le CECR 4.1 peut servir de référence	
Quels sont les thèmes de communication que les candidats doivent être capables de traiter ? ➤ Les listes du CECR 4.2 peuvent servir de textes de référence	
Quelles activités communicatives les candidats doivent-ils être capables d'effectuer ? ➤ Les listes du CECR 4.3 peuvent servir de référence	
Quels types de textes attend-on que les candidats soient capables de traiter ? ➤ Les listes du CECR 4.6.2 et 4.6.3 peuvent servir de référence	
Le CECR ne présente pas d'échelle pour la Traduction. En généralisant à partir des échelles de Compréhension de l'écrit, Traiter un texte et Production écrite, dites et justifiez à quel(s) niveau(x) le sous test devrait se situer	Niveau :
	Justification (y compris références documentaires)

Les fiches pour les Aspects des Compétences communicatives langagières (CECR Chapitre 5) sont présentées ci-dessous. Cette partie est organisée de manière différente. Après la présentation des descripteurs du CECR relatifs aux « Paramètres qualitatifs pertinents pour la Réception » (Tableau 4.3), on demande aux utilisateurs de compléter la fiche sur les « Aspects de la compétence langagière en réception » (Fiche A19) sur la base de l'analyse des épreuves ou de l'examen étudié.

Seules les fiches pertinentes pour les Aspects de la communication évalués dans les épreuves ou l'examen en question seront complétées. Les utilisateurs compareront les épreuves concernées avec la sous échelle la plus pertinente du CECR. Ces échelles sont présentées à la suite de chaque fiche.

La même démarche sera appliquée pour l'« Interaction » en utilisant le Tableau 4.4 et la Fiche A20 ; pour la « Production » à l'aide du Tableau 4.5 et de la Fiche A21 et pour la « Médiation » à l'aide de la Fiche A22. Pour la Médiation, comme il n'y a pas d'échelle dans le CECR, on demande aux utilisateurs de se reporter aux descripteurs de Réception et de Production.

Les fiches sont présentées dans l'ordre suivant :

- Réception
- Interaction
- Production
- Médiation.

RECEPTION

Les échelles du CECR les plus pertinentes pour la Réception ont permis de créer le Tableau 4.3. Ce Tableau ne comprend pas de descripteurs pour des « niveaux plus ». Les échelles originales pertinentes dont certaines définissent des « niveaux plus » comprennent :

Compétence linguistique	
➤ Etendue linguistique générale	p. 87
➤ Etendue du vocabulaire	p. 88
Compétence sociolinguistique	
➤ Correction sociolinguistique	p. 95
Compétence pragmatique	
➤ Développement thématique	p. 97
➤ Cohérence et cohésion	p. 98
➤ Précision	p. 101
Compétence stratégique	
➤ Reconnaître des indices et faire des déductions	p. 60

TABLEAU 4.3 : FACTEURS QUALITATIFS PERTINENTS POUR LA RECEPTION

		STRATEGIQUES Reconnaitre des indices et faire des déductions		
		PRAGMATIQUES D'après « Développement thématique » et « Précision »		
		SOCIOLOGIQUES D'après « Correction sociolinguistique »		
		LINGUISTIQUES D'après « Etendue linguistique générale » et « Etendue du vocabulaire »		
C2	<p>Peut comprendre avec précision une gamme très étendue de discours, apprécier l'insistance et la discrimination. Ne montre aucun signe d'incompréhension.</p> <p>Possède une bonne maîtrise d'un vaste répertoire lexical d'expressions idiomatiques et courantes avec la conscience du niveau de connotation sémantique</p>	<p>Manifeste une bonne maîtrise des expressions idiomatiques et dialectales avec la conscience des niveaux connotatifs de sens.</p> <p>Apprécie complètement les implications sociolinguistiques et socioculturelles de la langue utilisée par les locuteurs natifs et peut réagir en conséquence.</p>	<p>Peut comprendre avec précision des nuances de sens assez fines en utilisant une gamme étendue de procédés de modalisation (par exemple, adverbess exprimant le degré d'intensité, propositions restrictives)</p> <p>Peut comprendre l'insistance et la différenciation sans ambiguïté.</p>	<p>Comme C1</p>
C1	<p>Possède une bonne maîtrise d'un vaste répertoire lexical.</p> <p>Bonne maîtrise d'expressions idiomatiques et familières.</p>	<p>Peut reconnaître un large éventail d'expressions idiomatiques et dialectales et apprécier les changements de registre ; peut devoir toutefois confirmer tel ou tel détail, en particulier si l'accent n'est pas familier.</p> <p>Peut suivre des films utilisant largement l'argot et des expressions idiomatiques.</p> <p>Peut comprendre la langue avec efficacité et souplesse dans des relations sociales, y compris pour un usage affectif, allusif ou pour plaisanter.</p> <p>Peut, avec quelque effort, suivre des discussions rapides et familières.</p>	<p>Peut comprendre des descriptions et des récits compliqués avec des thèmes secondaires et certains plus développés.</p> <p>Peut comprendre avec précision les qualificatifs des opinions et des affirmations relatifs aux degrés, par exemple, de certitude/doute, croyance/doute, similitude, etc.</p>	<p>Est habile à utiliser les indices contextuels, grammaticaux et lexicaux pour en déduire une attitude, une humeur, des intentions et anticiper la suite.</p>
B2	<p>Possède une gamme assez étendue de langue pour comprendre des descriptions, des points de vue et des arguments sur la plupart des sujets pertinents pour sa vie quotidienne tels que la famille, les loisirs et centres d'intérêt, le travail, les voyages et l'actualité.</p>		<p>Peut comprendre une description ou un récit, reconnaître les points saillants, des détails et des exemples.</p> <p>Peut comprendre une information détaillée de façon fiable.</p>	<p>Peut utiliser différentes stratégies de compréhension dont l'écoute des points forts et le contrôle de la compréhension par les indices textuels.</p>
B1	<p>Possède suffisamment de moyens linguistiques pour se débrouiller et un vocabulaire suffisant pour comprendre la plupart des textes sur des sujets tels que la famille, les loisirs et centres d'intérêt, le travail, les voyages et l'actualité.</p>	<p>Peut répondre à un large éventail de fonctions langagières en utilisant leurs expressions les plus courantes de manière neutre.</p> <p>Peut reconnaître les règles de politesse importantes.</p> <p>Est conscient des différences les plus significatives entre les coutumes, les usages, les attitudes, les valeurs et les croyances qui prévalent dans la communauté concernée et celles de sa propre communauté et en recherche les indices.</p>	<p>Peut, avec une exactitude relative, comprendre un récit ou une description linéaire.</p> <p>Peut comprendre les points principaux d'une idée ou d'un problème avec une certaine précision.</p>	<p>Peut identifier des mots inconnus à l'aide du contexte sur des sujets relatifs à son domaine et à ses intérêts.</p> <p>Peut extrapoler du contexte le sens de mots inconnus et en déduire le sens de la phrase à condition que le sujet en question soit familier.</p>
A2	<p>Possède un vocabulaire suffisant pour se débrouiller dans des situations courantes au contenu prévisible et pour répondre à des besoins simples de type concret.</p>	<p>Peut se débrouiller dans des échanges sociaux très courts en utilisant les formes quotidiennes polies d'accueil et de contact. Peut faire des invitations, des excuses et y répondre.</p>	<p>Peut comprendre une histoire ou une description consistant en une succession de points.</p> <p>Peut comprendre un échange d'information limité, simple et direct sur des sujets familiers et habituels.</p>	<p>Peut utiliser le sens général d'un texte ou d'un énoncé court sur des sujets quotidiens concrets pour déduire du contexte le sens probable de mots inconnus.</p>
A1	<p>Possède un choix élémentaire d'expressions simples pour les informations sur soi et les besoins de type courant.</p>	<p>Peut comprendre les formes de politesse les plus élémentaires d'accueil et prise de congé, de présentation ; dire merci, s'il vous plaît, excusez-moi, etc.</p>	<p>Pas de descripteur disponible</p>	<p>Pas de descripteur disponible</p>

Fiche A19 : Aspects de la compétence langagière pour la réception

Compétence linguistique	Brève description et/ou référence
<p>Quelle étendue de la compétence lexicale et grammaticale attend-on que les candidats soient capables de maîtriser ?</p> <p>➤ Les listes du CECR 5.2.1.1 et 5.2.1.2 peuvent servir de référence</p>	
<p>Après avoir pris connaissance de l'échelle de Compétence linguistique du Tableau 4.3, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer</p>	Niveau
	Justification (y compris références documentaires)
Compétence sociolinguistique	Brève description et/ou référence
<p>Quelles compétences sociolinguistiques attend-on que les candidats soient capables de mettre en œuvre : marqueurs linguistiques, règles de politesse, adéquation des registres, dialectes/accent, etc. ?</p> <p>➤ Les listes du CECR 5.2.2 peuvent servir de référence</p>	
<p>Après avoir pris connaissance de l'échelle de Compétence sociolinguistique du Tableau 4.3, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.</p>	Niveau
	Justification (y compris références documentaires)
Compétence pragmatique	Brève description et/ou référence
<p>Quelles compétences pragmatiques attend-on que les candidats soient capables de mettre en œuvre : compétences discursives, fonctionnelles ?</p> <p>➤ Les listes du CECR 5.2.3 peuvent servir de référence</p>	
<p>Après avoir pris connaissance de l'échelle de Compétence pragmatique du Tableau 4.3, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.</p>	Niveau
	Justification (y compris références documentaires)
Compétence stratégique	Brève description et/ou référence
<p>Quelles compétences stratégiques attend-on que les candidats soient capables d'utiliser ?</p> <p>➤ Les listes du CECR 4.4.2.4 peuvent servir de référence</p>	
<p>Après avoir pris connaissance de l'échelle de Compétence stratégique du Tableau 4.3, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.</p>	Niveau
	Justification (y compris références documentaires)

INTERACTION

Les échelles du CECR les plus appropriées pour l'Interaction ont permis de créer le Tableau 4.4 auquel on peut se reporter dans cette partie. Le Tableau 5 ne comporte pas de descripteurs de « niveaux plus ». Les échelles originales pertinentes dont certaines définissent des « niveaux plus » comprennent :

Compétence linguistique

- Etendue linguistique générale p. 87
- Etendue du vocabulaire p. 88
- Maîtrise du vocabulaire p. 89
- Correction grammaticale p. 90

Compétence sociolinguistique

- Correction sociolinguistique p. 95

Compétence pragmatique

- Souplesse p. 97
- Tours de parole p. 97
- Aisance à l'oral p. 101
- Précision p. 101

Compétence stratégique

- Tours de parole p. 70
- Coopérer p. 71
- Faire clarifier p. 71
- Compensation p. 54
- Contrôle et correction p. 54

Fiche A20 : Aspects de la compétence langagière en interaction

Compétence linguistique	Brève description et/ou référence
<p>Quelle étendue de la compétence lexicale et grammaticale attend-on que les candidats soient capables de maîtriser ?</p> <ul style="list-style-type: none"> ➤ Les listes du CECR 5.2.1.1 et 5.2.1.2 peuvent servir de référence 	
<p>Quelle étendue de la compétence phonologique et orthographique attend-on que les candidats soient capables d'utiliser ?</p> <ul style="list-style-type: none"> ➤ Les listes du CECR 5.2.1.4 et 5.2.1.5 peuvent servir de référence 	
<p>Après avoir pris connaissance des échelles « Etendue » et « Correction » du Tableau 4.4, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.</p> <ul style="list-style-type: none"> ➤ Les échelles pour la Maîtrise du système phonologique du CECR 5.2.1.4 et pour la Maîtrise de l'orthographe (5.2.1.4) peuvent aussi servir de référence 	<p>Niveau</p> <p>Justification (y compris références documentaires)</p>
Compétence sociolinguistique	Brève description et/ou référence
<p>Quelles compétences sociolinguistiques attend-on que les candidats soient capables de mettre en œuvre : marqueurs linguistiques, règles de politesse, adéquation des registres, dialectes/accent, etc. ?</p> <ul style="list-style-type: none"> ➤ Les listes du CECR 5.2.2 peuvent servir de référence 	
<p>Après avoir pris connaissance de l'échelle de Compétence sociolinguistique du Tableau 4.4, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.</p>	<p>Niveau</p> <p>Justification (y compris références documentaires)</p>
Compétence pragmatique	Brève description et/ou référence
<p>Quelles compétences pragmatiques attend-on que les candidats soient capables de mettre en œuvre : compétences discursives, fonctionnelles ?</p> <ul style="list-style-type: none"> ➤ Les listes du CECR 5.2.3 peuvent servir de référence 	
<p>Après avoir pris connaissance de l'échelle pour l'Aisance du Tableau 4.4, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.</p>	<p>Niveau</p> <p>Justification (y compris références documentaires)</p>
Compétence stratégique	Brève description et/ou référence
<p>Quelles compétences stratégiques attend-on que les candidats soient capables d'utiliser ?</p> <ul style="list-style-type: none"> ➤ Le débat du CECR 4.4.3.5 peut servir de référence 	
<p>Après avoir pris connaissance de l'échelle pour l'Interaction du Tableau 4.4, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.</p>	<p>Niveau</p> <p>Justification (y compris références documentaires)</p>

PRODUCTION

Les échelles du CECR les plus appropriées pour la Production ont permis de créer le Tableau 4.5 auquel on peut se reporter dans cette partie. Le Tableau 6 ne comporte pas de descripteurs de « niveaux plus ». Les échelles originales pertinentes dont certaines définissent des niveaux plus sont :

Compétence linguistique

- Etendue linguistique générale p. 87
- Etendue du vocabulaire p. 88
- Maîtrise du vocabulaire p. 89
- Correction grammaticale p. 90

Compétence sociolinguistique

- Correction sociolinguistique p. 95

Compétence pragmatique

- Souplesse p. 97
- Développement thématique p. 97
- Cohérence et cohésion p. 97
- Aisance à l'oral p. 101
- Précision p. 101

Compétence stratégique

- Planification p. 53
- Compensation p. 54
- Contrôle et correction p. 54

Fiche A21 : Aspects de la compétence langagière en production

Compétence linguistique	Brève description et/ou référence
Quelle étendue de la compétence lexicale et grammaticale attend-on que les candidats soient capables de maîtriser ? ➤ Les listes du CECR 5.2.1.1 et 5.2.1.2 peuvent servir de référence	
Quelle étendue de la compétence phonologique et orthographique attend-on que les candidats soient capables d'utiliser ? Les listes du CECR 5.2.1.4 et 5.2.1.5 peuvent servir de référence	
Après avoir pris connaissance des échelles de Etendue et Correction du Tableau 4.4, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer. ➤ Les échelles pour la Maîtrise du système phonologique en CECR 5.2.1.4 et pour la Maîtrise de l'orthographe en 5.2.1.4 peuvent aussi servir de référence	Niveau
	Justification (y compris références documentaires)
Compétence sociolinguistique	Brève description et/ou référence
Quelles compétences sociolinguistiques attend-on que les candidats soient capables de mettre en œuvre : marqueurs linguistiques, règles de politesse, adéquation des registres, dialectes/accent, etc. ? ➤ Les listes du CECR 5.2.2 peuvent servir de référence	
Après avoir pris connaissance de l'échelle de Compétence sociolinguistique du Tableau 4.4, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.	Niveau
	Justification (y compris références documentaires)

Compétence pragmatique	Brève description et/ou référence
<p>Quelles compétences pragmatiques attend-on que les candidats soient capables de mettre en œuvre : compétences discursives, fonctionnelles ?</p> <ul style="list-style-type: none"> ➤ Les listes du CECR 5.2.3 peuvent servir de référence 	
<p>Après avoir pris connaissance de l'échelle pour la Compétence pragmatique du Tableau 4.4, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.</p>	Niveau
	Justification (y compris références documentaires)
Compétence stratégique	Brève description et/ou référence
<p>Quelles compétences stratégiques attend-on que les candidats soient capables d'utiliser ?</p> <ul style="list-style-type: none"> ➤ Le débat du CECR 4.4.1.3 peut servir de référence 	
<p>Après avoir pris connaissance de l'échelle de Compétence stratégique du Tableau 4.4, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.</p>	Niveau
	Justification (y compris références documentaires)

Fiche A21 : Aspects de la compétence langagière en Production

TABLEAU 4.4 : PARAMETRES QUALITATIFS PERTINENTS POUR L'INTERACTION

	ETENDUE LINGUISTIQUE D'après « Etendue linguistique générale », « Etendue du vocabulaire », « Souplesse »	CORRECTION LINGUISTIQUE D'après « Correction grammaticale » et « Maîtrise du vocabulaire »	SOCIOLOGUE D'après « Correction sociolinguistique »	AISANCE Aisance, Souplesse	INTERACTION D'après « Tours de parole », « Coopérer », « Faire clarifier »
C2	Montre une grande souplesse dans la reformulation d'idées en les présentant sous des formes linguistiques variées pour accentuer l'importance, marquer une différence et lever l'ambiguïté. Possède aussi une bonne maîtrise d'un répertoire d'expressions idiomatiques et courantes	Peut maintenir constamment un niveau élevé de correction grammaticale même lorsque l'attention se porte ailleurs (par exemple, la planification ou l'observation des réactions de l'autre)	Apprécie complètement les implications sociolinguistiques et socioculturelles de la langue utilisée par les locuteurs natifs et peut réagir en conséquence. Peut jouer efficacement le rôle de médiateur entre les locuteurs de la langue cible et de celle de sa communauté d'origine en tenant compte des différences socioculturelles et sociolinguistiques	Peut s'exprimer longuement dans un discours naturel et sans effort en évitant ou en contournant les difficultés de sorte que l'interlocuteur ne s'en rend pas compte	Peut intervenir habilement et avec facilité en utilisant des expressions non-verbales ou l'intonation apparemment sans effort. Peut relier naturellement sa propre contribution à celle d'autres interlocuteurs en prenant la parole à son tour, faisant des références et des allusions, etc.
C1	Possède une bonne maîtrise d'un répertoire lui permettant de choisir la façon de s'exprimer clairement de manière appropriée sur un large éventail de sujets académiques, professionnels ou de loisirs sans restrictions sur ce qu'il/elle veut dire	Peut maintenir un niveau élevé de correction grammaticale ; les erreurs sont rares, difficiles à repérer et généralement corrigées rétrospectivement	Peut utiliser la langue avec efficacité et souplesse dans des relations sociales, y compris pour un usage affectif, allusif ou pour plaisanter	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 la garder et relier habilement sa propre contribution à celles de ses interlocuteurs
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 et en utilisant des phrases complexes	A un niveau relativement élevé de correction grammaticale. Ne fait pas de fautes conduisant à des malentendus et peut corriger la plupart de ses fautes	Peut, avec quelque effort, comprendre et participer à des échanges dans un groupe même si le discours est rapide et familier. Peut poursuivre une relation suivie avec des locuteurs natifs sans les amuser ou les irriter sans le vouloir ou les mettre en situation de se comporter autrement qu'avec un locuteur natif	Peut s'adapter aux changements de sujet, de style et de ton rencontrés normalement dans une conversation. Peut parler relativement longtemps avec un débit assez régulier ; bien qu'il/elle puisse hésiter pour chercher tournures et expressions, on remarque peu de longues pauses	Peut commencer un discours, prendre la parole au bon moment et terminer la conversation quand il/elle le souhaite bien que parfois sans élégance. Peut faciliter le développement de la discussion sur un terrain connu en confirmant sa compréhension, en invitant les autres à participer, etc.
B1	Possède suffisamment de moyens linguistiques 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	Peut se servir avec une correction suffisante d'un répertoire de tournures et expressions fréquemment utilisées et associées à des situations plutôt prévisibles	Peut s'exprimer et répondre aux fonctions langagières de base telles que l'échange d'information et la demande et exprimer simplement une idée et une opinion. Est conscient des règles de politesse importantes et se conduit de manière appropriée	Peut exploiter avec souplesse une gamme étendue de langue simple afin d'exprimer l'essentiel de ce qu'il/elle veut dire. Peut s'exprimer avec une certaine aisance. Malgré quelques problèmes de formulation ayant pour conséquence pauses et impasses, est effectivement capable de continuer à parler sans aide	Peut commencer, poursuivre et terminer une simple conversation en tête-à-tête sur des sujets familiers ou d'intérêt personnel. Peut reformuler en partie les dires de l'interlocuteur pour confirmer une compréhension mutuelle
A2	Peut utiliser des modèles de phrases élémentaires et communiquer des informations limitées dans des situations courantes de la vie quotidienne à l'aide de phrases mémorisées, de groupes de mots et d'expressions toutes faites	Peut utiliser des structures simples correctement mais commet encore systématiquement des erreurs élémentaires	Peut se débrouiller dans des échanges sociaux très courts en utilisant les formes quotidiennes polies d'accueil et de contact. Peut faire des invitations, des excuses et y répondre	Peut se faire comprendre dans une brève intervention même si la reformulation, les pauses et les faux démarrages sont très évidents. Peut développer des expressions apprises par la simple recombinaison de leurs éléments	Peut indiquer qu'il/elle suit ce qui se dit mais est rarement en mesure de comprendre suffisamment pour poursuivre la conversation. Peut attirer l'attention
A1	Possède un répertoire élémentaire de mots isolés et d'expressions simples relatives à soi et à des situations concrètes particulières	A un contrôle limité de structures syntaxiques et de formes grammaticales appartenant à un répertoire mémorisé	Peut établir un contact social de base en utilisant les formes de politesse les plus élémentaires : accueil et prise de congé ; présentation et dire merci, s'il vous plaît, excusez-moi, etc.	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 intervenir simplement mais la communication repose entièrement sur la répétition, la reformulation et la remédiation

TABLEAU 4.5 : PARAMETRES QUALITATIFS PERTINENTS POUR LA PRODUCTION

	ETENDUE LINGUISTIQUE D'après « Etendue linguistique générale », « Etendue du vocabulaire »	CORRECTION LINGUISTIQUE D'après « Correction grammaticale », « Maîtrise du vocabulaire », « Maîtrise du système phonologique »	SOCIOLINGUISTIQUE D'après « Correction sociolinguistique »	PRAGMATIQUE Aisance à l'oral, Souplesse	PRAGMATIQUE Développement thématique, Précision, Cohérence et cohésion	STRATEGIQUE Compensation, Contrôle et correction
C2	Possède une grande souplesse pour reformuler les idées de différentes manières afin d'exprimer avec précision des nuances fines de sens pour insister, discriminer et lever l'ambiguïté. Possède également une bonne maîtrise d'expressions idiomatiques et courantes	Peut maintenir constamment un niveau élevé de correction grammaticale même lorsque l'attention se porte ailleurs (par exemple, la planification ou l'observation des réactions de l'autre)	Apprécie complètement les implications socioculturelles de la langue utilisée par les autres locuteurs et peut réagir en conséquence	Peut s'exprimer longuement avec spontanéité dans une langue courante, en évitant ou contournant les difficultés de telle sorte que l'interlocuteur ne s'en tienne pas compte	Peut créer un texte cohérent et cohésif en utilisant de manière complète et appropriée les structures organisationnelles adéquates et une grande variété d'articulateurs	Peut substituer à un mot qui lui échappe un terme équivalent de manière si habile que l'on s'en rende à peine compte
C1	Possède une bonne maîtrise d'une vaste étendue de langue lui permettant de choisir la formulation appropriée pour s'exprimer clairement de manière appropriée sur une gamme importante de sujets généraux, académiques, professionnels ou sur les loisirs sans avoir à restreindre ce qu'il/elle veut dire	Peut maintenir constamment un niveau élevé de correction grammaticale ; les erreurs sont rares, difficiles à repérer et généralement corrigées aussitôt	Peut utiliser la langue avec efficacité et souplesse dans des relations sociales, y compris pour un usage affectif, allusif ou pour plaisanter	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 produire un texte clair, fluide et bien structuré, démontrant un usage contrôlé de moyens linguistiques de structuration et d'articulation. Peut faire des descriptions et des récits compliqués avec des thèmes secondaires et certains plus développés et arriver à une conclusion adéquate	Peut contourner une difficulté rencontrée et reformuler ce qu'il/elle veut dire sans interrompre complètement le fil du discours
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 et en utilisant des phrases complètes	Possède un niveau relativement élevé de correction grammaticale. Ne fait pas de fautes conduisant à des malentendus et peut corriger la plupart d'entre elles rétrospectivement	Peut s'exprimer de façon appropriée à la situation et éviter des erreurs grossières de formulation	Peut parler relativement longtemps avec un débit assez régulier ; bien qu'il/elle puisse hésiter en cherchant tournures et expressions, on remarque peu de longues pauses	Peut faire une description ou un récit clair en développant et argumentant les points importants à l'aide de détails et d'exemples significatifs. Peut utiliser un nombre limité d'articulateurs pour relier ses énoncés en un discours clair et cohérent bien qu'il puisse y avoir quelques « sauts » dans une longue intervention	Peut utiliser des périphrases et des paraphrases pour dissimuler des lacunes lexicales et structurales. Peut relever ses erreurs habituelles et surveiller consciemment son discours afin de les corriger
B1	Possède suffisamment de moyens linguistiques et d'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	Peut se servir avec une correction suffisante d'un répertoire de tournures et d'expressions fréquemment utilisées et associées à des situations plutôt prévisibles	Pas de descripteur disponible	Peut exploiter avec souplesse une gamme étendue de langue simple afin d'exprimer l'essentiel de ce qu'il/elle veut dire. Peut discourir de manière compréhensible même si les mots pour chercher ses mots et ses phrases et pour faire ses corrections sont évidentes, notamment dans les séquences plus longues de production libre	Peut relier une série d'éléments courts, simples et distincts afin de raconter ou de décrire, avec une relative aisance, quelque chose de simple et de linéaire	Peut utiliser un mot simple signifiant quelque chose de semblable au concept recherché et solliciter une « correction ».
A2	Peut utiliser des modèles de phrases élémentaires et communiquer à l'aide de phrases mémorisées de groupes de mots et d'expressions toutes faites pour transmettre des informations limitées sur de simples situations quotidiennes	Peut utiliser correctement des structures simples mais commet encore systématiquement des erreurs élémentaires	Pas de descripteur disponible	Peut se faire comprendre dans une brève intervention, même si la reformulation, les pauses et les faux démarrages sont très évidents. Peut développer des expressions apprises par la simple recombinaison de leurs éléments	Peut relier des groupes de mots avec des connecteurs simples tels que « et », « mais » et « parce que »	Pas de descripteur disponible
A1	Possède un répertoire élémentaire de mots et d'expressions élémentaires relatifs à soi et à des situations concrètes particulières	A un contrôle limité de structures syntaxiques et de formes grammaticales simples appartenant à un répertoire mémorisé	Pas de descripteur disponible	Peut se débrouiller avec des énoncés très courts, isolés, généralement stéréotypés, avec de nombreuses pauses pour cher ses mots, pour prononcer les moins familiers et pour remédier à la communication	Peut relier des groupes de mots avec des connecteurs très élémentaires tels que « et » ou « alors »	Pas de descripteur disponible

MEDIATION

C'est de la nature de la médiation que vont dépendre les échelles du CECR qui seront le plus appropriées.

En situation de langue étrangère, on met naturellement l'accent sur savoir-faire en langues. Pour des activités de médiation effectuées à partir de la langue étrangère vers la langue maternelle, les savoir-faire requis seront essentiellement du domaine de la réception tandis que pour la médiation effectuée à partir de la langue maternelle vers la langue étrangère, ce sont les savoir-faire en production qui seront nécessaires. En ce qui concerne la Médiation entièrement en langue étrangère, on fera appel à des savoir-faire à la fois de réception et de production.

Variables	Type de compétence	Descripteurs
a. dans une langue étrangère	Pour la Réception et la Production	Tableaux 3 et 5
b. d'une langue étrangère à une autre	Pour la Réception et la Production	Tableaux 3 et 5
c. d'une langue étrangère vers la langue maternelle	Pour la Réception	Tableau 4.3
d. de la langue maternelle vers la langue étrangère	Pour la Production	Tableau 4.5

Les autres paramètres à prendre en considération sont les variables de la compétence (de la réception de l'oral ou de l'écrit vers la production orale ou écrite) et les variables des activités de médiation – selon un registre formel ou informel – comme indiqué dans le CECR 4.4.4.1 (médiation orale) et 4.4.4.2 (médiation écrite).

Ainsi, bien qu'il n'y ait pas de descripteurs pour la Médiation en tant que telle dans le CECR, toutes les échelles de descripteurs du Chapitre 5 du CECR auxquelles s'ajoutent les échelles pour les Stratégies de Réception et de Production (incluses respectivement dans les Tableaux 3 et 5) sont pertinentes.

Si l'examen comprend de la Médiation, consulter les Tableaux 3, 4 et/ou 5 pour remplir la Fiche A24

MEDIATION

Fiche A22 : Aspects de la compétence langagière en médiation

Compétence linguistique	Brève description et/ou référence
Quelle étendue de la compétence lexicale et grammaticale attend-on que les candidats soient capables de maîtriser ? ➤ Les listes du CECR 5.2.1.1 et 5.2.1.2 peuvent servir de référence	
Quel type de relations sémantiques attend-on que les candidats soient capables de traiter ? ➤ Les listes du CECR 5.2.1.4 peuvent servir de référence	
Quelle étendue de la compétence phonologique et orthographique attend-on que les candidats soient capables de maîtriser ? ➤ Les listes du CECR 5.2.1.4 et 5.2.1.5 peuvent servir de référence	

<ul style="list-style-type: none"> ➤ L'échelle pour la Maîtrise de l'orthographe (CECR 5.2.1.5) peut également servir de référence 	Niveau
	Justification (y compris références documentaires)
Compétence sociolinguistique	Brève description et/ou référence
Quelles compétences sociolinguistiques attend-on que les candidats soient capables de mettre en œuvre : marqueurs linguistiques, règles de politesse, adéquation des registres, dialectes/accent, etc. ? <ul style="list-style-type: none"> ➤ Les listes du CECR 5.2.2 peuvent servir de référence 	
Après avoir pris connaissance de l'échelle de Compétence sociolinguistique du Tableau 4.4, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.	Niveau
	Justification (y compris références documentaires)
Compétence pragmatique	Brève description et/ou référence
Quelles compétences pragmatiques attend-on que les candidats soient capables de mettre en œuvre : compétences discursives, fonctionnelles ? <ul style="list-style-type: none"> ➤ Les listes du CECR 5.2.3 peuvent servir de référence 	
Après avoir pris connaissance de l'échelle pour la Compétence pragmatique du Tableau 4.5, dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.	Niveau
	Justification (y compris références documentaires)
Compétence stratégique	Brève description et/ou référence
Quelles compétences stratégiques attend-on que les candidats soient capables d'utiliser ? <ul style="list-style-type: none"> ➤ Le débat du CECR 4.4.2.4 et 4.4.1.3 peut servir de référence 	
Après avoir pris connaissance de l'échelle de Compétence stratégiques des Tableaux 4 et 6 dites et justifiez à quel(s) niveau(x) l'examen devrait se situer.	Niveau
	Justification (y compris références documentaires)

Fiche A22 : Aspects de la Compétence langagière en Médiation

4.3. Conclusion : présentation graphique de la relation de l'Examen avec les niveaux du CECR

Une fois l'examen analysé en fonction des catégories du CECR, la relation doit être présentée sous forme de graphique. Cette représentation permet de visualiser le profil du contenu de l'examen étudié, selon les échelles appropriées du CECR pour les Activités de communication langagière et pour les Aspects de la compétence linguistique.

Dans le graphique ci-dessous, l'axe Y (vertical, à gauche) représente les niveaux du CECR. Sur l'axe X on représentera la Compétence langagière et les activités de communication langagière ainsi que les Aspects de la compétence en langue. Chaque colonne aura comme intitulé les catégories pertinentes du CECR. Les cases qui correspondent à l'examen ou aux épreuves traités seront ombrées. Si l'examen exige un niveau plus élevé dans certaines catégories, on le montrera en ombrant les cases correspondantes comme dans l'exemple de la Fiche A23 ci-dessous.

L'intitulé des colonnes de la Fiche A23 ne sera pas forcément le même que celui des épreuves de l'examen. Certaines colonnes peuvent correspondre aux épreuves mais on peut également en ajouter d'autres. Il se peut, par exemple, que l'examen ne propose pas d'épreuve distincte pour la compétence linguistique mais que le producteur de l'examen veuille indiquer aux utilisateurs le niveau de compétence linguistique attendu.

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

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

Chapitre 5

Standardisation des évaluations

Ce chapitre présente une suite de démarches destinées à aider ceux qui sont impliqués dans la mise en relation des examens et du CECR à parvenir à une compréhension commune des niveaux du CECR, à vérifier que l'on a atteint cette compréhension commune et à en garantir la pérennité.

La standardisation de la mise en relation avec les niveaux du CECR implique :

- l'atteinte d'une compréhension convenable des niveaux du CECR et de la capacité de relier les épreuves des examens locaux et des performances à ces niveaux
- l'assurance que cette compréhension est partagée par tous les participants et qu'elle est mise en œuvre de façon cohérente
- la mise en place de standards, c'est-à-dire la définition des seuils fonctionnels entre les différents niveaux de compétence, seuils fondés sur les décisions que l'on peut prendre quant aux performances de l'apprenant et/ou des items ou activités permettant de classer ou de sélectionner

Par rapport au premier point, un certain nombre de démarches de standardisation ont déjà été mises en œuvre dans différents contextes européens afin de relier les niveaux des examens existants avec ceux du CECR. La plupart de ces démarches sont des adaptations de celles élaborées et utilisées pour la conception et la validation des niveaux du CECR (voir l'Annexe B du CECR) ainsi que pour la validation des énoncés d'auto-évaluation de DIALANG (voir l'Annexe C du CECR). Les grandes lignes de ce chapitre s'appuient sur l'expérience décrite dans les rapports d'applications de ces différentes approches et procédures. Certains d'entre eux feront l'objet d'un compte rendu en annexe sous forme d'Etudes de cas.

De façon à garantir une mise en œuvre commune et cohérente des niveaux du CECR par des individus et des organismes différents, des techniques de formation d'examineurs, de rédacteurs d'items et de mise en réseau ont été développées. Ici aussi, les grandes lignes de ce chapitre s'inspirent de l'expérience apportée par ces différentes approches et procédures dans un certain nombre de contextes.

Les procédures de définition de standards de ce chapitre sont dues à la mise en commun des opinions d'experts au sujet de l'estimation de la pertinence et de la difficulté des tâches et des items aux différents niveaux du CECR, ainsi qu'à l'établissement de seuils fonctionnels permettant de reporter les notes obtenues sur l'échelle du CECR. Les procédures de vérification et d'ajustement éventuel de ces seuils fonctionnels sont présentées dans le chapitre 6.

5.1. Pré-requis à la standardisation

Les projets¹⁷ déjà en cours pour relier des examens avec les niveaux du CECR ont montré qu'avant toute tentative de standardisation, il est nécessaire de s'assurer que l'examen que l'on va relier au CECR répond à toutes les exigences de validité et de fiabilité. Cela suppose que l'on ait démontré que la Spécification était convenable en répondant aux exigences des étapes énumérées dans le Chapitre 4 de ce Manuel (Spécification). Cela présuppose également que l'examen à l'étude a été élaboré et

¹⁷ Kaftandjieva, F. et Takala, S. (2003) Development and Validation of Scales of Language Proficiency. In Vagle, W. (sous la direction de) Vurdering av sprakferfighet (Evaluation of Language Proficiency) Institutt for spak- og kommunikationsstudier, NTNU, Trondheim, 31-38, fondé sur une communication au Séminaire ALTE de Barcelone, 2001, en relation avec l'Année européenne des langues.

Alderson (2002) Relating national examinations to the Common European Framework: The Hungarian Experience.

Van Hest, De Jong et Stocks (2001) Relating the CEF with school leaving exams in the Netherlands.

Jones (2002) Relating the ALTE Framework to the Common European Framework of Reference

validé dans le respect des propriétés attendues d'un test valide, telles qu'elles sont énumérées dans la littérature spécialisée du domaine¹⁸.

S'appuyant sur les conditions générales requises d'une évaluation fiable et valide, Alderson (2002) signale que :

- 1) Si le résultat d'un examen n'est pas fiable, on ne peut pas comparer cet examen au CECR.
- 2) Si, chaque fois que l'on produit une nouvelle version de l'examen, il varie dans son contenu et son niveau de difficulté, il devient très difficile de comparer cet examen au CECR puisqu'il ne présente pas un standard stable.
- 3) Si les résultats de l'examen varient d'un enseignant à l'autre, d'un établissement scolaire à l'autre, on ne peut pas les relier au CECR.

Il est donc important que, parallèlement à la conduite de formations et aux procédures de mise en place de standards permettant de relier les examens au CECR, on prenne des mesures pour valider et – le cas échéant – améliorer des examens déjà existant. Cette tâche contribuera à satisfaire les exigences du Chapitre 6 : Validation empirique.

5.2. Organisation de la standardisation

Avant d'entrer dans le processus proprement dit, le coordinateur doit lire attentivement ce Manuel et prendre connaissance des ouvrages de référence pertinents dans ce contexte.

Afin de faciliter la visualisation du travail de formation à la standardisation, on trouvera un Tableau récapitulatif (Tableau 5.11) à la fin de ce chapitre. Les institutions peuvent utiliser le Tableau 5.11 afin de faire une estimation du montant du budget à prévoir pour le processus de standardisation et de décider à quelle hauteur elles veulent s'engager dans ce processus. Le tableau peut également servir aux coordinateurs qui l'utiliseront comme un aide-mémoire fonctionnel pour contrôler le processus et l'intégrer ensuite, dûment rempli, dans le rapport détaillé.

L'ordre dans lequel sont présentées les étapes du processus de standardisation n'est pas aléatoire. On le recommande comme étant le plus efficace mais on peut, bien évidemment, le modifier selon les besoins de la situation.

5.2.1. Planification préalable

Les coordinateurs sont responsables de l'organisation de la standardisation. Cela signifie qu'ils doivent répondre :

- de la logique à suivre, basée sur ce Manuel et sur les références appropriées
- de la structure du projet
- des décisions quant aux types d'expertises auxquels il faut faire appel ; des personnes à impliquer, de leurs rôles et de l'étape du processus où elles interviendront ;
- de la mobilisation d'experts locaux habitués à :
 - travailler avec le CECR ;
 - produire des programmes et des spécifications d'examens ;
 - évaluer des compétences de production en fonction de critères définis ;
 - travailler dans le domaine de l'évaluation en langue, notamment à la rédaction d'items et à la définition de standards ;
 - former des groupes d'enseignants et d'examineurs ;
- des instructions qu'ils donneront aux professionnels qui devront recueillir, dans un format standard :

¹⁸ Alderson et al. 1995, Bachman 1990 ; Bachman et Palmer 1996.

- Les échantillons d'écrits et les vidéos de performances orales qui seront utilisées pour calibrer les performances locales sur des échantillons standards du CECR et sur le CECR lui-même (Etape 3)
 - Les épreuves d'examen ou les items de tests qui serviront de documents de travail dans les stages sur l'évaluation
- de la préparation, de l'élaboration et de la reproduction du matériel qui sera utilisé aux différentes étapes de la démarche :
 - Les descripteurs de niveaux du CECR
 - Les tableaux du CECR et les outils d'évaluation
 - Une sélection (dans la batterie fournie) d'échantillons de performances standards du CECR et – si possible – de modèles d'épreuves et d'items calibrés
 - Une sélection d'échantillons de performances locales et/ou d'épreuves d'examen
 - Les fiches de compte rendu et les documents utilisés pour recueillir de l'information au cours des formations
 - Du recueil et de l'analyse des données issues des stages de standardisation, de la présentation et de la reproduction de résultats significatifs (par exemple, la difficulté empirique de la valeur des épreuves ; la notation d'autres groupes sur des échantillons) afin de les réutiliser éventuellement dans d'autres stages au moment approprié.
 - De l'organisation même des stages de façon la plus adaptée à la situation locale. Les coordinateurs devront 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/concepteurs d'épreuves), les sessions auxquelles ils participeront, et les implications sur la préparation des stages selon le public concerné ;
 - la planification du temps, l'ambiance et le regroupement le plus adéquat 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 et le résumé des conclusions.

Si l'on envisage d'y consacrer toute une journée, on peut organiser deux stages, un le matin et l'autre l'après-midi. Selon le nombre et la formation antérieure des participants, les activités prendront plus ou moins de temps. Voir les propositions de durée à la fin de chaque sous partie.

 - La vérification du nombre suffisant de locaux pour les travaux de groupe et de la disponibilité du matériel.
 - L'organisation de la documentation et du compte rendu du travail effectué durant les stages afin d'assurer la crédibilité du système et de fournir un support à des stages et à des sessions ultérieures.
 - De la planification du contrôle continu, de la diffusion et des actions de suivi.

5.2.2. Animation des stages

Les différentes étapes du processus de standardisation doivent avoir lieu au cours de stages où, en suivant les directives et les documents donnés par le coordinateur, les participants se familiarisent avec le CECR, analysent et évaluent des performances ou des épreuves et parviennent à un consensus afin de leur assigner un niveau du CECR.

Les responsables de la coordination des démarches de formation le sont aussi de la conception et de la planification du temps ainsi que de l'organisation selon les options proposées. Ils ont également la responsabilité de préparer le matériel nécessaire pour les stages. Pendant les stages, les coordinateurs doivent :

- Vérifier que les participants arrivent à une bonne compréhension de base de ce qu'est le CECR et jusqu'à quel point ils prennent conscience de la façon dont le CECR peut contribuer à améliorer leur travail et leur vie professionnelle. 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 :
 - Amorçage et exemplification
 - Évaluation individuelle
 - Évaluation en tandem
 - Discussion en grand groupe
- Recueillir des informations et donner régulièrement un retour d'information de manière aussi claire et visuelle que possible,
- Vérifier que, comme cela est précisé dans les instructions, 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 :

- d'assurer la diffusion d'un dossier regroupant le matériel utile (procès verbaux, échantillons standards de performances, modèles d'épreuves ou d'items) ;
- de faire ou de faire faire les analyses nécessaires ;
- de faciliter et de contrôler le suivi du réseau afin d'encourager la coopération et le perfectionnement professionnels ;
- de contrôler la bonne marche du système.

5.2.3. Enregistrements et rapports

Les coordinateurs devront recueillir les informations propres au processus et les accompagner de pièces justificatives. Une fois les stages terminés, ils en rédigeront un bref compte rendu et compléteront la Fiche « Rapport de standardisation » (Fiche B1) pour chacun d'entre eux.

Des copies de ces documents devront être inclus dans le rapport rédigé à l'intention des professionnels.

5.2.4. Diffusion

Nombreux sont ceux qui peuvent prétendre qu'ils disposent déjà d'un système d'élaboration de tests qui suit les procédures standards et en garantit la validité et la fiabilité. Toutefois, même dans de tels cas, le fait d'avoir un test stable et d'avoir réalisé les démarches de mise en relation décrites dans ce chapitre, n'est pas en soi une garantie que le test gardera toujours sa relation avec le CECR. Il faut donc mettre en place un système de contrôle pour s'assurer que les examens qui ont été reliés au CECR à un moment donné continuent à l'être en dépit des années qui passent et du personnel qui change.

Il existe un certain nombre de moyens pour que les organismes puissent garantir la qualité, la fiabilité et la validité des liens de leur examen au CECR. On recommande de combiner les démarches suivantes :

Le suivi

Le premier pas à faire pour consolider la standardisation est de s'assurer que le travail et l'expérience sont pris en compte et utilisés. Il y a différentes façons de procéder, qui doivent inclure :

- La documentation : les stagiaires formés doivent avoir accès aux matériels utilisés pour la formation afin de pouvoir s’y référer dans leur travail quotidien :
 - Echantillons standards de performances
 - Outils d’évaluation (Tableaux 5.4, 5.5, 5.8)
 - Modèles d’épreuves ou d’items
 - Matériel de formation
 - Descriptions des démarches
 - Exemples de procédures correctes

- Les réunions : les anciens participants au processus de standardisation devraient exploiter ce matériel lors de réunions régulières de groupes d’examineurs, d’enseignants et de rédacteurs d’items, dans leur établissement et entre institutions afin de poursuivre le débat sur les activités et les performances et accroître ainsi l’implication et le professionnalisme ainsi que les possibilités de comparaison.

- La formation continue : les nouveaux rédacteurs d’items et les examinateurs qui rejoignent les institutions après que le processus de standardisation a eu lieu devraient être formés à l’aide du CECR et des échantillons standards de performances et de modèles d’items, en utilisant les techniques décrites dans ce chapitre.

Maillage du réseau

Les organismes doivent encourager l’organisation de contacts entre des groupes locaux et entre des groupes du secteur éducatif régional s’occupant de la langue concernée ainsi qu’entre des groupes éducatifs régionaux s’occupant d’autres langues, afin de s’assurer que l’interprétation des niveaux du CECR est réellement partagée. La mise en réseau de groupes du même secteur linguistique/éducatif de pays voisins facilite la mise en relation des systèmes et est également souhaitable.

Visites de contrôle

Les coordinateurs, comme les formateurs d’enseignants et les inspecteurs scolaires qui ont participé au processus de standardisation programmeront des visites d’écoles afin d’accompagner le fonctionnement du système local, d’organiser des stages de formation et de recyclage, de contrôler le déroulement de la passation des examens, la correction, la notation, etc. Ces visites stimuleront également l’échange d’informations entre les établissements et les individus.

Fiche B1 : Standardisation

Standardisation : Fiche de synthèse		
Lieu	Date	
Coordinateur	Nom	Institution/Projet
Etape	Familiarisation Formation Calibrage des performances Etablissement de standards	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Domaine(s)	Evaluation d’échantillons de production orale Evaluation d’échantillons de production écrite Epreuve/item : <ul style="list-style-type: none"> • Compréhension de l’oral • Compréhension de l’écrit • Travail en langue Autre :	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

Participants	Nombre	Fonctions
Activités réalisées	Familiarisation Illustration avec des exemplaires standards Pratique libre/contrôlée avec des exemplaires calibrés Calibrage d'échantillons de performances locales Réunion de synthèse des évaluateurs Formation avec des items calibrés Evaluation de la difficulté des items Retour d'information sur la difficulté réelle des items Autre	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Matériel utilisé	Echantillons standards de performances du CECR Outils d'évaluation du CECR (Tableaux 5.4, 5.5, 5.8) Echantillons de performances locales Outils d'évaluation adaptés (à annexer) Epreuves et items de tests calibrés sur le CECR Epreuves et items de tests locaux Autre : _____	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Information sur les épreuves et les items		
Autres commentaires		
Démarches de diffusion envisagées		

Fiche B1 : Standardisation: Fiche de synthèse

5.3. Les étapes du processus de standardisation

Concevoir, élaborer et diffuser des procédures de **standardisation** exige **l'implication, à différents niveaux**, de professionnels expérimentés exerçant des fonctions diverses. Bien que des contextes différents puissent entraîner des approches différentes, les étapes décrites dans cette partie doivent être **suivies pas à pas** :

1 Familiarisation → 2 Formation → 3 Calibrage de performances → 4 Définition des standards
--

Figure 5.1 Le processus de standardisation

- **Première étape : la Familiarisation** devra comprendre les activités exposées en détail dans le Chapitre 3 : examen des divers aspects du CECR, tri des échelles de descripteurs, utilisation d'outils d'auto-évaluation, etc. Ces activités sont conçues pour s'assurer que les participants à un processus de standardisation ont une compréhension approfondie du CECR, de sa structure, des catégories de description et des niveaux communs de référence. Avant de commencer la formation à la standardisation, les participants doivent être totalement familiarisés avec les niveaux du CECR.
- **Deuxième étape : la Formation** doit se faire en différentes étapes selon les savoir-faire langagiers traités. On peut distinguer deux grandes sous-étapes : la première pour les compétences de production et la seconde pour les compétences de réception et les compétences de base.

Si l'on ne possède pas d'échantillons standards de performances et/ou d'épreuves ou d'items pour la langue dont on s'occupe, on utilisera des échantillons d'une autre langue (par exemple l'anglais) pour cette étape. Pour les langues moins enseignées, cela signifie que les participants au processus de standardisation devront partager un bon niveau de compétence (au minimum C1) dans une des langues pour laquelle des copies standards sont disponibles.

- **Performances** : la formation des experts en évaluation des compétences de production suppose que l'on utilise des échantillons de performances orales et écrites associées à ce Manuel et déjà standardisées selon les niveaux du CECR, de façon à évaluer les performances en fonction des niveaux. Au cours de cette étape, les participants évaluent les échantillons, débattent sur les niveaux auxquels ils se situent et acquièrent l'expérience de mise en relation de descripteurs écrits avec des niveaux donnés de performance. Au cours de cette étape, il est essentiel d'atteindre un niveau élevé de consensus, les échantillons étant déjà standardisés par rapport aux outils d'évaluation constitués de descripteurs du CECR. Il est vital que l'interprétation des niveaux du CECR à laquelle on arrive dans cette situation soit en adéquation avec les interprétations données dans d'autres situations et que des groupes différents ne développent pas leur propre interprétation exclusive.
- **Tâches et items** : la formation des experts en évaluation des compétences de réception suppose que l'on évalue la difficulté des épreuves ou des items en fonction des niveaux du CECR en utilisant des items calibrés associés à ce Manuel et déjà standardisés selon les niveaux du CECR. Au cours de cette étape, les participants évaluent des épreuves ou des items, débattent pour savoir pourquoi certains se situent à un niveau plutôt qu'à un autre et acquièrent l'expérience de la mise en relation des descripteurs écrits avec des réalisations particulières dans des textes et des épreuves ou items. A cette étape également, il est essentiel d'atteindre un niveau élevé de consensus parce que ces items calibrés sont déjà standardisés par rapport aux niveaux du CECR.
- **Troisième étape : le Calibrage des performances** se fait en appliquant le consensus auquel on est arrivé dans l'étape de Formation à l'évaluation des échantillons de performances locales préalablement recueillis. Les « échantillons locaux » que l'on utilisera peuvent être sélectionnés pour :
 - une langue cible pour laquelle on n'a pas encore d'échantillons de performances standards ;
 - un secteur éducatif qui n'est pas représenté convenablement par des échantillons standards, par exemple une langue de spécialité ;
 - le système éducatif local qui couvrirait, par exemple, le premier et le second cycles du secondaire ainsi que l'enseignement aux adultes ;
 - les tâches utilisées dans les examens locaux.

Ce processus entraîne des démarches semblables à celles de la Deuxième étape et avec des activités semblables afin d'évaluer le niveau des échantillons locaux en fonction du CECR. Le résultat, à condition d'atteindre un consensus satisfaisant, sera un ensemble d'échantillons locaux de performances orales et écrites standards reliées aux niveaux du CECR.

- **Quatrième étape : la Définition des standards** est le processus qui consiste à définir les « seuils » des différents niveaux d'un test. Dans le contexte de ce Manuel, cela signifie que l'on définira les notes seuils ou les scores correspondant aux niveaux du CECR déclarés dans l'examen ou le test. Le point de départ de ce processus est constitué par l'évaluation de la difficulté des épreuves et des items isolés. Ces jugements initiaux sur la difficulté doivent alors être confirmés par la preuve empirique de la difficulté réelle, dans la pratique, des épreuves et des items. Enfin, définir des standards pour les examens ou les tests en fonction du CECR implique également que l'on examine la validité interne et externe, comme cela est exposé dans le Chapitre 6.

5.4. Première étape : Familiarisation

Il est absolument essentiel de commencer tout processus de standardisation en faisant une sélection parmi les exercices de familiarisation parmi ceux décrits au Chapitre 3.

L'expérience semble indiquer qu'il est indispensable de faire suivre une activité introductive (choisie parmi les activités *a* à *c*) par au moins deux autres activités qui supposent une démarche plus détaillée (choisie parmi les activités *d* à *f*).

5.5. Deuxième étape : Formation

L'objectif des activités exposées dans cette partie est de garantir que les praticiens impliqués dans ce processus de formation peuvent relier leur interprétation des niveaux du CECR avec des échantillons standards de performance et des épreuves ou des items calibrés. C'est la condition pour qu'ils puissent ensuite évaluer, par rapport aux niveaux du CECR, des échantillons pertinents de performances locales et des épreuves ou des items produits localement (Troisième étape). Ce travail entraînera la mise en pratique de leur compréhension « abstraite » des niveaux du CECR par une application sur des critères définis tirés du CECR.

On recommande aux utilisateurs de travailler avec des échantillons standards de performances du CECR avant de passer à l'examen d'épreuves et d'items produits localement, même s'ils ne disposent pas d'échantillons de performances locales et même si l'examen n'évalue pas les compétences de production. L'observation de performances de niveaux différents aide aussi les rédacteurs d'items à avoir une idée plus claire de ces niveaux.

Durée

Le temps nécessaire pour mener à bien la Deuxième étape dépendra :

- de la connaissance que les participants ont de la question pour avoir participé à des sessions *antérieures* de formation d'évaluateurs ;
- de leur habitude à utiliser des échelles d'évaluation ;
- de leur expérience de la rédaction d'items et de l'évaluation de la difficulté des épreuves ou des items.

Avec des stagiaires expérimentés, on pourra éventuellement atteindre les objectifs de la première partie de la Deuxième étape (Compétences de production orale et écrite) en une journée, en consacrant la matinée à la production orale et l'après-midi à la production écrite. On peut alors passer à la Troisième étape, le calibrage des performances, en travaillant le jour suivant avec des échantillons de performances locales. Alternativement, on peut consacrer exclusivement le premier jour à la formation et à la standardisation des activités sur des performances orales et le second jour sur des performances écrites. Il faut chaque jour commencer le travail avec des échantillons standards de performances

(Deuxième étape) le matin et passer au calibrage des échantillons locaux (Troisième étape) l'après-midi.

Le temps nécessaire à la seconde partie de la Deuxième étape (Compétences de réception) dépendra non seulement de la plus ou moins grande habitude que les participants ont du processus de sélection et de rédaction d'épreuves ou d'items mais aussi du nombre de compétences à évaluer.

On peut suivre un déroulement identique pour chaque compétence. Si – comme on le recommande dans ce Manuel – on travaille d'abord la Compréhension de l'écrit, la standardisation d'items calibrés peut avoir lieu le matin, suivie, l'après-midi, de l'évaluation des items de tests locaux (Quatrième étape : Définition de standards). Afin de calculer au mieux la durée de chaque étape et de planifier le processus dans son ensemble, le coordinateur pourra préparer un tableau semblable au Tableau 5.1 :

Tableau 5.1 : Planification du processus de standardisation

Compétence	Deuxième étape
Production orale	Une session ¹⁹
Production écrite	Une session
Compréhension de l'écrit
Compréhension de l'oral
Utilisation de la langue. Grammaire ²⁰
Utilisation de la langue. Vocabulaire

Echantillons nécessaires

Le processus commence par l'analyse et l'évaluation d'échantillons standards d'Interaction orale et de Production (vidéos : à produire²¹) du CECR, et se poursuit (le cas échéant) par des échantillons standards d'Interaction écrite et de Production (textes à produire²²). Les échantillons standards de production orale (vidéos) suivent une démarche classique qui comprend, de la part du candidat, une phase de Production (un monologue suivi dans lequel le candidat explique quelque chose à un autre candidat qui pose des questions) et une phase d'Interaction (dans laquelle les deux candidats débattent spontanément d'un problème)²³.

Pour l'évaluation de la Production écrite, il est important de prendre connaissance d'échantillons d'Interaction écrite (notes, lettres) ainsi que de Production écrite (descriptions, narrations, comptes rendus) d'un candidat. Ceci est particulièrement important aux niveaux inférieurs.

Il peut arriver que l'on ne dispose pas d'échantillons standards de performances et/ou d'épreuves calibrées pour la langue dont on s'occupe. Dans ce cas, la conclusion escomptée sera le résultat du travail effectué sur des échantillons disponibles en anglais, utilisés comme catalyseurs (à condition que les participants possèdent un bon niveau de compétence dans cette langue, au minimum un niveau C1). Si tel est le cas, il faudra l'annoncer comme une formation indirecte dans la documentation d'accompagnement.

¹⁹ Une session est une période de 4-5 heures, y compris une pause de 20-30 minutes.

²⁰ La durée de Utilisation de la langue : vocabulaire et Utilisation de la langue : grammaire est évaluée sur la base de l'utilisation des échelles de DIALANG déjà disponibles pour ces composantes sur l'utilisation de la langue.

²¹ Echantillons standards du CECR pour la production orale et l'interaction. En cours de production : une vidéo pour l'anglais, début 2004. Les vidéos pour d'autres langues pourraient faire l'objet d'un projet Socrates

²² Echantillons standards du CECR de production écrite et l'interaction. Il n'y a pas pour l'instant de projet concret de recueil d'échantillons écrits.

²³ Ce déroulement standard a été adopté par le projet de recherche suisse qui a élaboré l'échelle de descripteurs du CECR. Il limite les interventions/effets de l'interlocuteur. La cassette vidéo qui accompagne cette version expérimentale du Manuel se compose d'enregistrements d'apprenants suisses dont les performances ont été mathématiquement calibrées sur les niveaux du CECR.

Le récapitulatif d'échantillons nécessaires (Tableau 5.2) pourra aider les coordinateurs.

Arriver au consensus et le vérifier

Tout au long du stage, les animateurs stimulent commentaires et discussions et résumés des évaluations de la façon la mieux appropriée au contexte afin de parvenir à un consensus véritable.

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 du coordinateur, mais n'est pas donné aux stagiaires avant leur évaluation. Il ne sera donné qu'à la fin du stage, par le coordinateur. Contrairement à la phase suivante (calibrage d'échantillons locaux), 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 une réponse correcte en appliquant les critères.

Tableau 5.2. Echantillons de performances pour la standardisation	
Compétence	Nombre d'échantillons
Production orale	Echantillons d'Interaction orale et de Production : 9 au minimum, 12 au maximum
Production écrite	Echantillons d'Interaction écrite et de Production : comme pour l'Oral
Compréhension de l'écrit	Un minimum de 15 items (certains inclus dans un mini test)
Compréhension de l'oral	Idem ?
Utilisation de la langue. Grammaire ²⁴	Idem ?
Utilisation de la langue. Vocabulaire	Idem ?....

L'animateur doit savoir (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 d'accorder le temps nécessaire à la formation (Deuxième étape) avant de passer au travail sur les échantillons locaux non encore standardisés.

Il y a deux écoles sur la façon de conduire le groupe vers le consensus qui convient.

La première est une approche humaine où l'on évite d'embarrasser les participants en respectant l'anonymat des évaluations. Cette approche garantit que les participants rapportent leur évaluation réelle, ne sont pas « malmenés » et que le consensus qui émerge est un consensus authentique. Avec cette approche, les individus sont influencés par les évaluations des autres : si un participant est « marginal », il s'en rend compte et se recentre.

- C'est ce qui se passe avec le vote électronique (comme dans certains programmes de télévision). Comme il est interactif, les « marginaux » sont amenés à reconsidérer leur opinion de sorte que l'écart diminue sans gêne.
- La discrétion est également préservée si le coordinateur distribue des bulletins d'évaluation sans faire de commentaires. Si l'on veut identifier les évaluateurs dans un recueil de données pour des analyses ultérieures on peut utiliser des surnoms (Astérix par exemple) ou des numéros de code imprimés sur les bulletins. La projection rapide des bulletins anonymes sur un rétro-projecteur ou dans un tableau de synthèse expose les « marginaux » sans les gêner – à moins qu'ils ne décident d'argumenter !

²⁴ La durée de Utilisation de la langue : vocabulaire et Utilisation de la langue : grammaire est évaluée sur la base de l'utilisation des échelles de DIALANG déjà disponibles pour ces composantes sur l'utilisation de la langue.

La deuxième école préconise une approche plus vigoureuse : les opinions divergentes doivent s'exprimer et faire l'objet d'un débat si l'on veut parvenir à un vrai consensus. Le consensus sera ici plus délibéré, résultat d'une discussion argumentée - et peut être obtenu par un stagiaire convaincant. C'est la raison pour laquelle il est bon que l'animateur s'assure que les participants connaissent les échantillons standards ainsi que la raison pour laquelle on a attribué tel niveau à tel échantillon ainsi que son rapport avec les descripteurs.

- L'annonce conjointe, à haute voix, par chacun, de son évaluation est une technique efficace (et amusante) qui permet aussi aux participants les plus timides ou les moins sûrs d'eux de sauver la face.
- Les stagiaires apprécient le travail en tandem ou en petits groupes. L'animateur peut circuler et écouter les discussions, ramener éventuellement un groupe dans la bonne direction et demander qu'un compte rendu soit fait par un rapporteur de chaque groupe. L'avantage principal du travail en tandem ou en petits groupes est d'obliger de façon naturelle les participants à utiliser les critères définis pour justifier leurs jugements.
- La façon la plus simple d'enregistrer les résultats du groupe est de les recueillir au fur et à mesure et de les présenter, sur une grille, au rétroprojecteur.

Quel que soit le type de fonctionnement choisi, le coordinateur devra calculer le pourcentage de stagiaires qui s'accordent sur les différentes évaluations ou les coefficients de corrélation entre évaluateurs. Le coordinateur devra décider de l'opportunité de communiquer ces chiffres aux participants pour contribuer à la formation et à une meilleure convergence des évaluations.

Il est également intéressant de présenter un schéma de dispersion des résultats. Une façon simple de le faire est de produire un graphique (en tuyaux d'orgue) sur Microsoft Excel. Une autre consiste à utiliser les boîtes à moustaches produites par SPSS (logiciel de statistiques). Pour un exemple de ce type de graphique, voir la Figure 6.5 du Chapitre 6.

5.5.1. Formation avec des échantillons standards de Production

Interaction orale et Production

Il est essentiel pour ce stage, que les participants utilisent la grille d'évaluation de performance orale (CECR Tableau 3²⁵ reproduit ici comme Tableau 5.5), une échelle globale simplifiée d'après ce tableau (Tableau 5.4) et une fiche standard d'évaluation pour noter leurs commentaires et le niveau attribué à chaque performance (exemples donnés dans les Fiches B2 et B3).

Ce stage est organisé en trois phases :

- **Phase 1 : Illustration.** Le coordinateur commence le stage par deux ou trois performances orales standards pour exemplifier les niveaux. Le premier de ces exemples doit présenter une performance de profil relativement « plat » parmi les catégories du Tableau 3 du CECR/ Tableau 5.5 du Manuel (à savoir, un locuteur qui serait par exemple de niveau B1 dans toutes les catégories Étendue, Correction, Aisance, Interaction, Cohérence).

Le coordinateur projette la cassette vidéo et invite les stagiaires à commenter la performance avec leurs voisins. A un moment donné, le coordinateur reconstitue le grand groupe et lui fait expliciter pourquoi cette performance illustre le niveau décrit sur la grille du Tableau 3 du CECR (Tableau 5.5 du Manuel) et non un niveau inférieur ou supérieur.

- On recommande que l'un de ces échantillons standards introductifs montre un profil moins régulier, par exemple si le locuteur est au niveau B1 dans certaines catégories mais en B2 ou au moins en B1+ dans d'autres. En effet, si la question des « profils inégaux » n'est pas traitée assez tôt au cours de la formation, elle peut poser ultérieurement un problème.

²⁵ Tableau 3 – Niveaux communs de compétence : Aspects qualitatifs de l'utilisation de la langue parlée. p. 28.

- Afin de mettre en évidence le fait que certains candidats peuvent avoir des profils très inégaux et qu'il faut examiner séparément les différentes catégories (Etendue, Correction, Aisance, Interaction, Cohérence), les coordinateurs peuvent envisager d'évaluer plusieurs performances dans la même catégorie. Cela neutralise la tendance naturelle des évaluateurs à laisser leur impression générale avoir une influence sur leur jugement dans chaque catégorie (« effet de halo »).
 - On peut demander aux participants d'utiliser d'abord l'échelle globale (Tableau 5.4) qui simplifie la grille du Tableau 3 du CECR (Tableau 5.5) afin de se rendre parfaitement compte de leur impression globale sur le niveau des candidats avant d'examiner les catégories de la grille du Tableau 3 du CECR (Manuel Tableau 5.5). S'étant fait une première impression sur le niveau de la performance, ils devraient alors consulter les descripteurs détaillés de ce Niveau sur la grille du Tableau 3 du CECR (Manuel Tableau 5.5), lire les descripteurs pour les niveaux immédiatement supérieur et inférieur de chaque catégorie et utiliser la grille pour tracer le profil de la performance du candidat.
 - Ils devraient alors utiliser les descripteurs de cette grille comme fil directeur de la discussion avec leur voisin.
 - Le Tableau 3 du CECR (Tableau 5.5) présente les critères : Etendue, Correction, Aisance, Interaction, Cohérence.
 - Dans un contexte d'évaluateurs monolingues, il peut être souhaitable de remplacer une des deux dernières colonnes par l'échelle de Maîtrise du système phonologique du CECR 5.2.1.4.
 - Avec des échantillons vidéos de langues maternelles variées et des évaluateurs uniquement habitués au discours de l'apprenant par des locuteurs de leur langue maternelle, les aspects phonologiques peuvent provoquer la confusion et il peut être préférable de ne pas essayer de les évaluer.
- **Phase 2 : Pratique.** Dans cette seconde phase –où l'on utilisera de nouveau trois échantillons – le rôle du coordinateur 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 bulletins, le coordinateur utilisera une fiche de synthèse sur transparent (par exemple la Fiche B3) pour enregistrer les évaluations. Tout au long de cette phase, le coordinateur doit faire visualiser aux participants leur comportement en tant que groupe et animer la discussion comme indiqué plus haut, sans embarrasser les individus.
 - Les échantillons standards du CECR montrent des groupes de deux apprenants réalisant d'abord une tâche de Production suivie d'une Interaction. Il est possible de connaître le niveau des deux apprenants de la cassette mais il n'est pas judicieux de demander aux évaluateurs en formation d'évaluer les deux locuteurs. L'expérience montre que des évaluateurs inexpérimentés surestiment le niveau du meilleur locuteur et sous estiment l'autre. En conséquence, il est préférable de se concentrer sur un seul des deux locuteurs, en montrant sa phase de Production puis la phase d'Interaction entre les deux locuteurs. A la fin, on peut organiser une discussion ouverte sur le niveau du second apprenant.
 - Les coordinateurs doivent décider à l'avance s'ils continueront à utiliser l'échelle globale (Tableau 5.4) après la phase d'Illustration. Elle est utile en ce sens (a) qu'elle donne aux participants un point de départ pour la lecture de la grille (Tableau 3 du CECR ; Tableau 5.5 du Manuel) et, (b) qu'elle aide les participants à faire la part de leur première impression par rapport à un jugement réfléchi – en particulier si l'on a consigné séparément les deux réactions dans la fiche d'enregistrement proposée (Fiche B2).
 - Si l'on n'a pas utilisé le vote anonyme, une technique efficace consiste en l'écoute des 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 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 . La grande majorité des participants devrait en effet être d'accord sur le niveau avec une dispersion inférieure à un niveau et demi. Par exemple, pour une performance que l'on s'accorde à situer en B1+, la dispersion des résultats ne devrait pas excéder la gamme de B1 à B2 ; pour une performance que l'on s'accorde à placer en B1, la dispersion devrait être de A2+ à B1+.

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

- Au cours des discussions, et afin de contribuer à une meilleure compréhension du niveau, le coordinateur décidera s'il est pertinent d'utiliser d'autres échelles de production orale du CECR et de justifier plus en détail l'affectation de niveau.
- On recommande d'analyser au moins une performance par niveau du CECR au cours du stage, de l'évaluer et d'en discuter.
- Il est bon de disposer de plus d'une cassette d'échantillons vidéos du CECR et de les avoir préparées de sorte à démarrer tout de suite sur le visionnement des échantillons choisis pour la formation. On gagne du temps à ne pas embobiner/rembobiner. Si ce n'est pas possible, il est alors essentiel de noter au préalable le numéro inscrit au compteur pour le démarrage de chaque échantillon afin d'accéder très rapidement à la bonne séquence.

Tableau 5.3 : Planification du temps pour l'évaluation d'échantillons vidéos de performances

Recommandation : groupe de 30 participants au maximum	
Première étape : Familiarisation	60 minutes
Deuxième étape : Travail sur des échantillons standards:	60 minutes
<i>Phase 1 : Illustration avec environ trois performances standards</i>	60 minutes
<i>Pause</i>	
<i>Phase 2 : Pratique sous contrôle du coordinateur avec environ trois performances standards</i>	
<i>Phase 3 : Etape libre avec environ trois performances standards</i>	60 minutes
	60 minutes
Déjeuner	
Troisième étape : Calibrage des échantillons locaux :	
<i>Evaluation individuelle et discussion de groupe sur environ trois performances</i>	60 minutes
<i>Evaluation individuelle d'environ cinq performances supplémentaires</i>	60 minutes
<i>Pause</i>	
<i>Planification du suivi et mise en réseau</i>	60 minutes
<i>Synthèse et clôture</i>	30 minutes
Documents et matériel à préparer	
<i>Photocopies pour tous les participants :</i>	
<ul style="list-style-type: none"> • Grille d'évaluation du CECR Tableau 3/du Manuel Tableau 5.5 • Echelle d'évaluation simplifiée ci-dessus : Tableau 5.4 (si nécessaire) • Feuilles d'évaluation pour les participants : exemples des Fiches B2 et B3 • Choix d'échelles complémentaires pertinentes ou Tableau 4.4 	
<i>Auxquels s'ajoutent</i>	
<ul style="list-style-type: none"> • Des vidéos de performances standards • le Manuel • des fiches de synthèse pour le coordinateur et des transparents (Fiche B 4) • des vidéos locales (enregistrées et/ou sélectionnées selon les instructions des Etudes de cas) 	

Production écrite

Le deuxième stage sur des échantillons *standards* est consacré à l'analyse et à l'évaluation de performances écrites *standards*. On recommande ici une démarche semblable à celle effectuée pour les performances orales. La grille d'évaluation à laquelle se reporter est fournie par le Tableau 5.8. Cette grille combine les descriptifs généraux du Tableau 5.4 (Echelle d'évaluation générale de l'oral) et les critères détaillés d'Etendue, Cohérence et Correction. S'y ajoutent – à droite – deux colonnes sur la Description et sur l'Argumentation qui ne doivent être utilisées que pour les textes de ce type.

- **Phase 1 : Illustration** : Le stage commence par deux ou trois performances écrites standards que le coordinateur utilise pour exemplifier les niveaux. Le premier de ces exemples doit présenter une performance de profil relativement « plat » parmi les catégories du Tableau 5.8 du Manuel (à savoir, un locuteur qui serait par exemple de niveau B1 dans toutes les catégories Etendue, Correction, Aisance, Interaction, Cohérence et également d'un bon niveau pour décrire et argumenter). Comme pour les exemples de performance orale, le coordinateur peut envisager d'évaluer plusieurs performances écrites dans une même catégorie afin que les participants se rendent compte de « l'effet de halo ».

Le coordinateur invite les stagiaires, individuellement ou en tandems, à lire les textes et à examiner la performance par rapport aux critères du Tableau 5.8. A un moment donné, le coordinateur reconstitue le grand groupe et lui fait expliciter comment cette performance illustre le niveau décrit sur la grille du Tableau 5.5 et pourquoi elle n'est pas du niveau inférieur ou supérieur.

- On recommande que l'un de ces premiers échantillons standards montre un profil moins régulier, par exemple que le locuteur soit au niveau B1 dans certaines catégories mais en B2 ou au moins en B1+ dans d'autres. En effet, si la question des « profils inégaux » n'est pas traitée assez tôt au cours de la formation, elle peut poser ultérieurement un problème.
- **Phase 2 : Pratique** : Dans cette seconde phase – où l'on utilisera de nouveau trois échantillons – le rôle de l'animateur 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 bulletins, le coordinateur utilisera une fiche de synthèse (par exemple la Fiche B3) pour rapporter les évaluations sur transparent.

Tout au long de cette phase, le coordinateur doit 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 en l'écoute des 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 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. C'est-à-dire que la grande majorité des participants devraient être d'accord presque général sur le niveau avec une dispersion inférieure à un niveau et demi. Par exemple, pour une performance que l'on s'accorde à situer en B1+, la dispersion des résultats ne devrait pas excéder la gamme de B1 à B2 ; pour une performance que l'on s'accorde à placer en B1, la dispersion devrait être de A2+ à B1+.

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

- Comme au cours des discussions sur la production orale et les échantillons d'interaction, le coordinateur peut décider d'utiliser des échelles spécifiques²⁶ pour faciliter un accord et mieux justifier l'affectation à un niveau donné. Le coordinateur peut également distribuer le Tableau 4.4 du Chapitre 5 sur la Spécification.

On recommande d'analyser au moins une performance par niveau du CECR au cours du stage, de l'évaluer et d'en discuter.

TABLEAU 5.4 : ECHELLE DE PRODUCTION ORALE GENERALE

C2	<i>Transmet, avec naturel et précision des nuances de sens subtiles.</i> Est capable de s'exprimer spontanément et avec beaucoup d'aisance, de communiquer facilement avec habileté et de discriminer avec précision des nuances de sens subtiles. Peut produire des descriptions claires, régulières et bien structurées.
C1	<i>S'exprime spontanément et avec aisance dans un discours clair et bien structuré.</i> Est capable de s'exprimer spontanément et avec aisance, presque sans effort, dans un discours régulier. Peut faire des descriptions claires et détaillées de sujets complexes. Niveau élevé de correction ; les erreurs sont rares.
B2+	
B2	<i>Exprime ses opinions sans effort notable.</i> Est capable de communiquer sur une gamme étendue de sujets et de produire des énoncés sur un rythme assez régulier. Peut faire des descriptions claires et détaillées sur une vaste étendue de sujets relatifs à son centre d'intérêt. Ne commet pas de fautes qui provoquent des malentendus.
B1+	
B1	<i>Rapporte de façon compréhensible ce qu'il/elle tient à dire</i> Est capable de tenir un discours compréhensible même si les pauses pour rechercher des mots ou des phrases sont très évidentes ainsi que la remédiation. Peut relier des éléments discrets simples en un paragraphe articulé pour faire des descriptions simples sur des sujets familiers variés dans son domaine. Utilisation raisonnablement juste d'un répertoire de base associé aux situations les plus prévisibles.
A2+	
A2	<i>Rapporte des informations de base sur, par exemple, le travail, la famille, les loisirs, etc.</i> Est capable de communiquer dans un échange simple et direct d'informations sur des sujets courants. Peut se faire comprendre dans de très courts énoncés même si les pauses, les hésitations et la reformulation sont très évidentes. Peut décrire en termes simples ses conditions de vie, ses études, son métier présent ou passé. Utilise correctement des structures simples mais peut commettre des fautes de base systématiquement.
A1	<i>Fait des énoncés simples sur des détails personnels et des sujets très familiers.</i> Est capable de se faire comprendre de façon simple, de poser des questions sur des détails personnels et d'y répondre à condition que l'interlocuteur parle lentement et clairement et soit prêt à aider. Peut se débrouiller avec des énoncés très courts, isolés, le plus souvent stéréotypés. De nombreuses pauses pour chercher ses mots et prononcer les moins familiers.
Au-dessous de A1	N'est pas dans la norme A1
<ul style="list-style-type: none"> • <i>Utiliser cette échelle pour les deux ou trois premières minutes d'un échantillon de production orale afin de décider approximativement du niveau auquel on pense que le locuteur se trouve.</i> • <i>Puis passer au Tableau 5.5 (Tableau 3 du CECR) et évaluer plus en détail la performance par rapport aux descripteurs de ce niveau.</i> 	

²⁶ Production écrite générale, Ecriture créative, Essais et rapports.

**TABLEAU 5.5 : GRILLE DES CRITERES D'EVALUATION DE L'ORAL
(CECR Tableau 3)**

	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 liaisons 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+					
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+					
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".

Fiche B2 : Evaluation analytique

NOM DE L'APPRENANT

Niveaux : **B, W, W+, T, T+, V, V+, E, M**

1. Impression initiale
Classement - échelle globale

2. Analyse détaillée / Estimation à l'aide de la grille

ETENDUE	CORRECTION	AISANCE	INTERACTION	COHERENCE

3. Classement final

*Fiche B2 : Fiche analytique d'évaluation
Eurocentres (North 1991/1992)/Projet suisse (Schneider and North 2000)*

Fiche B3 : Evaluation globale (DIALANG)

Compétence : _____	Niveau estimé	Commentaires
Echantillon d'activité 1		
Echantillon d'activité 2		
Echantillon d'activité 3		
Echantillon d'activité 4		
Echantillon d'activité 5		
Echantillon d'activité 6		
Echantillon d'activité 7		
Echantillon d'activité 8		

Exemple d'une fiche d'évaluation simple qui demande au participant d'évaluer globalement le niveau de chaque échantillon ou tâche. On peut l'utiliser pour évaluer des performances ou des épreuves d'examen et items de tests.

Fiche B4 : Synthèse de l'évaluation globale

	<i>Astérix</i>	<i>Idéfix</i>	<i>Bécassine</i>	<i>Tintin</i>	<i>Henri 4</i>	<i>Hercule Poirot</i>	<i>Autre nom de code</i>	<i>Autre nom de code</i>	<i>Autre nom de code</i>
Premier item									
Deuxième item									
Troisième item									
Quatrième item									

Fiche B4 : Fiche de synthèse de l'évaluation globale (DIALANG)

Tableau 5.6 : Planification du temps pour l'évaluation d'échantillons de performances écrites

Recommandation : groupe de 30 participants au maximum	
<i>Activités d'introduction (Familiarisation)</i>	60 minutes
Travail sur des échantillons standards : <i>Phase 1 : Illustration avec environ trois performances standards</i> <i>Pause</i> <i>Phase 2 : Pratique sous contrôle avec environ trois à cinq performances standards</i> <i>Phase 3 : Etape libre avec environ trois performances standards</i>	60 minutes 60 minutes 60 minutes
Déjeuner	
Calibrage des échantillons locaux : <i>Evaluation individuelle et discussion de groupe de performances d'un niveau élevé, moyen et faible</i> <i>Evaluation individuelle d'environ cinq performances supplémentaires</i>	60 minutes 60 minutes

Tableau 5.7 : Documents et matériel à préparer pour l'évaluation de la production écrite

<p>Documents et matériel à préparer</p> <p><i>Photocopies pour tous les participants :</i></p> <ul style="list-style-type: none"> • <i>Grille d'évaluation (Tableau 5.8)</i> • <i>Feuilles d'évaluation pour les participants : exemples des Fiches B2 et B3</i> • <i>Choix d'échelles complémentaires pertinentes ou Tableau 4.4</i> <p><i>Auxquels s'ajoutent</i></p> <ul style="list-style-type: none"> • <i>des textes de performances standards</i> • <i>des fiches de synthèse pour le coordinateur et des transparents (Fiche B 4)</i> • <i>des textes produits localement (enregistrées et/ou sélectionnées selon les instructions des Etudes de cas)</i>
--

TABEAU 5.8 : GRILLE DES CRITERES D'EVALUATION DES PRODUCTIONS ECRITES

	Vue d'ensemble	Etendue	Coherence	Correction	Description	Argumentation
C2	Est capable d'écrire des textes élaborés, limpides, fluides et parfaitement corrects dans un style personnel, approprié et efficace et qui transmette des nuances fines de sens. Peut utiliser une structure logique qui aide le destinataire à remarquer les points importants.	Manifeste une grande souplesse pour formuler des idées sous des formes linguistiques différentes afin de transmettre avec précision des nuances fines de sens, pour insister et pour lever l'ambiguïté. Possède aussi une bonne maîtrise d'expressions idiomatiques et familières.	Est capable de créer des textes cohérents et articulés en faisant un usage complet et adéquat d'une variété de modèles d'organisation et un choix étendu de connecteurs et autres articulateurs.	Garde une maîtrise cohérente et extrêmement juste des formes de la langue même les plus complexes. Les fautes sont rares et portent sur des formes rarement utilisées.	Est capable de rédiger des histoires claires, fluides et très intéressantes ainsi que de décrire des expériences dans un style approprié avec le genre choisi.	Est capable de produire des comptes rendus, des articles et des essais élaborés, clairs et fluides pour présenter un cas ou donner une appréciation critique de propositions ou d'œuvres littéraires. Peut fournir une structure logique efficace et appropriée qui aide le lecteur à trouver les points importants.
C1	Est capable de rédiger des textes bien structurés et corrects dans l'ensemble sur des sujets complexes. Peut souligner les points pertinents les plus saillants, étendre et confirmer des points de vue de manière élaborée par l'intégration d'arguments secondaires, de justifications et d'exemples pertinents pour parvenir à une conclusion appropriée.	A une bonne maîtrise d'une gamme étendue de langue qui lui permet de s'exprimer clairement dans un style approprié sur une vaste étendue de sujets généraux, académiques, professionnels ou de loisirs sans devoir limiter ce qu'il/elle veut dire. La souplesse de style et de ton est un peu limitée.	Est capable de produire des textes clairs et fluides, bien structurés, montrant un usage maîtrisé de modèles d'organisation, de connecteurs et autres articulateurs.	Garde constamment un niveau élevé de correction grammaticale ; fautes occasionnelles en grammaire, formes familières et idiomatiques.	Est capable de rédiger des descriptions et des textes créatifs clairs, détaillés et bien construits dans un style personnel, naturel et affirmé approprié au lecteur visé.	Est capable d'exposer clairement par écrit de façon bien structurée des sujets complexes en relevant les points saillants importants. Peut étendre et confirmer des points de vue de manière élaborée par l'intégration d'arguments secondaires, de justifications et d'exemples pertinents.
B2	Possède une étendue de langue suffisante pour pouvoir faire des descriptions claires, exprimer des opinions sur les sujets les plus généraux en utilisant des formes grammaticales complexes pour le faire. Néanmoins, le discours manque d'expressivité et l'utilisation de formes plus élaborées reste stéréotypée.	Possède une étendue de langue suffisante pour pouvoir faire des descriptions claires, exprimer des opinions sur les sujets les plus généraux en utilisant des formes grammaticales complexes pour le faire. Néanmoins, le discours manque d'expressivité et l'utilisation de formes plus élaborées reste stéréotypée.	Est capable d'utiliser un nombre limité d'articulateurs pour relier ses phrases en un texte clair et cohérent bien qu'il puisse y avoir quelques « sauts » dans un texte un peu long.	Manifeste un degré relativement élevé de maîtrise de la grammaire. Ne commet pas de fautes qui causent des malentendus.	Est capable d'écrire des descriptions claires et détaillées d'événements réels ou imaginaires en établissant la relation entre des idées clairement articulées et en suivant les conventions en vigueur du genre en question. Peut faire des descriptions claires et détaillées sur un certain nombre de sujets relatifs à son centre d'intérêt. Peut écrire le compte rendu d'un film, d'un livre ou d'une pièce.	Est capable de rédiger un essai ou un rapport qui développe systématiquement une argumentation avec un éclairage approprié des points importants ainsi que des détails secondaires pertinents. Peut évaluer des idées ou des solutions différentes à un problème. Peut écrire un essai ou un rapport qui développe une argumentation, justifier ou rejeter une opinion particulière et expliquer les avantages et les inconvénients de choix variés. Peut faire la synthèse d'arguments et d'informations empruntés à des sources diverses.
B1	Possède une langue suffisante pour se débrouiller avec un vocabulaire suffisant pour s'exprimer avec quelques périphrases sur des sujets tels que la famille, les loisirs et les centres d'intérêt, le travail, les voyages et l'actualité.	Possède une langue suffisante pour se débrouiller avec un vocabulaire suffisant pour s'exprimer avec quelques périphrases sur des sujets tels que la famille, les loisirs et les centres d'intérêt, le travail, les voyages et l'actualité.	Est capable de lier une série d'éléments discrets courts en un texte linéaire articulé.	Utilise de façon raisonnablement correcte un répertoire de clichés et d'expressions associés aux situations les plus courantes. Commet occasionnellement des erreurs que le lecteur peut habituellement interpréter correctement en s'appuyant sur le contexte.	Est capable de rendre compte d'expériences, de décrire des sentiments et des réactions dans des textes simplement articulés. Peut faire la description d'un événement, d'un voyage récent – réel ou imaginaire. Peut raconter une histoire. Peut faire des descriptions sur un certain nombre de sujets courants proches de son centre d'intérêt.	Est capable d'écrire de courts essais simples sur des sujets d'intérêt général. Peut résumer avec quelque assurance des informations factuelles nombreuses, en rendre compte et donner son opinion sur des sujets courants ou pas, dans son domaine. Peut écrire sous une forme classique des comptes rendus très courts pour transmettre des informations factuelles courantes et justifier des actions.
A2	Utilise des formules de base d'expressions toutes faites mémorisées, de groupes de quelques mots et expressions afin de communiquer une information limitée dans des situations simples de la vie quotidienne.	Utilise des formules de base d'expressions toutes faites mémorisées, de groupes de quelques mots et expressions afin de communiquer une information limitée dans des situations simples de la vie quotidienne.	Est capable de lier des mots avec des connecteurs simples tels que « et », « mais » et « parce que ».	Utilise correctement des structures simples mais commet encore systématiquement des fautes élémentaires. Les erreurs peuvent quelquefois provoquer des malentendus.	Est capable d'écrire de brèves biographies simples et des poèmes simples sur les gens. Peut écrire des descriptions élémentaires très brèves d'événements, d'activités passées et d'expériences personnelles.	
A1	Possède un répertoire élémentaire de mots et d'expressions simples relatives à des questions personnelles et à des situations concrètes particulières.	Possède un répertoire élémentaire de mots et d'expressions simples relatives à des questions personnelles et à des situations concrètes particulières.	Est capable de lier des mots ou groupes de mots avec des connecteurs très élémentaires tels que « et » et « alors ».	Ne montre qu'une maîtrise limitée de quelques structures grammaticales simples et de clichés mémorisés. Les erreurs peuvent provoquer des malentendus.	Est capable d'écrire des phrases et des expressions simples sur des gens réels ou imaginaires, où ils vivent et ce qu'ils font.	

5.5.2. Formation à l'aide d'échantillons calibrés pour les Compétences de réception et la compétence linguistique

L'objectif des activités décrites dans cette partie est de s'assurer que les praticiens impliqués dans le processus de formation peuvent établir le lien entre leur interprétation des niveaux du CECR et celle d'items calibrés de sorte à pouvoir ultérieurement construire sur cette compréhension commune afin de :

- relier les épreuves ou les items pertinents produits localement aux niveaux du CECR
- acquérir une compétence pour l'élaboration d'épreuves ou d'items pouvant éventuellement se réclamer d'une liaison avec les Niveaux du CECR.

Les techniques décrites pour les épreuves et les items évaluant des compétences de réception 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 compétences 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. 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.

Une formation à l'évaluation des épreuves est nécessaire. Elle permet aux rédacteurs d'affiner leur sélection de textes et de tâches et de produire des items de tests ciblés sur un niveau spécifique du CECR. Elle permet également d'aider les producteurs d'examens et de tests à estimer le niveau des items du CECR. Comme cela a été fait avec les échantillons de performances, on mettra d'abord en place une formation avec des prototypes, des items types dont le coefficient de difficulté est connu, avant de passer au processus d'analyse d'épreuves ou d'items produits localement et dont le coefficient de difficulté est également connu (5.6.2, Troisième étape). Le Chapitre 6 sur la Validation empirique donne des informations importantes sur les différentes façons de recueillir des données adéquates pour le calibrage et de les analyser.

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

1. se rendre pleinement compte de l'étendue des sous échelles de descripteurs du CECR pour des domaines particuliers (voir Chapitre 4) ;
2. identifier la pertinence du contenu des items analysés en fonction de ce que recouvre le construct par rapport aux niveaux et aux échelles du CECR ;
3. estimer le niveau de l'item en fonction des descripteurs pertinents du CECR ;
4. examiner les raisons possibles de divergences entre les niveaux estimés et les niveaux empiriques ;
5. confirmer le niveau de difficulté par confrontation avec les données empiriques.

Il est essentiel de commencer la formation avec la compréhension de l'écrit. De même qu'il est plus facile de travailler sur des performances orales et écrites (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'épreuves ou d'items imprimés (que l'on peut voir) que d'écouter ou de visionner à de nombreuses reprises des textes et des items de compréhension de l'oral (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 l'aptitude à la compréhension de l'oral et de travailler sur des textes de compréhension de l'oral car les stagiaires auront déjà l'habitude du travail à faire.

Le coordinateur doit décider de l'organisation des stages et estimer leur durée qui dépendra de la situation et de la formation antérieure des participants. Voir la Partie 5.2.1 sur la Planification.

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 de standards, une activité de tri de descripteurs de la compétence que l'on étudie doit avoir lieu.

Le CECR fournit des échelles globales générales (par exemple « Réception », « Compréhension générale de l'écrit », « Compréhension générale de l'oral ») mais aussi des échelles spécialisées 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 »).

Les coordinateurs devront décider des échelles les plus pertinentes pour les objectifs de l'examen dans le contexte où on le passe. Le travail doit toujours commencer par l'analyse et la discussion des échelles générales (par exemple « Compréhension générale de l'oral »). Les animateurs regrouperont ensuite les sous échelles les plus pertinentes pour la compétence concernée dans la situation donnée (par exemple « Comprendre en tant qu'auditeur ») ou utiliseront les reformulations des descripteurs du CECR pour l'auto-évaluation, utilisées dans le projet DIALANG (CECR, Annexe C) et demanderont aux stagiaires de trier les descripteurs selon les 6 niveaux du CECR (voir Partie 3.2, activité f).

La standardisation des épreuves ou des items qui évaluent des compétences linguistiques devra se faire selon une approche sensiblement différente de celle adoptée pour la compréhension de l'oral ou pour la compréhension de l'écrit parce qu'il est nécessaire de préciser le type de composantes que l'on peut s'attendre à trouver aux différents niveaux. Le CECR fournit des descripteurs généraux pour des éléments de la compétence de communication langagière (CECR Partie 5.2 ; Manuel Tableau 4.3) mais les spécifications linguistiques de ce type sont propres à chaque langue. La série des spécifications du Conseil de l'Europe, Waystage, Threshold et Vantage auxquels s'ajoutent Profile Deutsch et son équivalent français apportent des informations de départ pour l'élaboration de spécifications de ce type. Le projet DIALANG a aussi élaboré un ensemble de spécifications pour 14 langues, comprenant des conseils aux rédacteurs d'épreuves ou d'items.

Le processus de standardisation se déroule également en trois phases :

- **Phase 1 : Illustration** : Première évaluation du niveau d'un texte et de l'épreuve ou 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 l'épreuve ou des items** qui l'accompagnent. Un texte ne peut pas avoir 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 être utilisé dans un examen ou un test visant un niveau donné.

Les utilisateurs se rapporteront aux fiches complétées du Chapitre 4 (Spécifications) et examineront la difficulté du texte et de la tâche par rapport aux parties appropriées du CECR. Pour la Compréhension de l'écrit, par exemple, on utilisera la Fiche A10. Le Tableau 5.9 indique les parties du CECR auxquelles se reporter :

Tableau 5.9 : Sources de références dans le CECR

Domaine	Référence dans le CECR
Situations, catégories de contenu, domaines	Tableau 5, chapitre 4, 4.1
Thèmes de communication	Listes, chapitre 4, 4.2
Activités de communication	Listes, chapitre 4, 4.3
Activités de communication et stratégies	Listes, chapitre 4, 4.4.2.2
Textes et types de textes	Listes, chapitre 4, 4.6.2 et 4.6.3
Caractéristiques du texte : longueur de la tâche, cohérence et structure des tâches	Informations, chapitre 7, 7.3.2.2
Activités	Description, chapitre 7, 7.1, 7.2 et 7.3

Cette activité doit se faire individuellement dans un premier temps. Le coordinateur, de même que précédemment, devra faire prendre conscience aux stagiaires des convergences ou des divergences de leurs évaluations.

- Il est très important que les stagiaires lisent réellement le texte et répondent individuellement aux questions qui s’y rapportent avant d’évaluer en groupe la difficulté des questions traitées et le niveau du CECR qu’elles représentent le mieux.
- Après avoir répondu aux questions, ils devraient être en mesure de comparer leur propre réponse et la réponse correcte à chaque question. Pour s’assurer d’une compréhension claire du corrigé type ou du barème, il est bon qu’une discussion précède l’évaluation de la difficulté des questions. Les participants procèdent à une évaluation individuelle avant de justifier leur décision en petits groupes ou par deux.
- Il est également essentiel que le coordinateur donne des consignes claires sous forme d’instructions précises distribuées aux stagiaires. L’épreuve ou l’item est la mise en œuvre d’un descripteur du CECR et le problème est de savoir à quel niveau l’apprenant doit se trouver pour être capable de faire ce qui est décrit (c’est-à-dire de répondre correctement).

L’instruction précise que les évaluateurs doivent appliquer individuellement pour chaque item est :

« A partir de quel niveau du CECR un candidat peut-il donner une réponse correcte à l’épreuve ou à l’item ? »

- Les participants notent individuellement leur évaluation des items, et ensuite par deux ou en petits groupes, justifient leurs décisions.
 - A la fin de l’activité, le coordinateur révèle « le » niveau sur lequel la ou les question(s) sont effectivement calibrées.
- **Phase 2 : Pratique suivie :** Une fois les phases d’illustration et de discussion achevées et un accord obtenu sur la une commune 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.

De même que pour l’évaluation des échantillons de production orale et écrite, on peut poursuivre le travail avec 4 à 6 questions ou 2 ou 3 mini tests (un texte avec plus d’une question). On demandera aux stagiaires de :

- Lire les textes et répondre aux questions correspondantes puis remplir une grille (voir ci-dessous) donnant leur évaluation de chaque item afin de :
- Identifier les descripteurs du CECR que l’item met en œuvre
- Classer chaque question sur l’un des six niveaux du CECR :

Fiche B5 : Fiche d’évaluation des questions/items (DIALANG)

Compétence _____	Descripteur mis en œuvre (Enumérer les sous échelles et le niveau)	Niveau attribué	Commentaires (En incluant les références à la Fiche A9)
Question/item 1			
Question/item 2			
Question/item 3			
Question/item 4			
Question/item 5			
Etc.			

Le travail de groupe portera sur :

- Le type de question (réponse sélectionnée, réponse construite) et comment cela affecte la difficulté de l'item.
- La mise en œuvre de divers descripteurs du CECR dans le texte et dans la tâche.
- La preuve dont on dispose pour justifier le calibrage de chaque item sur un des niveaux du CECR.
- Les autres aspects pertinents des caractéristiques de la question, du texte, de la réponse, portés dans la colonne « commentaires ».

Il faut remarquer à cet égard que les évaluateurs ont tendance à surestimer la difficulté des questions à réponse sélectionnée (par exemple à choix multiple) et de sous estimer la difficulté des réponses construites (par exemple, répondre à une question, finir une phrase). Demander aux participants de répondre effectivement aux questions avant de se lancer dans des discussions sur la difficulté peut permettre de réduire le problème. Quoi qu'il en soit, se centrer sur l'interaction entre un texte et un type d'item pour définir la difficulté – en fonction de la mise en œuvre d'un descripteur du CECR – est une sensibilisation nécessaire à ce moment de la formation.

Les coordinateurs susciteront à nouveau des commentaires et des discussions. Ils feront une synthèse claire des évaluations et les présenteront sous forme schématique non seulement pour que les stagiaires puissent les visualiser mais aussi pour en garder trace dans la documentation.

- **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. De même que pour les performances de production écrite et orale, on recommande de travailler sur des groupes de 4 à 6 items. Cela permet de centrer la discussion sur la standardisation plutôt que sur les propriétés des épreuves ou des textes. Le dernier groupe d'items devrait faire l'objet d'un large consensus.

On recommande aux participants de travailler de la même façon qu'avec les échantillons de performances, en utilisant la grille pour consigner leurs évaluations jusqu'à ce le groupe atteigne une dispersion des résultats de un niveau et demi (par exemple de A2+ à B1+).

Le coordinateur peut utiliser une Fiche d'évaluation globale comme celle qui se trouve en 5.5 (Fiche B4) afin de reporter sur un transparent les évaluations faites par les stagiaires et de faire apparaître visuellement les variations de leur accord. Cette fiche sera nécessaire à la documentation de la Deuxième étape.

5.6. Troisième étape : Calibrage des performances

Une fois que l'on a achevé la formation à la standardisation (Deuxième étape) et que l'on considère que le consensus sur l'évaluation des échantillons standards est satisfaisant (dispersion égale à un niveau et demi, c'est-à-dire de A2+ à B1+), la phase de travail sur les performances des apprenants locaux peut commencer.

On appliquera des techniques semblables à celles utilisées antérieurement pour la formation à la standardisation.

Durée

La phase de calibrage des échantillons locaux doit avoir lieu aussitôt 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 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.

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. Une fois calibrés sur le CECR, il y a des chances pour que ces échantillons prennent tout leur sens en termes de référence. On recommande en conséquence de faire un choix réfléchi des épreuves ou des items pour garantir la qualité, la représentativité (en ce qui concerne les candidats) et le contenu couvert par l'examen.

Le processus de collecte peut être très semblable à celui du processus de production d'épreuves ou 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.

Il est essentiel que les échantillons locaux de performances utilisés pour le calibrage comprennent, pour les mêmes candidats, des discours de types différents.

Pour une vidéo, cela suppose une activité avec des phases qui illustrent différents types de discours. La technique de tournage pour filmer les échantillons standards a été conçue pour éviter l'influence de l'interlocuteur locuteur natif et pour fournir un échantillon équilibré de production aussi bien que d'interaction orale.

Pour la production écrite, différents types de textes sont suggérés. Il est préférable que les échantillons d'écrit incluent à la fois des productions libres (par exemple, une lettre amicale, une description) et des activités plus formelles où les candidats suivent un modèle appris (par exemple, la lettre de confirmation d'une réservation d'hôtel). Ceci est particulièrement important, notamment aux niveaux élémentaires.

Il est essentiel de veiller à ce que les échantillons de productions recueillis pendant le processus de production, d'administration, d'enregistrement ou de documentation soient de bonne qualité et utilisables. Dans le cas des vidéos, cela implique un son et une image de qualité²⁷ ; dans le cas de textes écrits cela signifie que les performances n'ont pas été biaisées par des circonstances extérieures dues à une prolongation du temps imparti, l'usage de dictionnaires, une mauvaise écriture, etc.

Parvenir au consensus et le vérifier

En général, les procédures à suivre sont celles décrites dans la Deuxième étape : Standardisation (voir Partie 5.5 Phase 2 « Pratique » et Phase 3 « Evaluation individuelle »). Le processus à suivre avec des échantillons produits localement sera semblable à la phase de « Pratique » des ateliers sur les échantillons standards. Cela comprendra :

²⁷ Si une vidéo est ultérieurement copiée sur un « master » lui-même copié pour distribution, les utilisateurs auront alors une copie de troisième génération qui aggrave tous les défauts sonores. C'est la raison pour laquelle on recommande d'utiliser *toujours* un micro extérieur et *non* celui de la caméra. Avec un micro externe d'étendue moyenne, il est parfaitement possible d'obtenir une bonne qualité de son sans passer par un studio d'enregistrement.

- L'utilisation des mêmes outils que ceux utilisés pour la formation durant la Deuxième étape (Tableaux 5.4 et 5.5 ; Tableau 5.8 pour les performances d'apprenants ; les échelles du CECR et/ou les Tableaux 4.3 et 4.4 du Chapitre 4 pour les textes et items en langue et de réception).
- Une évaluation individuelle suivie de discussion (par deux) conduisant le 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).

Si l'on ne parvient PAS à un accord, le coordinateur doit discuter avec les stagiaires de la raison de ce problème incompatible avec leur maîtrise de l'évaluation des échantillons standards. Le coordinateur devra se prononcer sur la cause du problème et faire le nécessaire pour le résoudre. Parmi les raisons possibles et les solutions :

Problème

- Les échantillons locaux ne proposent qu'une tâche et cette tâche est trop différente des échantillons du CECR
- La grille d'évaluation (Tableau 5.5) ne semble pas appropriée pour les échantillons (par exemple contexte professionnel, tâche étroitement définie)
- Certains stagiaires appliquent d'autres normes quand ils évaluent « leurs » apprenants

Action possible

- ➔ Vérifier qu'une gamme assez large de discours est disponible. Trouver d'autres échantillons plus proches du CECR
- ➔ Réviser la grille en consultant les échelles du CECR
- ➔ Proposer un échantillon local et un échantillon du CECR pour essayer de forcer les praticiens à appliquer les mêmes normes

Analyse de données

Les évaluations des échantillons standards du CECR seront analysées statistiquement afin de (a) confirmer la relation avec les niveaux et, (b) calculer la fiabilité de chaque évaluateur (constance) et des évaluateurs entre eux.

Le degré de l'accord entre les participants sera évalué et le niveau moyen des échantillons confirmé par l'analyse des évaluations au cours du processus de calibrage. L'avantage principal est que les évaluateurs dont le comportement n'est pas constant peuvent être identifiés et qu'on peut exclure leurs évaluations de l'analyse.

Plusieurs méthodes permettent d'atteindre ce but. Outre les corrélations de fiabilité entre les évaluateurs il y a, par exemple, le modèle multiple de Rasch mis en œuvre dans des programmes tels que FACETS (Linacre 1989).

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. L'enregistrement sonore des débats du stage peut être un document utile pour préparer des notes de ce type sur chaque échantillon calibré. Le coordinateur peut 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 distribuer ces notes aux participants à l'issue du stage.

5.7. Quatrième étape : Définition des standards

Le but de la définition de standards dans le contexte de la mise en relation d'un examen avec le CECR est de définir des seuils entre les différents niveaux du CECR pour des épreuves données de l'examen dans son ensemble. Cela exige l'analyse des réponses des candidats et l'analyse des données de l'évaluation. La définition de standards ne saurait se réduire à un problème d'évaluation statistique ni à l'utilisation convenable des meilleures méthodes de la psychométrie. Il y a, de nos jours, un consensus sur le fait que les seuils doivent être définis par une démarche réfléchie de jugements experts sur le contenu de l'examen ; ils se fondent sur des données relatives à la performance des items. On ne peut pas définir les seuils en se contentant d'appliquer un algorithme.

En conséquence, la définition de standards comprend deux temps :

- Un processus d'évaluation (décrit en détail dans les parties ci-dessous).
- Une analyse de données (décrite en détail dans le Chapitre 6) permettant de valider l'exactitude des standards.

Puisque les standards sont définis plutôt que calculés statistiquement, il n'y a pas de critère absolu à l'aune duquel mesurer les résultats de la procédure de leur définition. Ce caractère de subjectivité est, en fait, présent dans toutes les décisions politiques en matière éducative, sociale et politique. Il n'y a pas non plus de manière évidente de méthode qui soit clairement supérieure pour définir des standards. Certes, depuis les années 60, un débat se poursuit sur cette question et on a pu proposer des douzaines de méthodes (voir l'Annexe B). Il existe par ailleurs quelques critères d'adéquation et de validité de seuils reconnus et ce malgré leur subjectivité. Ces critères sont examinés dans le Chapitre 6.

Ce qui est essentiel, c'est que les seuils soient définis de façon méthodique et systématique. Pour mettre en pratique une procédure de définition de standards méthodique et systématique, il faut notamment veiller à :

- La définition d'objectifs pour la procédure de décision. Ils influenceront la méthode de définition de standards choisie.
- La sélection des participants (évaluateurs)
- La formation des participants (évaluateurs)
- La définition du standard de performance
- La coordination des sessions d'évaluation
- Les procédures de recueil de données

Les sections suivantes proposent des informations et des démarches pour mener le processus d'évaluation à un degré de succès satisfaisant.

5.7.1. Méthodes de définition de standards

Les décisions relatives à la difficulté des épreuves ou des items sont prises à trois moments du processus d'élaboration de l'examen ou du test :

Activité	Moment de la démarche
a) Estimation du niveau de chaque épreuve ou item au moment de leur production ou de leur sélection ;	Rédaction des épreuves ou des items
b) Estimation de la difficulté de nombreux items ou épreuves au cours d'ateliers entre collègues en se référant à des échantillons et des items déjà calibrés ;	Première définition de standards
c) Amélioration de la précision de ces estimations à la lumière du retour d'information apporté par les réponses des candidats et par l'indice de difficulté venant des données recueillies lors de l'expérimentation	Vérification de la définition de standards

Le point (b) ci-dessus est traité dans cette partie. La confirmation de la définition de standards est traitée au Chapitre 6. Dans le contexte de la mise en relation d'un examen avec les niveaux du CECR, parmi toutes les méthodes disponibles de définition de standards, deux d'entre elles sont les plus pertinentes.

La première méthode présentée est la méthode classique d'Angoff et la seconde en est une variante telle qu'elle a été utilisée dans différents projets européens, y compris le projet DIALANG financé par l'U.E.

On parle généralement de la méthode d'Angoff comme d'une méthode « classique » parce que c'est la première méthode –encore très utilisée– de définition de standards. Elle se caractérise par les traits suivants :

- Des évaluateurs se prononcent non pas sur les performances de personnes réelles (par exemple des enseignants évaluant leurs élèves ou leurs étudiants), mais sur des personnes fictives et en particulier sur les performances d'une personne « limite ». Les performances de cette personne fictive se situent entre deux niveaux, par exemple à la frontière entre A2 et B1. Un tel sujet se situe au niveau B1 le plus faible que l'on puisse imaginer.
- Pour chaque épreuve de l'examen ou item du test, l'évaluateur doit se prononcer sur la probabilité d'une réponse correcte donnée par cette personne limite.
- Le récapitulatif de ces probabilités pour toutes les épreuves ou tous les items livre le résultat attendu de cette personne limite.
- La moyenne de ces résultats attendus donne la note seuil.

L'avantage de cette méthode est qu'il n'y a pas de candidat réel impliqué dans les évaluations de sorte que l'on peut augmenter le nombre d'évaluateurs (tandis qu'habituellement, un seul enseignant évalue des élèves ou des étudiants réels). Avec un nombre important d'évaluateurs, les tendances répétées de certains (tels que l'indulgence) se verront éliminées. Il est également possible, par des méthodes statistiques, d'identifier les évaluateurs au comportement marginal et d'éventuellement les exclure.

L'inconvénient de la méthode est que l'on demande des estimations de probabilité à des évaluateurs qui peuvent ne pas avoir une idée très nette de ce qu'on entend par probabilité. En outre, des études sur la prise de décision montrent que l'homme n'est pas très bon dans l'estimation de la probabilité, quelle que soit sa formation sur le sujet. Il semble qu'il y ait une tendance à sous estimer les probabilités extrêmes, c'est-à-dire que les probabilités proches de un sont généralement sous estimées et celles près de zéro sont surestimées.

La méthode d'Angoff originale connaît un certain nombre de variantes dont une par Angoff lui-même dans laquelle il propose des évaluations de type Oui/Non. Nedelsky a également adapté la méthode pour l'utiliser avec des questions à choix multiple. Certaines de ces méthodes sont détaillées dans la section B du dossier « Références complémentaires ».

Une variante de type Oui/Non de la méthode d'Angoff a été appliquée au tout début du projet DIALANG pour la définition de standards liés au CECR. La grande différence avec la méthode d'Angoff d'origine est qu'on ne demande pas d'estimation de probabilité aux évaluateurs. La tâche de l'évaluateur est de décider pour chaque question de l'examen ou item du test :

« A partir de quel niveau du CECR un candidat peut-il répondre correctement à la question suivante ? »

Remarquez les différences avec la méthode d'Angoff classique :

- La notion de personne « limite » n'existe pas. Elle est remplacée par un candidat fictif qui se situe, selon la perception de l'expert, à un niveau donné du CECR.
- Il n'est pas question de probabilité. Comme cela a été exposé au paragraphe 5.5.2 sur le travail avec des items standards l'item est conçu comme une mise en œuvre d'un descripteur du CECR. La question est de savoir à quel niveau l'apprenant doit être pour être capable de « faire cela » (c'est-à-dire de répondre correctement à la question). Evaluer comme « B1 » un

item particulier signifie que, dans l'esprit de l'évaluateur, un candidat de niveau B1 « est capable » de faire ce que l'item exige.

Au niveau pratique, la procédure suit un schéma déjà présenté dans les parties précédentes de ce Manuel :

- Familiarisation avec le CECR
- Formation (avec retour d'information) sur la façon d'évaluer des items anciens et nouveaux ainsi que des performances en fonction des niveaux du CECR
- Evaluation des performances et des items relatifs à l'examen ou au test qui doit être mis en relation avec le CECR

Si cette interprétation est claire et bien comprise par les évaluateurs, le recueil d'items étiquetés par exemple « B1 » peut alors être considéré comme une exigence minimale pour que quelqu'un soit positionné au niveau B1, c'est-à-dire pour qu'il ait passé le seuil A2/B1.

Bien évidemment, si nous avons dix évaluateurs, en règle générale nous aurons dix ensembles différents d'exigences minimales et il nous faudra les résumer de façon rationnelle. Une façon simple de procéder consiste à compter le nombre d'items dans chaque ensemble et à faire la moyenne. On arrondit alors la moyenne au *chiffre inférieur* pour obtenir la note seuil du test : si la moyenne est de 17,8 (que l'on considérera comme la note minimale pour être classé en B1), la note seuil est 17. (La raison pour laquelle c'est 17 et non 18 est expliquée en détail dans la Partie B des Références complémentaires).

Lorsqu'un examen ou un test est prévu pour couvrir plusieurs niveaux, les questions ou les items varieront en fonction de l'estimation de leur niveau dans le CECR. Par exemple, dans un examen ou un test qui rendrait compte de résultats aux niveaux A1 et A2, on peut appliquer directement le même principe de comptage du nombre d'items qu'on a évalués à certains niveaux pour définir à combien d'items il faut répondre correctement pour atteindre :

- un niveau de compétence au-dessous de A1
- le seuil pour A1
- le seuil pour A2 (et, en conséquence, le plafond de A1)
- Etc.

On trouvera plus de détails sur l'application pratique dans l'Annexe B. Pour une description détaillée de cette approche, on peut aussi consulter le Manuel de définition de standards pour DIALANG de F. Kaftandjieva, N. Verhelst et S. Takala.

5.7.2. Procédures pour la définition des standards

Les procédures d'évaluation sont semblables à celles de la Deuxième étape. Il faut toutefois se souvenir que la mise en relation avec le CECR peut être réalisée de différentes façons. On peut distinguer, dans l'administration de l'examen ou du test, trois principaux scénarios de complexité croissante :

- Un seul examen ou test administré lors d'une session donnée pour un niveau donné ;
- Différentes épreuves ou items venant de sessions différentes et représentant le même niveau ;
- Différentes épreuves ou items venant d'examens ou de tests représentant des niveaux différents.

Le deuxième et le troisième scénarios se prêtent à l'élaboration d'une banque d'items.

Les techniques mises en œuvre seront semblables à celles que l'on a appliquées dans la formation à la standardisation lors de la Deuxième étape. Comme dans la Deuxième étape, il faut planifier plusieurs

séances pour l'évaluation des épreuves ou des items concernant les compétences ou les copies d'examen. On recommande vivement de travailler dans un ordre logique, du plus au moins visible :

1. Questions de compréhension de l'écrit, en commençant par les textes ;
2. Questions de compréhension de l'oral ;
3. Items testant l'utilisation de la langue.

Il faut s'assurer, par des analyses appropriées, que les items sélectionnés pour la définition de standards proviennent d'examens ou de tests de bonne qualité.

Le processus de définition de standards est plus long et prend plus de temps qu'il n'est difficile ou compliqué. Comme il est crucial de le respecter de manière méthodique et systématique, on trouvera ci-dessous la liste des procédures à suivre. La liste des étapes est suivie d'une estimation du nombre d'items et du temps exigés. Cette indication est donnée afin de faciliter la planification. Les parties suivantes apportent plus de détails sur les démarches concrètes avant, pendant et après les stages d'évaluation proprement dits.

1. choisir une méthode de définition de standards, préparer le matériel nécessaire et fixer l'ordre du jour pour l'évaluation.
2. choisir un assez grand nombre d'évaluateurs experts en la matière.
3. sélectionner les échelles appropriées du CECR et les mettre à la disposition des évaluateurs
4. former le groupe à la méthode de définition des standards choisie.
5. animer les discussions du groupe : examiner les données de performances réelles (par exemple, indice de difficulté des items, indice de discrimination, analyse graphique des items avec analyse des distracteurs, courbes caractéristiques des items).
6. animer les stages d'évaluation (où l'on évalue le niveau de difficulté des items).
7. recueillir les évaluations des items faites par les évaluateurs.
8. organiser, si possible, une discussion sur le processus, recueillir toute information utile à l'élaboration du test (par exemple, à l'aide d'un questionnaire).
9. recueillir les évaluations et analyser comportement des évaluateurs.
10. définir les notes / scores seuils.
11. réunir les preuves de la validité et la documentation technique.
12. donner aux évaluateurs un retour d'information sur les données (par exemple comment les résultats d'un examen donné se sont situés par rapport aux niveaux du CECR).
13. fournir des statistiques descriptives, sur la fidélité de chaque évaluateur et des évaluateurs entre eux.

Durée

La phase d'application de la compréhension des niveaux du CECR pour la définition des standards doit avoir lieu juste après la formation à la standardisation (Deuxième étape). Il est vivement recommandé de la faire au cours du même stage, l'après-midi ou le jour suivant. Le coordinateur est à même de juger si cela est faisable ou s'il vaut mieux le faire plus tard.

Si la définition de standards 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 de leur rappeler la discussion qui a eu lieu.

Le coordinateur peut utilement compléter un tableau comme celui proposé ci-dessous afin de planifier la durée d'ensemble nécessaire. Le temps nécessaire à la formation (Deuxième étape) est indiqué en italiques au cas où il serait nécessaire de recommencer.

Tableau 5.10 Planification de la standardisation

Compétence	Deuxième partie Formation à la standardisation	Troisième partie Définition de standards
Compréhension de l'écrit	<i>Une session</i> ²⁸	Une session ou plus ²⁹
Compréhension de l'oral	<i>Une session</i>	Une session ou plus
Utilisation de la langue : grammaire ³⁰
Utilisation de la langue : vocabulaire
Etc.		

Items nécessaires

Le groupe d'items nécessaires sera différent selon le scénario choisi.

Dans le cas du premier scénario où la mise en relation se fonde sur un test donné à un niveau donné, la définition de standards sera basée sur tous les items (auxquels s'ajoute un échantillon représentatif de performances pour les composantes de production écrite et orale).

Dans le cas du deuxième et du troisième scénarios, et même si cela retarde le projet, il est important d'investir du temps et de l'énergie pour recueillir un jeu d'échantillons représentatifs de bonne qualité qui représente l'étendue complète de la compétence évaluée.

Une fois l'évaluation achevée, l'analyse montrera quelles performances et quels items s'avèrent représentatifs d'un niveau donné du CECR (c'est-à-dire qu'il y a un consensus important entre les évaluateurs).

Après la confirmation empirique de la difficulté des items, ces items peuvent servir de modèles aux rédacteurs d'items ou « d'ancrage » pour faciliter la mise en relation des nouveaux examens ou tests et du CECR (voir la Théorie de réponse aux items dans le Chapitre 6).

5.7.3. Le processus d'évaluation

Avant les stages d'évaluation

L'organisation des stages d'évaluation suppose un certain nombre d'activités que le coordinateur doit planifier avec soin :

a) Préparation des dossiers contenant les items à évaluer.

Il est important de donner un code aux items pour en faciliter l'identification et l'utilisation ultérieure. Si l'on doit utiliser ces codes dans la constitution d'une banque d'items, il se peut que le nombre de critères de description soit limité à cause des limites du logiciel (par exemple, compétence, type d'item, etc.). Les critères utilisés se retrouvent dans le code d'identification (par exemple, COA1S0024, qui signifie Item de compréhension orale – C pour « compréhension » – ayant pour cible le niveau A1 – A1 -, type d'item à réponse sélectionnée – S -, item isolé indépendant (0), numéro de l'item (024).

Il est vivement recommandé que les échantillons locaux que l'on va utiliser couvrent une gamme adéquate de types de textes et de types d'items ainsi que de niveaux du CECR. Idéalement,

²⁸ Une session est une période 4 à 5 heures, y compris une pause de 20 à 30 minutes.

²⁹ Une session est une période 4 à 5 heures, y compris une pause de 20 à 30 minutes.

³⁰ Le temps nécessaire à l'Utilisation/Usage de la langue (vocabulaire et grammaire) est estimé sur la base des échelles DIALANG pour ces composantes.

l'animateur devrait proposer un minimum de 20 items à analyser, mieux encore 8 à 10 mini-examens ou mini-tests.

b) Préparation des fiches d'évaluation sur mesure pour chaque dossier

Le coordinateur donne aux évaluateurs un numéro de code unique (par exemple, pour le français : FR01, FR02, FR03, etc.) On ne doit pas changer ces codes.

Le coordinateur fait une liste indiquant le code à gauche et le nom de l'expert à droite. On peut s'y référer au cours d'un stage. A l'issue du stage, on peut également en avoir besoin pour l'analyse de données. Pendant le stage, il est important de vérifier que les évaluateurs indiquent leur numéro de code sur tous les formulaires.

c) Organisation des lieux et du matériel nécessaire

d) Recrutement des évaluateurs

Le critère de sélection le plus important est le niveau d'expertise de l'évaluateur. Les évaluateurs doivent avoir un minimum de 5 ans d'expérience de l'enseignement.

Il est bien connu que l'évaluation de la difficulté des items est quelque chose de difficile. C'est pourquoi le nombre d'évaluateurs impliqués dans la définition des standards a un effet plus grand sur la fidélité qu'il ne l'aurait dans le cas, par exemple, de l'évaluation de performances. En conséquence, un minimum de 10 évaluateurs est requis. Il est préférable d'en recruter 15 ou plus.

Bien que ce ne soit pas indispensable, il est recommandé d'utiliser les mêmes évaluateurs pour toutes les compétences lors des démarches d'évaluation. Quand un évaluateur a accepté de faire le travail pour une compétence ou plus, il doit évaluer toutes les épreuves ou tous les items de cette compétence.

La formation est propre à chaque compétence et devra donc être reprise pour chacune des compétences même s'il s'agit des mêmes évaluateurs.

e) Fixer un budget et se mettre d'accord sur les rémunérations

Evaluer des items est un travail difficile, exigeant et qui prend du temps. Bien que de nombreux praticiens puissent se porter volontaires pour le faire bénévolement, il faut prévoir un budget minimum. Il faut annoncer clairement que la participation au processus d'évaluation est une activité professionnelle et non un service que l'on rend au coordinateur.

f) Préparer et répéter les exposés introductifs

Comme le processus est long, plus le coordinateur sera prêt avant de commencer les stages plus ils se dérouleront vite et sans à coups. La présentation des activités à réaliser fera l'objet d'une répétition de façon à ce qu'elles soient claires et rapidement comprises.

g) Essayer les démarches

Si les démarches ne sont pas familières, le coordinateur, ses assistants et quelques évaluateurs « cobayes » les répéteront. On peut éviter de nombreux problèmes sous-jacents potentiels même avec un essai à très petite échelle.

h) Préparer un formulaire pour le retour d'informations et les commentaires des évaluateurs ou pour conduire les entretiens de rapport de stage

Cela permet de recueillir de nombreuses informations utiles et intéressantes pour les utiliser ultérieurement lors de l'élaboration de l'examen. La première impression sera perdue si l'on reporte la collecte du retour d'informations.

Pendant les stages d'évaluation

Pendant les stages d'évaluation, le coordinateur doit insister sur l'importance de suivre les consignes données et doit vérifier que le processus est mené à bien comme prévu. Il est tout particulièrement important de contrôler que les formulaires d'évaluation et de retour d'informations sont dûment remplis afin que l'analyse des données ne soit pas entravée par le comportement inadéquat d'un évaluateur négligent.

Il est recommandé d'ouvrir le stage en expliquant la procédure et de répondre aux questions. L'évaluation elle-même devra se faire individuellement, sans discussion ou consensus atteint par la discussion ou la négociation.

Les fiches utilisées pour la définition de standards peuvent être une adaptation de celle utilisée pour la compréhension de l'oral.

Définition de standards	Evaluateur (code et nom)	Date :
Compétence : Compréhension de l'oral		
<p>Consignes : <i>Votre tâche consiste à évaluer -----questions items pour la compréhension de l'oral. Les questions / items sont présentés dans le dossier joint, pratiquement sous la même forme que le candidat les trouverait. Les seules exceptions sont que (a) on a ajouté un code, (b) les réponses correctes ont été cochées pour les questions à choix multiple et les items Vrai/Faux et, (c) les barèmes pour les réponses ouvertes sont indiqués. Les consignes indiquent si le document est écouté une seule ou deux fois.</i></p> <p>Ecoutez l'enregistrement et, pour chaque item, dites à partir de quel niveau du CECR un candidat peut répondre correctement à la question. Entourez ce niveau sur la fiche.</p>		

Fiche B6 : Consignes d'évaluation (Finlande)

Les évaluateurs peuvent écouter les questions de compréhension de l'oral soit en groupe soit individuellement. Elles peuvent être écoutées sur des bandes audio, des vidéos, des cédéroms ou des DVD, dans un laboratoire de langues ou sur un magnétophone avec un bon haut-parleur. On peut prévoir des pauses dans le processus d'évaluation pour avoir le temps d'évaluer. L'avantage est que le début et la fin peuvent être planifiés en avance avec exactitude. Seul l'enregistrement original est nécessaire (avec une copie de sécurité). Si on utilise des baladeurs, il faut que chaque évaluateur ait une copie de la cassette. Ce modèle peut aussi s'adapter à l'utilisation de vidéos.

Code de l'item	Niveau du CECR	Code de l'item	Niveau du CECR	Code de l'item	Niveau du CECR
COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6
COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6
COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6
COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6
COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6
COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6	COxxxxxxxxx	1 2 3 4 5 6

Fiche B7 : Evaluation des items (Finlande)

Les fiches de ce chapitre sont également disponibles sur le site Internet www.coe.int/lang/fr

Après les stages d'évaluation

Une fois l'évaluation de tous les items du dossier effectuée, l'animateur relève toutes les fiches d'évaluation et, le cas échéant, les formulaires de retour d'informations et vérifie que les consignes ont été respectées. Il commence alors à préparer l'analyse des données de l'évaluation. En même temps, il prépare également les fiches nécessaires à la session de retour d'informations pour les évaluateurs (voir ci-dessous l'analyse du comportement de l'évaluateur). Ceci doit être fait en étroite collaboration avec l'analyste des données, qui élaborera les fichiers de données nécessaires pour fixer les seuils et pour poursuivre l'analyse de données dès que les résultats du calibrage de l'examen seront disponibles.

5.7.4. Analyse des données de l'évaluation

Il y a deux résultats souhaités du processus d'analyse des données de l'évaluation. Premièrement, il est possible de réinjecter dans le processus de formation, une grande quantité d'informations concernant le comportement des évaluateurs. Deuxièmement, l'analyse révèle la difficulté des questions / items dans la pratique. Cette information est de toute évidence nécessaire pour améliorer l'exactitude des notes seuils. Mais on oublie parfois que cette information apporte un retour d'informations essentiel pour les évaluateurs les moins expérimentés. Au fur et à mesure qu'on les informe de la façon dont leur évaluation est reliée à la réalité empirique, de nombreux évaluateurs peuvent ajuster leur évaluation initiale de manière significative.

Analyse du comportement de l'évaluateur (N.B. : utiliser le plus de schémas possibles pour leur valeur illustrative)

L'animateur choisira une procédure pour le retour d'informations et la documentation parmi celles présentées ci-dessous. Comme cela a été exposé en Partie 5.5, on peut conseiller de n'utiliser que le code des évaluateurs (et non leur nom) de sorte qu'ils soient seuls à se reconnaître. On évite ainsi que l'atmosphère ne soit tendue.

Le retour d'informations peut se faire à partir d'un certain nombre d'analyses réalisées sur données de l'évaluation.

- Analyse conventionnelle des données d'évaluation pour l'ensemble du groupe
 - Distribution des fréquences
 - Moyennes / médianes / modes
 - Ecart types – de préférence aussi sous forme de boîtes à moustaches
- Vérification des modèles individuels d'évaluation – illustré par des histogrammes ou des boîtes à moustaches
 - Indulgence / sévérité relatives
 - Etendue de la fourchette utilisée (large ou étroite)

Comparaison entre les résultats estimés et les résultats observés

Une fois les épreuves et les items pré-testés ou administrés et que les données empiriques sont disponibles, on peut mettre en œuvre les procédures suivantes. Se reporter au Chapitre 6 pour un exposé plus détaillé des procédures recommandées.

- Lorsque les données de la performance ont été analysées, il est possible de comparer les niveaux de difficulté estimés et empiriques (par exemple, en utilisant les corrélations comparées au degré exact du consensus).
- la fidélité de l'évaluateur et celle des évaluateurs entre eux doivent être étudiées. Dans la plupart des cas, l'intérêt est le degré de différence entre les valeurs de l'échelle estimées et celles de l'échelle observée.
- Les variations des évaluateurs doivent être examinées :

- a) En fonction du type d'item (certains peuvent mieux réussir dans les choix multiples que dans les questions ouvertes et vice versa ; on recommande des graphiques individuels sous forme de boîte à moustaches avec le type d'item juxtaposé) ;
- b) En fonction de l'aptitude ou de la compétence (il peut y avoir différents profils d'évaluateurs concernant la compréhension de l'écrit, la compréhension de l'oral, la grammaire et le vocabulaire) ; on recommande des graphiques individuels sous forme de boîte à moustaches avec les compétences juxtaposées ;
- c) En fonction du niveau de compétence (certains évaluateurs peuvent être meilleurs sur certains niveaux de l'échelle) ; on recommande des graphiques sous forme de boîte à moustaches centrés sur le niveau.

Afin d'aider les rédacteurs de tests à développer leur compétence professionnelle, donner un retour d'informations sur les résultats – de préférence en préservant l'anonymat– et en examiner les implications pour le travail ultérieur.

Résumé

Le processus de standardisation décrit dans ce chapitre est résumé dans le tableau suivant (Tableau 10). Ce tableau fournit une vue d'ensemble des différentes étapes sous forme de liste de contrôle. Les producteurs d'examen peuvent l'utiliser afin de faciliter la visualisation du travail que la formation à la standardisation entraîne. Ce tableau peut également être utile aux coordinateurs qui l'utiliseront comme une liste de travail pour contrôler le processus et l'intégrer, dûment rempli, dans le rapport final à l'intention des professionnels.

Une fois la standardisation menée à bien, on encouragera les producteurs d'examens à passer à l'étape suivante du processus de mise en relation de leurs examens et du CECR en choisissant parmi les procédures de validation empirique décrites au Chapitre 6. Certaines de ces procédures exigent l'approche professionnelle d'un psychométricien mais d'autres appliquent des principes fondamentaux de façon plus modeste. Même les producteurs d'examen dont les ressources et l'expertise sont limitées devraient être en mesure de choisir une technique qui réponde à leurs besoins parmi les procédures examinées.

Tableau 5.11 : Récapitulatif pour la formation à la standardisation

Activité	Matériel nécessaire	Durée	Effectif	Suggestions
FAMILIARISATION	<ul style="list-style-type: none"> • Questionnaires de contrôle fondé sur des rappels du cadre de référence • Photocopies de ces listes • Photocopies des Tableaux 1 et 2 du CECR • Versions abrégées du Tableau 2 du CECR, autres échelles 	2 heures	Coordinateur Possibilité de grands groupes	Utiliser le programme d'auto formation en ligne s'il est disponible
FORMATION (Compétences de production)	<ul style="list-style-type: none"> • Vidéos de performances standards (au minimum 8) • Ecrits standards (au minimum 8) • Photocopies d'échelles spécialisées de compétence : Tableau 3 du CECR/ Tableaux 5.4 et 5.5 (performance orale) • Tableau 5.8 (performance écrite) <p>Photocopies de</p> <ul style="list-style-type: none"> • Fiches de notation des stagiaires (Fiches B2 et 3) • Fiches de notation du coordinateur (Fiche B4) <p>Photocopies d'autres échelles supplémentaires si pertinentes</p> <p>Photocopies d'échelles spécialisées de compétence :</p> <ul style="list-style-type: none"> • Compréhension générale de l'écrit • Compréhension générale de l'oral <p>Photocopies de</p> <ul style="list-style-type: none"> • Fiches de notation des stagiaires (Annexe 2) • Fiches de notation du coordinateur (Annexe 3) <p>Photocopies d'autres échelles supplémentaires si pertinentes</p> <ul style="list-style-type: none"> • Modèles d'items calibrés 	3 à 4 heures par compétence : Introduction, 30 min Calibrage, 90 min Examens locaux, 90 min	Coordinateur 30 stagiaires maximum	Traiter deux compétences par jour ou passer une demi-journée sur la formation et une demi-journée sur le calibrage d'une seule compétence.
FORMATION (Compétences de réception)	<ul style="list-style-type: none"> • Vidéos locales (au minimum 8) • Ecrits produits localement (au minimum 8) • Photocopies d'échelles spécialisées de compétence Tableau 3 du CECR/ Tableaux 5.4 et 5.5 (performance orale) • Tableau 5.8 (performance écrite) <p>Photocopies de</p> <ul style="list-style-type: none"> • Fiches de notation des stagiaires (Fiches B2 et 3) • Fiches de notation du coordinateur (Fiche B4) <p>Photocopies d'autres échelles supplémentaires si pertinentes</p> <p>Photocopies d'échelles spécialisées de compétence :</p> <ul style="list-style-type: none"> • Compréhension générale de l'écrit • Compréhension générale de l'oral <p>Echelles de compétence en langue</p> <p>Photocopies de</p> <ul style="list-style-type: none"> • Fiches de notation des stagiaires (Annexe 2) • Fiches de notation du coordinateur (Annexe 3) <p>Photocopies d'autres échelles supplémentaires si pertinentes</p> <ul style="list-style-type: none"> • Items locaux expérimentés 	3 à 4 heures par compétence : Introduction, 30 min Calibrage, 90 min Examens locaux, 90 min	Coordinateur 30 stagiaires maximum	Il est possible de traiter deux compétences par jour car les stagiaires seront maintenant familiarisés avec les niveaux du CECR et les activités de standardisation
CALBRAGE D'ECHANTILLONS DE PERFORMANCES (Production)	<ul style="list-style-type: none"> • Vidéos locales (au minimum 8) • Ecrits produits localement (au minimum 8) • Photocopies d'échelles spécialisées de compétence Tableau 3 du CECR/ Tableaux 5.4 et 5.5 (performance orale) • Tableau 5.8 (performance écrite) <p>Photocopies de</p> <ul style="list-style-type: none"> • Fiches de notation des stagiaires (Fiches B2 et 3) • Fiches de notation du coordinateur (Fiche B4) <p>Photocopies d'autres échelles supplémentaires si pertinentes</p> <p>Photocopies d'échelles spécialisées de compétence :</p> <ul style="list-style-type: none"> • Compréhension générale de l'écrit • Compréhension générale de l'oral <p>Echelles de compétence en langue</p> <p>Photocopies de</p> <ul style="list-style-type: none"> • Fiches de notation des stagiaires (Annexe 2) • Fiches de notation du coordinateur (Annexe 3) <p>Photocopies d'autres échelles supplémentaires si pertinentes</p> <ul style="list-style-type: none"> • Items locaux expérimentés 	3 à 4 heures par compétence : Introduction, 30 min Calibrage, 90 min Examens locaux, 90 min	Coordinateur 30 stagiaires maximum	Traiter deux compétences par jour ou passer une demi-journée sur la formation et une demi-journée sur le calibrage d'une seule compétence
DEFINITION DE STANDARDS POUR LES EPREUVES ET LES ITEMS (Compétences de réception=	<ul style="list-style-type: none"> • Vidéos locales (au minimum 8) • Ecrits produits localement (au minimum 8) • Photocopies d'échelles spécialisées de compétence Tableau 3 du CECR/ Tableaux 5.4 et 5.5 (performance orale) • Tableau 5.8 (performance écrite) <p>Photocopies de</p> <ul style="list-style-type: none"> • Fiches de notation des stagiaires (Fiches B2 et 3) • Fiches de notation du coordinateur (Fiche B4) <p>Photocopies d'autres échelles supplémentaires si pertinentes</p> <p>Photocopies d'échelles spécialisées de compétence :</p> <ul style="list-style-type: none"> • Compréhension générale de l'écrit • Compréhension générale de l'oral <p>Echelles de compétence en langue</p> <p>Photocopies de</p> <ul style="list-style-type: none"> • Fiches de notation des stagiaires (Annexe 2) • Fiches de notation du coordinateur (Annexe 3) <p>Photocopies d'autres échelles supplémentaires si pertinentes</p> <ul style="list-style-type: none"> • Items locaux expérimentés 	3 à 4 heures par compétence : Introduction, 30 min Calibrage, 90 min Examens locaux, 90 min	Coordinateur 30 stagiaires maximum	Il est possible de traiter deux compétences par jour car les stagiaires seront maintenant familiarisés avec les niveaux du CECR et les activités de standardisation

Chapitre 6

Validation empirique

Comme annoncé dans l'Introduction, la mise en relation d'un test et du Cadre européen commun de référence (CECR) est une entreprise complexe. L'existence d'une telle relation n'est pas un fait qu'il suffirait d'observer mais une affirmation pour laquelle le producteur du test doit apporter des preuves suffisantes, théoriques autant qu'empiriques. On peut résumer par le mot **validation** les procédures qui apportent ces preuves.

Ce chapitre a pour but de faire un bref récapitulatif de quelques concepts et approches de base qu'il faut examiner pour mettre en relation les tests et le Cadre de référence. Cette présentation est complétée par une autre, plus détaillée et plus technique, qui figure en Annexe. Les Annexes serviront régulièrement de référence à la présentation ci-dessous. La principale raison d'être de ce chapitre et des Annexes qui viennent l'appuyer est que, si ce domaine est certes technique, on peut en même temps utiliser les techniques statistiques mises en œuvre de façon presque courante (grâce à l'existence de logiciels puissants), même quand de telles applications ne semblent pas appropriées du tout. Bien évidemment, il existe déjà des travaux spécifiques sur le sujet en linguistique appliquée mais ils sont généralement conçus ou bien du point de vue de l'aide à apporter pour concevoir la recherche sur l'acquisition d'une langue seconde et autres recherches appliquées³¹ ou comme un guide général qui n'entre pas dans le détail technique des procédures mises en œuvre³². On conseille aussi très vivement aux utilisateurs de ce Manuel de consulter les *standards du bon usage* que l'on trouve dans la littérature spécialisée et comme le recommandent des organismes comme ALTE (Association des centres d'évaluation en langue en Europe, une association de la définitions qui proposent des examens de langue dans la langue de leur pays)³³ ou ILTA (International Language Testers Association : une association d'évaluateurs individuels en langue)³⁴ et l'AREA (American Educational Research Association)³⁵.

Si le travail de validation empirique basée sur les données de la performance exige une part non négligeable de savoir-faire, tous ceux qui sont impliqués dans l'évaluation en langue peuvent tirer profit de la compréhension des principes de base. Il n'y a guère de sens à présenter aux utilisateurs des « trucs » à suivre machinalement. Au cours d'une validation empirique, il faut analyser les données et les interpréter attentivement en étant conscient qu'il existe des sources possibles d'incertitude et d'erreur. Comme une action réfléchie de ce type est loin d'être facile, ce chapitre se focalisera essentiellement sur la prise de conscience du bon usage en appliquant divers outils et approches psychométriques. Toutefois, on veillera à illustrer concrètement les concepts clés chaque fois que cela sera possible.

Ce chapitre est organisé de telle sorte que l'on y trace brièvement les grandes lignes de base de la démarche. Un exposé plus technique est également disponible dans les Annexes. On espère que cette approche aidera ceux qui s'impliquent dans des projets de mise en relation à aller des principes généraux à une familiarisation plus détaillée si leur rôle dans le projet l'exige.

³¹ Par exemple, Woods, A., Fletcher, P, et Hughes A. 1989: *Statistics in Language Studies*, Cambridge, CUP; Hatch, E. et Lazarson, A. 1991: *The Research Manual: Design and Statistics for Applied Linguistics*; New York, Newbury House.

³² Par exemple Alderson, J.C., Clapham, C. et Wall, D. 1995: *Language Test Construction and Evaluation*, Cambridge, CUP; Weir, C. 1993: *Understanding and Developing Language Tests*, Hemel Hempstead UK, Prentice Hall.

³³ La Charte de ALTE fournit à ses membres une liste de contrôle qui couvre l'élaboration, la passation, la correction et la notation du test ainsi que l'analyse et la révision qui suivent.

³⁴ ILTA Footnote.

³⁵ AERA/APA/NCME 1999 *Standards for educational and psychological testing*. Washington, D.C: American Educational Research Association. (ISBN 0-935302-25-5)

Le plan de ce chapitre est le suivant :

- La Partie 6.1 introduit brièvement les concepts de validation interne et externe
- La Partie 6.2 traite les aspects et les démarches de validation interne en détail
- La Partie 6.3 traite les aspects et les démarches de validation externe en détail

6.1 Validation empirique : validation interne et externe

De manière générale, les activités exécutées en rapport avec la **Spécification** et la **Standardisation** font partie du processus de validation. Elles rassurent l'utilisateur et lui garantissent que la définition et la production du test ont été entreprises de façon méthodique et réfléchie. Toutefois, toutes ces activités peuvent avoir lieu sans que l'on observe comment se comportent les personnes qui appartiennent à la population cible lors de la passation du test. Dans un sens plus étroit alors, la validation peut être comprise comme la mise en œuvre de procédures qui montrent que les hypothèses formulées lors des phases de Spécification et de Standardisation peuvent effectivement être confirmées lorsque les données du test deviennent disponibles. Ce chapitre traitera de ces procédures sous l'intitulé de **Validation empirique**.

Considérons l'exemple d'un test de production écrite en langue étrangère. Le principal résultat de la passation du test est un nombre qui reflète la qualité de la performance au test, par exemple le nombre de réponses correctes. Ce nombre est appelé **score au test**. On peut confirmer l'hypothèse que ce test mesure la capacité d'écrire en calculant la corrélation du score au test avec une évaluation (par des pairs ou des enseignants) de la compétence à écrire. En particulier, on peut confirmer l'hypothèse que les résultats du test sont en cohérence avec les niveaux du CECR en réalisant cette évaluation en relation avec les descripteurs correspondant à différents niveaux et aspects de la capacité à écrire telle qu'elle est conçue dans le CECR (voir la grille d'évaluation de la Production écrite dans le tableau 5.8). Mais, une telle hypothèse exige que l'on fasse passer le test et que l'on recueille les évaluations sur un échantillon de candidats avant de pouvoir calculer la corrélation. Une corrélation élevée est une preuve réelle de la validité de l'hypothèse, tandis qu'une faible corrélation ne l'est pas. Il reste que l'un des problèmes est de préciser ce qu'on entend par corrélation élevée.

Le problème de la validation empirique se divise en deux parties. Dans l'exemple ci-dessus, la validation a eu lieu en reliant les résultats du test proposé à une mesure externe. Toutes les procédures qui visent à montrer les relations avec d'autres mesures existantes seront appelées **validation externe**. De telles procédures sont au cœur même de la validation externe et doivent être considérées comme indispensables.

Pourtant, avant de pouvoir entrer dans ces procédures de validation externe, il faut démontrer que le test proposé est élaboré par des professionnels. Ce professionnalisme doit être manifeste non seulement dans le soin apporté, par exemple, à la rédaction des items, mais aussi dans le travail empirique réalisé, soit, dans cet exemple, l'analyse des réponses des candidats à ces items. On peut résumer ces procédures sous l'intitulé de **validation interne**.

La validation interne est un préalable à une mise en relation acceptable avec le CECR et répond à des questions sur la qualité intrinsèque du test telles que :

- Est-ce que l'évaluation « fonctionne » à un degré acceptable de qualité ?
 - Les items sont-ils de qualité acceptable ?
 - Les meilleurs élèves répondent-ils convenablement aux questions difficiles ?
 - Les items se tiennent-ils ?
 - Le test est-il au bon niveau ?
 - Les items sont-ils vraiment au(x) niveau(x) où ils sont censés être ?
 - Le test cible-t-il correctement le(s) niveau(x) des élèves ?
 - Les résultats sont-ils fidèles ?
 - Les résultats du test sont-ils fidèles ?
 - Les résultats issus d'autres versions du test sont-ils équivalents ?

- Les résultats annoncés par des examinateurs différents sont-ils comparables ?
- Les limites établies entre niveaux sont-elles stables d'une session à une autre ?
- Peut-on être sûr que le test mesure ce qu'il prétend mesurer ?
 - Les sous-tests supposés tester des choses différentes apportent-ils des informations différentes ou tous les sous-tests sont-ils fortement corrélés ?
 - Les apprenants se centrent-ils sur ce qu'on teste ou sur quelque chose de tout à fait différent ?
 - Dans quelle mesure le test mesure-t-il les compétences souhaitées et dans quelle mesure les résultats reflètent-ils la capacité de bien faire dans une méthode particulière d'évaluation ?
- Est-ce que les différentes procédures utilisées pour le processus d'évaluation fonctionnent toutes convenablement ?
 - La sécurité du test est-elle assurée (confidentialité) ?
 - Les examinateurs et les correcteurs suivent-ils les instructions ?
 - Les examinateurs savent-ils obtenir efficacement de bonnes performances ?
- Etc.

Ces questions concernent la qualité du test et sont relatives à la validité interne : confirmer la qualité intrinsèque du test. Les méthodes quantitative et qualitative peuvent se compléter et leur apporter des réponses. D'une part, il nous semble adéquat de commencer par une analyse statistique simple qui utilise des méthodes classiques afin de définir les caractéristiques de base du test, et d'appliquer peut-être ensuite des méthodes qualitatives si les données suggèrent qu'une enquête détaillée plus poussée est souhaitable. D'autre part, il n'y a pas de raison d'entreprendre des analyses statistiques complexes sans vérifier que le comportement des candidats, des administrateurs, des examinateurs et des correcteurs/évaluateurs reflète ce que préconisent les spécifications.

On traite de ces questions relatives à la validité interne dans la Partie 6.2.

Une fois que l'utilisateur de ce manuel a le sentiment qu'on les a traitées, il peut s'occuper du problème principal auquel il est confronté : à savoir, le lien entre le test et le CECR. La question fondamentale peut se formuler comme suit :

- La performance d'un candidat à un test donné est résumée dans le score au test ;
- Le principal problème est alors de savoir comment relier ce score à un niveau du CECR. Cette note de test correspond-elle, par exemple, à un niveau de compétence B1 ou A2 ou B2 ?

Le problème n'est pas ici de nature statistique ou psychométrique mais relève essentiellement du jugement ou de la prise de décision, à savoir, la définition de standards. (je ne suis pas convaincu par la traduction de *standard-setting* par *définition de normes* dans la mesure où l'on parle bien de standards linguistiques par exemple)

La question de base a deux volets :

Un premier volet d'ordre général : comment peut-on mettre en place une démarche rationnelle qui convertisse un score à un test nouveau ou existant en niveau du CECR tel que B1 ? S'il n'y avait qu'un cas à examiner comme, par exemple, un test d'anglais pour des locuteurs suédois, ce ne serait pas un très gros problème car il y a assez de place pour élaborer des procédures de la définition de standards acceptables et plausibles et qui pourraient avoir une légitimité sociale. On a décrit des méthodes de la définition de standards plus directes dans la Partie 5.7 sur la Définition de standards. Il existe d'autres méthodes qui sont développées en Annexe, section B.

Cependant, la situation est beaucoup plus compliquée. Pour rendre la relation au CECR crédible partout en Europe, nous devons élaborer des démarches telles qu'un B1 pour l'anglais en Suède signifie la même chose qu'un B1 pour le français au Portugal. C'est ce volet de comparabilité des

niveaux du CECR d'une langue d'enseignement à l'autre et d'une langue cible à l'autre qui exige que l'on contrôle la définition de standards entrepris.

Comme on l'a exposé dans la Partie 5.7, même après les démarches de mise en relation décrites dans le Chapitre 4 (Spécification) et le Chapitre 5 (Standardisation) – et après avoir démontré la validité interne autant que les ressources disponibles le permettaient – le producteur d'examen se retrouve encore avec une série de questions potentiellement délicates :

- Les démarches utilisées pour définir les « seuils » des différents niveaux du CECR fonctionnent-elles convenablement ?
- L'interprétation des niveaux du CECR est-elle comparable à l'interprétation qu'on en fait ailleurs ?
- Y a-t-il un audit indépendant pour vérifier que les résultats reliés aux niveaux du CECR sont raisonnablement proches des résultats que la même population a obtenu dans une autre évaluation de la (des) même(s) compétence(s) ?
- Si des apprenants portugais de français devaient passer un test suédois de français, se retrouveraient-ils au même niveau que lorsqu'ils passent leur test portugais de français ?
- Etc.

Ces questions se rapportent à la **validité externe** et sont traitées dans la Partie 6.2.3.

L'étude de la validité externe est un processus permanent qui peut avoir lieu pendant un certain temps et qu'il faudra renouveler régulièrement. Le CECR est lui-même un cadre de référence ouvert qui peut évoluer avec le temps. Le processus de mise en relation systématique des examens avec le CECR dans des situations différentes en Europe, que l'on connaît par des études de cas et que ce Manuel essaie d'appuyer, peut aussi dévoiler des domaines dans lesquels il faudrait poursuivre le développement du CECR. Ces développements ultérieurs eux-mêmes devront être pris en compte lors d'une réactualisation appropriée de la mise en relation.

On trouvera en Annexe une version plus complète des idées exposées dans ce chapitre ainsi qu'une explication plus détaillée des techniques statistiques auxquelles on se réfère. On conseille aux utilisateurs de lire ce chapitre puis de se reporter aux Annexes qui correspondent à leurs besoins.

6.2 Validation interne

Démontrer la qualité intrinsèque d'un test – validation interne – est un pré requis à la mise en relation effective. La validation empirique a deux exigences :

- D'une part, il faut disposer des données de test obtenues dans un échantillon de population de même origine que la population à laquelle le test est destiné (la population cible) ;
- D'autre part, il faut une théorie comme cadre de référence général dans lequel on puisse résumer et évaluer les données recueillies.

Au cours de l'élaboration d'un test, les données empiriques peuvent trouver leur place à différentes étapes. Il faut expérimenter les items avec de « vrais » étudiants, vérifier que la formulation est correcte, recueillir des commentaires généraux, etc. Toutefois, à une certaine étape, il faut prendre des décisions quant à la version du test qui paraîtra. Normalement, ces décisions sont fondées sur les réponses d'un échantillon de sujets à un échantillon d'items ; un sous-échantillon de ces items constituera la version publiée du test ou, le cas échéant, les deux (ou plus) versions parallèles du test. On appelle généralement **échantillon de calibrage** l'échantillon de population réuni dans ce but. On utilisera les réponses de cet échantillon de calibrage pour caractériser le test mais aussi la population cible. A cet égard, deux principes sont importants :

1. la précision des hypothèses relatives au test et à la population dépend de la taille de l'échantillon de calibrage.
2. l'échantillon de calibrage doit être représentatif de la population cible.

Les deux principes sont valables pour les méthodologies de validation aussi bien externe qu'interne. En gros, on peut dire que les méthodes qualitatives sont utilisées pour établir la validité de ce qu'on mesure. On étudie en profondeur de petits échantillons avec une analyse des transcriptions et des différentes formes de catégories. La théorie de référence tendra à être descriptive en général et linguistique en particulier. Mais il n'en demeure pas moins que les échantillons doivent être représentatifs et que l'exactitude des hypothèses formulées reste dépendante de la taille de l'échantillon, puisque toute hypothèse sur la possibilité de généraliser les résultats suppose des preuves statistiques.

Les fondements théoriques de toute mesure, y compris l'évaluation en langue étrangère, sont exposés dans les théories de l'évaluation. Ces théories sont des théories statistiques ce qui signifie essentiellement que toutes les déductions et les conclusions fondées sur les résultats observés des tests comportent nécessairement une marge d'erreur. En tant que théories mathématiques, elles partent d'hypothèses et, si celles-ci sont remplies, on peut en tirer des conclusions par un raisonnement mathématique strict. Les procédures décrites dans ce chapitre sont fondées sur deux classes principales de théories statistiques de l'évaluation. La première se fonde sur des hypothèses très faibles. C'est la **Théorie classique des tests (TCT)** qui était, jusqu'à très récemment, la seule dont on disposait. La seconde appartient à une classe très large et plus récente et c'est la **Théorie de la réponse à l'item (TRI)**. Ses hypothèses sont plus rigoureuses que celles de la Théorie classique des tests mais ses résultats sont aussi plus forts.

Le reste de cette section sur la validation interne comporte cinq parties :

- dans la première partie sont présentés des éléments de la Théorie classique des tests, l'accent étant mis sur l'application pratique de la théorie. On trouvera plus de détails sur les techniques et les procédures de la TCT dans l'Annexe C.
- dans la deuxième partie, quelques-unes des principales techniques de l'analyse qualitative des tests sont brièvement introduites. Elles sont développées dans l'Annexe D.
- dans la troisième partie, c'est une théorie plus complexe qui est introduite (Théorie de la généralisabilité). On peut utiliser cette théorie avec des tests plus complexes, quand la réponse observée provient de l'intervention, dans l'évaluation, d'un ou plusieurs évaluateurs. On trouvera plus de détails sur cette théorie dans l'Annexe E.
- la quatrième partie signale un problème délicat de la TCT : que peut-on faire si le test ne mesure pas une seule « chose » (compétence) mais plusieurs ? La réponse qui a fait ses preuves jusqu'alors a été trouvée dans l'Analyse factorielle. Les méthodes de l'Analyse factorielle sont développées dans l'Annexe F.
- la dernière partie exposera la TRI avec un développement dans l'Annexe G.

6.2.1 La Théorie classique des tests

La Théorie classique des tests est utilisée depuis plus de cinquante ans comme guide à l'usage des concepteurs de tests pour comprendre les propriétés statistiques des scores aux tests et utiliser ces propriétés afin d'optimiser de différentes manières le test en cours d'élaboration. L'objectif principal de cette partie-ci est d'examiner les points principaux de la Théorie classique des tests et de mettre l'accent sur ce qu'on peut en attendre et ce qu'elle ne saurait apporter. Nous allons d'abord présenter quelques concepts fondamentaux puis continuer sur les procédures utilisées dans le cadre de référence de l'analyse classique des tests.

a) Concepts fondamentaux

Avant de mettre en œuvre des procédures, il est nécessaire de définir et comprendre les concepts fondamentaux de la Théorie classique des tests. Les concepts clés sont :

- Items
- Scores observées
- Scores vrais

- Erreur de mesure
- Variabilité : écart-type, variance
- Sources de variation
- Fidélité des scores au test

Certains de ces concepts sont plus familiers que d'autres. Néanmoins, ils peuvent tous entraîner des complications qui ne sont pas nécessairement toutes signalées dans la présentation, généralement assez courte, qu'en font les introductions à l'évaluation en langue. On conseille donc aux utilisateurs de ce Manuel de consulter la partie appropriée de l'Annexe C dans laquelle ces concepts sont définis et expliqués.

b) Procédures

Dans la Théorie classique des tests, certaines procédures sont pertinentes et fréquemment utilisées dans l'exploration de la validité interne :

- Indice p du niveau de difficulté de l'item
- Pouvoir discriminant de l'item
- Analyse graphique des items
- Estimation de la fidélité des scores aux tests
- Formule de Spearman-Brown (longueur et fidélité du test)
- Erreur standard de mesure
- Formule de Kelly (pour l'estimation des notes vraies)
- Résultats théoriques (sur la relation entre fidélité et validité)

Certaines de ces procédures sont plus courantes que d'autres (par exemple, indice p et pouvoir discriminant de l'item). D'autres sont moins fréquemment utilisées mais peuvent s'avérer utiles. Le concept de base de fidélité (ou, plus précisément, fidélité des notes de test) est plus compliqué que ne le laisse souvent croire la façon dont il est exposé dans des publications courantes. On conseille donc aux utilisateurs de ce Manuel de consulter le texte sur les procédures dans la partie appropriée de l'Annexe C dans laquelle ces concepts sont définis et expliqués.

Le **feedback pour les rédacteurs de tests** peut et doit être renvoyé à l'aide du graphique d'analyse d'item que de nombreux programmes d'analyse fournissent de nos jours. Dans ce graphique, l'information numérique utilisée pour donner les indices de discrimination est représentée par une ou plusieurs courbes comme on le voit dans les Figures 6.1 ci-après :

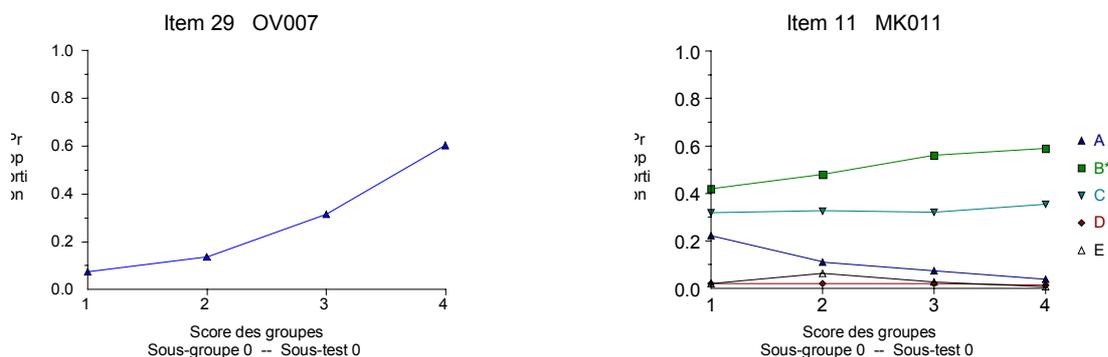


Figure 6.1 Analyse graphique d'items

Pour produire un graphique de ce type, l'échantillon total est divisé en un petit nombre de groupes homogènes (quatre dans les exemples) sur la base des scores au test. Les groupes 1 à 4 s'ordonnent du plus faible au plus fort. Dans chaque groupe, on calcule la proportion de réponses correctes que l'on reporte sur le graphique en face du numéro du groupe comme le montre la première des deux figures ci-dessus (Item 29).

On voit immédiatement que l'item 29 est relativement difficile : même dans le meilleur groupe (4) il n'y a que 60% des candidats environ qui donnent une réponse correcte. On voit aussi que l'item permet une bonne discrimination entre les meilleurs apprenants et les plus faibles. La corrélation item-test sera positive parce que plus le numéro du groupe est élevé, plus la proportion de réponses correctes est élevée. L'item est difficile mais possède un pouvoir discriminant satisfaisant.

Un graphique similaire (item 11) est proposé pour une question à choix multiple avec cinq alternatives (B est la réponse correcte, la « clé »). On voit ici immédiatement que l'item n'est pas de très bonne qualité. Son pouvoir discriminant est faible (la courbe pour l'alternative B monte très lentement). Les distracteurs D et E ne sont presque jamais choisis (et s'avèrent donc des distracteurs inutiles) tandis que le distracteur C attire des réponses à un niveau constant et assez élevé (plus de 30%) ce qui peut laisser penser que c'est un piège. Le distracteur C est trompeur d'une façon ou d'une autre. En conclusion, le graphique montre clairement que cet item mérite d'être révisé et ne saurait fonctionner comme « item type » pour aider à la formation des rédacteurs d'items.

Le **feedback sur** la définition de standards (Partie 5.6) doit être proposé en utilisant la difficulté empirique d'items individuels (indice p). Pour les items vrai/faux (items binaires ou dichotomiques), cette valeur est simplement la proportion de réponses correctes. Ainsi une valeur « p » de .99 signifie que 99% des candidats ont répondu juste à la question. Il est important de garder à l'esprit que les indices p sont dépendants de l'échantillon sur lequel ils sont calculés. Les items doivent donc être pré-testés sur un groupe approprié au niveau pour lequel le test est prévu. L'étendue précise des valeurs « p » considérée convenable dépendra de la philosophie de l'éducation et de l'enseignement sous-jacente. On obtient le maximum d'information sur un candidat lorsqu'il a une probabilité de 50% de répondre correctement (ce que l'on exploite dans la Théorie de la réponse à l'item). D'autre part, il faut équilibrer ces considérations sur la mesure avec la priorité éducative que l'évaluation doit renforcer l'apprentissage et, en conséquence, ne pas être démotivant.

Dans tous les cas, les indices p sont utiles pour identifier les items qui sont aux deux extrêmes de la difficulté (trop faciles ; trop difficiles). De tels items sont mal ciblés (ils ne sont pas au niveau auquel on les avait affectés lors de la définition de standards). Il est de bon usage d'en informer ceux qui définissent des standards dans un processus cyclique qui fait partie de leur formation continue. Souvent, l'analyse des graphiques des items peut aider ceux qui définissent les standards à voir, rétrospectivement, pourquoi tel item était mal ciblé. En conséquence, la combinaison de l'indice de difficulté et du feedback du graphique de l'item peut apporter, à terme, une contribution non négligeable à l'augmentation de la justesse des décisions relatives à la définition de standards.

Ainsi, bien que la TCT ait été remplacée par des techniques plus puissantes, on recommande toujours de la mettre en oeuvre comme première étape de la validation empirique. Une analyse fournissant les graphiques et indices p des items donnera un aperçu utile de chaque item et une indication sur la fidélité du test dans son ensemble. Ceci peut fournir un feedback très utile pour tous ceux qui sont impliqués dans le processus d'élaboration des tests.

6.2.2 Méthodes d'analyse qualitative

Les procédures d'analyse qualitative sont devenues de plus en plus populaires d'une part pour déterminer dans quelle mesure les procédures du processus de passation du test fonctionnent comme attendu et d'autre part pour étudier la qualité et la nature de l'interaction entre le candidat et la tâche ou entre le candidat et l'examineur.

On peut classer comme suit quelques-unes des méthodes les plus importantes :

a) Réflexion

- Rapports verbaux en temps réel enregistrant ce que pensent et ce que font les candidats

- Rapports verbaux en temps réel enregistrant les réactions des examinateurs aux performances et la façon dont ils interprètent et utilisent les échelles d'évaluation
- Carnets de bord des évaluateurs enregistrant comment ils interprètent et utilisent les barèmes, les échelles et les descripteurs
- Carnets de bord des apprenants

b) Analyse d'échantillons

- Analyse discursive/conversationnelle pour examiner la nature et la qualité de l'interaction, le comportement des examinateurs
- Analyse du discours du test généré par les différentes tâches, exigé pour des niveaux différents, produit par les examinateurs

c) Cadres de référence analytiques

- Cadres de référence pour concevoir le recueil d'instruments comme données (tests, questionnaires, enquêtes, etc.)
- Cadres de référence des caractéristiques des tâches

d) Méthodes de feedback

- Questionnaires
- Listes de contrôle
- Entretiens

L'analyse qualitative même d'un petit nombre de sujets peut éclairer le comportement et la manière de penser des candidats, des examinateurs, des correcteurs et des évaluateurs. Ceci aide à confirmer la nature de ce que l'on teste effectivement et contribue, en conséquence, à mettre en place la validité de construct. On trouvera le détail des techniques mentionnées ci-dessus dans l'Annexe D.

6.2.3 Théorie de la généralisabilité

Dans l'évaluation en langue, un ou plusieurs facteurs participent à une action de mesure. Ils impliquent toujours des items (tâches) et des candidats ; ils peuvent aussi impliquer des évaluateurs (et des situations encore plus complexes peuvent exister). La théorie de la généralisabilité fournit une théorie et une méthodologie par lesquelles on peut étudier ces « facettes » et établir la part que chacune d'entre elle explique du résultat. Cela peut aider à concevoir des dispositifs d'évaluation réels et à analyser leurs résultats.

Lorsqu'on n'utilise qu'une facette (les items), les résultats de la théorie de la généralisabilité sont très semblables à ceux de la Théorie classique des tests. Pour des tests dont les performances sont évaluées par des évaluateurs, la théorie de la généralisabilité peut aider à définir le nombre optimal d'évaluateurs nécessaire – permettant ainsi de faire des économies – et à identifier les meilleurs candidats pour devenir évaluateurs – améliorant alors l'efficacité. C'est pourquoi on recommande la conduite d'une étude de généralisabilité (Etude G) sur des tests qui emploient des évaluateurs pour voir comment l'ajout ou la suppression de tâches et/ou d'évaluateurs produira la fidélité la meilleure.

La Théorie de la Généralisabilité est moins connue et moins souvent utilisée pour l'évaluation des langues. Elle apporte toutefois des informations utiles que la Théorie classique des tests traditionnelle ne peut apporter. C'est la raison pour laquelle on encourage notamment les utilisateurs de ce Manuel qui partagent la responsabilité de concevoir et d'analyser des tests et des examens à consulter l'Annexe E.

6.2.4 Analyse factorielle

La performance à un test se résume habituellement à un seul nombre : le score au test. Ce score est composite parce qu'il correspond à la somme des scores des items. On est en droit de se demander si cela a un sens de réunir un certain nombre d'items dans un test et de laisser un seul nombre représenter

la performance. Qu'en est-il si le test se compose de la combinaison de deux sortes d'items dont chacun illustre un concept différent ? Un seul score a-t-il un sens ou devrait-on traiter ce test composite comme deux tests et donner deux scores différents ?

L'Analyse Factorielle (FA) est un modèle approprié pour repérer s'il y a d'autres dimensions responsables de la performance au test. En fait, ce modèle s'accorde bien avec la famille des modèles TRI qui seront présentés dans la partie suivante. On recommande de vérifier la dimensionnalité des tests que l'on essaie de relier au CECR au moyen d'une analyse factorielle. Cela aidera à décider s'il y a un sens à réunir un certain nombre d'items dans un test et de laisser un seul score représenter la performance ou si cela aurait plus de sens de rendre compte de deux scores.

Comme le modèle de l'analyse factorielle et les techniques mises en œuvre pour ces analyses sont largement utilisés – et à juste titre – mais comme il apparaît aussi malheureusement qu'il y a de nombreuses controverses à leur sujet, on recommande de consulter le texte sur l'Analyse factorielle dans l'Annexe F.

6.2.5 Théorie de réponse aux items

Si le concept de base de la Théorie classique des tests est le score vrai (à un test donné), dans la Théorie de la réponse à l'item, le concept hypothétique ou construct à mesurer est au centre de l'approche. Fondamentalement, on considère ce concept/construct comme une variable latente ou non observable (dans notre cas, la compétence en langue). Ce construct, que l'on ne peut observer directement, peut être de nature quantitative ou qualitative. S'il est qualitatif, les candidats appartiennent à des classes ou des types ; s'il est quantitatif, on peut représenter les candidats par des nombres ou des points sur une ligne droite. On ne traitera que ce dernier cas dans ce Manuel.

La Théorie classique des tests fonctionne sur des hypothèses faibles mais cela signifie que ses résultats sont aussi relativement faibles. La Théorie de réponse aux items utilise des hypothèses fortes, ce qui conduit à des résultats plus solides. Les méthodes TRI exigent que l'on vérifie le modèle relatif aux données et qu'on en démontre la validité. Si les prévisions faites sur le modèle correspondent bien aux données, c'est une preuve en faveur du modèle.

De même qu'une analyse graphique des items est un outil concret pour l'examen des items dans la Théorie classique des tests, ce que l'on a appelé les courbes de réponse aux items produites par l'Analyse TRI sont aussi très concrètes en ce sens qu'elles relient l'indice de la variable latente à la probabilité d'une réponse correcte. Plus simplement dans notre cas, elles montrent comment le niveau de compétence en langue d'un(e) candidat(e) est en rapport avec la probabilité que le (la) candidat(e) réponde correctement à chaque item en particulier.

Il existe plusieurs modèles TRI qui fonctionnent avec des paramètres différents et renferment ainsi des hypothèses différentes. Le modèle de Rasch qui n'avance qu'un seul paramètre (la difficulté) est le modèle le plus simple et le plus solide. Les modèles à deux paramètres opèrent avec les paramètres de difficulté et de discrimination considèrent que certains items doivent être plus discriminants que d'autres. Un modèle à deux paramètres appelé OPLM (One Parameter Logistic Model ; Modèle qui combine les propriétés mathématiques de Rasch avec la flexibilité du modèle logistique à deux paramètres) est particulièrement utile parce qu'il intègre la discrimination des items comme une constante qui a été fixée au cours du processus de construction du modèle, gardant ainsi la puissance du modèle de Rasch. (L'OPLM présente l'avantage supplémentaire de réduire les distorsions des scores élevées et faibles). Il existe aussi un modèle à trois paramètres qui comprend un troisième paramètre dit de conjecture. Toutefois, l'augmentation du nombre de paramètres entraîne une perte du pouvoir de mesure du modèle de Rasch plus économique et, par conséquent, plus solide.

Même quand on utilise un modèle plus solide (Rasch ; OPLM), le fait que tous les modèles TRI soient complexes signifie qu'ils avancent certaines hypothèses sur des données qu'il faudra confirmer pour éviter une distorsion des résultats. Des contrôles statistiques particuliers (= tests) sont sensibles à des

problèmes de ce type mais pas à d'autres. Il est donc de bon usage d'appliquer tous les contrôles statistiques disponibles du logiciel que l'on utilise.

L'atout pratique le plus important de l'usage de la Théorie de réponse aux items est que la comparaison des performances à des tests différents a vraiment du sens. Soit Jean qui passe un test de compréhension de l'écrit comprenant 30 questions et obtient une note brute de 22 ; Marie passe un autre test de compréhension de l'oral comprenant 34 questions et obtient une note brute de 24. Dans le cadre de référence de la Théorie classique des tests, il n'existe pas de raisonnement qui permette de déduire de ces deux informations si la compétence en compréhension de l'oral de Marie est inférieure ou supérieure à celle de Jean. Cela est néanmoins tout à fait possible avec la Théorie de réponse aux items, à la condition que les items des deux tests aient été calibrés conjointement. Bien que de nombreux logiciels permettent le calibrage conjoint de plusieurs tests, la qualité du résultat reposera fortement sur la procédure de recueil de données. On recommande vivement de prêter attention à cet aspect au moment de la constitution d'une banque d'items. On trouvera, exposés en Annexe G, quelques principes relatifs à la procédure de recueil de données.

On a besoin des méthodes de la Théorie de réponse aux items pour le calibrage des items et la construction de banques d'items. Il n'est pas rare en effet de faire des interprétations trop simplistes et injustifiées des méthodes TRI. Il faut donc les utiliser avec prudence. Pour ne citer qu'une méprise : un malentendu courant est de penser que si les données sont ajustées au modèle alors on a un niveau élevé de fidélité. Or ce n'est pas le cas : il faut toujours vérifier séparément la fidélité. Cette question ainsi que d'autres sont développées dans l'Annexe G.

Développer une banque de données est un processus plus coûteux que compliqué. Ci-dessous un aide mémoire des points principaux à ne pas oublier :

- Développer une spécification qui lui est propre
- Commander des items pour combler les lacunes en rapport avec la spécification
- Adopter un système de codage transparent pour les items
- Planifier une solide procédure de collecte des données pour permettre un « Calibrage vertical » (voir ci-dessous)
- Utiliser l'analyse pour rejeter les items mal ciblés ou autrement discutables

Une fois constituée une banque d'items d'une capacité suffisante et couvrant des niveaux différents, on peut :

- Produire des tests d'égale difficulté (= **Calibrage horizontal**)
- Produire des tests à des niveaux différents que l'on peut localiser sur la même échelle (= **Calibrage vertical**)
- Relier l'échelle de la banque d'items elle-même au CECR, évitant ainsi la nécessité de relier tous les tests séparément par un processus d'ancrage (= **Validation externe**)
- Réactualiser un test donné constitué d'items de la banque (par exemple avec des items nouvellement rédigés pour tenir compte d'un changement de programme) sans avoir à répéter la mise en relation au CECR pour les nouveaux items (= **Calibrage indirect**).

Puisque la Théorie de réponse aux items présente aussi tant d'avantages intéressants pour relier les examens de langue au CECR, on conseille de l'utiliser autant que possible. Toutefois, dans la mesure où cette méthode relativement nouvelle est de fait moins connue et risque d'être utilisée de manière moins réfléchie, on recommande vivement de lire avec soin le texte sur la Théorie de réponse aux items de l'Annexe G, en particulier à ceux qui sont impliqués dans la conception et l'analyse des données nécessaires au projet de mise en relation. Le lecteur du Manuel trouvera une explication simple pour « débutants » dans l'article de Baker (1997)³⁶. Par ailleurs, l'aide de spécialistes sera sûrement nécessaire pour une interprétation correcte des résultats.

³⁶ Baker, R. 1997 : Classical Test theory and Item Response Theory in Test Analysis. Special Report N°2 : Language Testing Update. Lancaster University, Department of Linguistics and Modern English Language.

Dans cette partie, on a exposé assez longuement la Validation interne. On y a présenté les concepts fondamentaux et discuté des procédures de base. Le but principal était de faire prendre conscience aux utilisateurs de ce Manuel des hypothèses sous-jacentes aux différents concepts en essayant de clarifier leurs fonctions et de signaler leurs limites. On espère, sur la base de ces informations, que l'on sera en mesure de prendre des décisions raisonnées et de réfléchir aux conclusions que l'on peut tirer des projets de mise en relation d'examens de langue avec le CECR.

6.3 Validation externe

Dans le contexte de ce Manuel, la validation externe porte sur la vérification de la définition de standards entreprise pour la mise en relation avec les niveaux du CECR. Les équipes d'élaboration d'un producteur d'examens ont élaboré des tests en rapport avec les spécifications (Chapitre 4) et ont défini des standards qui ont pour but de situer les candidats sur les niveaux du CECR en utilisant des procédures comme celles exposées dans la Partie 5.7 ou d'autres méthodes comme celles expliquées dans l'Annexe B. Toutefois, il reste à confirmer empiriquement qu'il y a effectivement une relation acceptable entre les seuils qui rendent compte des niveaux du CECR pour l'examen que l'on traite et les seuils du CECR eux-mêmes. C'est la seule façon de garantir une mise en relation exacte avec le CECR.

6.3.1 Principes généraux

La caractéristique principale d'un processus de validation externe de ce type est qu'elle est fondée sur des preuves **indépendantes**, fournies par le comportement face au test d'un échantillon de candidats issus de la population cible sur une mesure de critère externe. Ce critère peut être un test éprouvé et calibré sur le CECR (un test « ancre ») ou des résultats en relation aux descripteurs du CECR. Les deux possibilités seront étudiées.

La validation externe comprend deux aspects principaux :

1. **La Corrélation** : c'est la démonstration d'une relation acceptable entre les résultats du test en cours d'étude et un critère indépendant accepté comme représentant valide du concept mesuré. C'est la façon traditionnelle de prouver la validité de construct. La façon type d'exprimer une telle relation est d'utiliser un coefficient de corrélation, bien que la preuve puisse aussi s'exprimer de manière différente comme on le montrera ci-dessous. L'exigence la plus importante et la plus fondamentale est que l'on recueille les résultats au test en cours d'étude et au test-critère sur le même échantillon de candidats.
2. **Apparier les classifications avec les niveaux du CECR** : une caractéristique particulière du processus de mise en relation au CECR est la **conversion** d'un score quantitatif de test (disons allant de 0 à 40) en un nombre limité de catégories qualitatives représentées par les niveaux du CECR (A1, A2, B1, B2, etc.). Ce type de conversion est exactement ce que l'on fait à l'aide de l'une des procédures de la définition de standard examinées dans la Partie 5.7. La validation externe est la validation de cette procédure de la définition de standard. Cela revient à répondre à la question :

Dans quelle mesure peut-on être sûr qu'un candidat classé en B1 par le test-critère sera aussi classé dans la même catégorie avec le test que l'on étudie ?

Il faut se rendre compte clairement que :

- Le premier point examiné ci-dessus (la corrélation) est beaucoup plus important que le second. Si l'on trouve une corrélation de 0.40 entre le test et le critère, cela constitue une relation très faible. Dans un cas semblable, il y a probablement peu de chances de compenser cette relation très faible en ajustant la définition du standard (règle de conversion).

- D'un autre côté, la preuve convaincante d'une relation forte entre le test et le critère n'entraîne pas automatiquement que la définition du standard ait été optimale. Les évaluateurs habitués à définir un standard peuvent avoir été, par exemple, trop indulgents ou trop sévères.

Au niveau conceptuel, il est important de distinguer ces deux problèmes, à la fois pour l'évaluation du travail effectué et pour la façon dont on en rend compte.

Force de la corrélation : La Figure 6.2 donne un exemple de ce qui devrait être la première manifestation d'une validation externe : un diagramme de corrélation

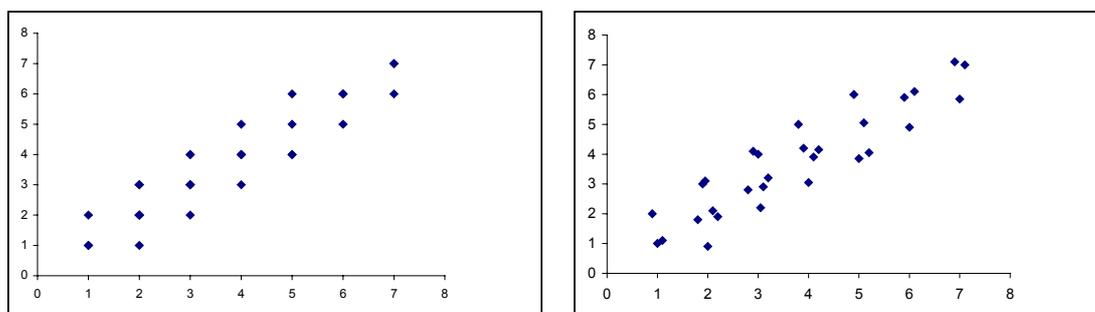


Figure 6.2. Diagrammes de corrélation

Ce diagramme permet d'apprécier graphiquement les résultats de l'étude de validation, une vision graphique facile à comprendre et qui apporte des informations précieuses. Chaque candidat est représenté par un point. Les coordonnées x représentent les résultats au test en cours d'étude, et les coordonnées y représentent les résultats au test-critère.

Un diagramme de corrélation de ce type apporte d'autant plus d'informations que les données et sur le test et sur les critères sont quantitatives (scores au test). Si les unes, ou les deux, n'ont que les niveaux du CECR – c'est-à-dire quelques catégories classées les unes après les autres, d'autres outils conviendront mieux. Deux de ces outils : les « Tables de décision » et les « Graphiques de type boîte à moustaches » seront examinés dans cette section.

Il est toujours bon, dans la mesure du possible, de produire dès le début un diagramme de corrélation pour l'étude d'une relation. Voici quelques conseils pratiques :

1. Assurez-vous que à chaque candidat correspond un point sur le graphique. Il se peut que plusieurs candidats aient obtenu le même score et il n'est pas pratique que nombre d'entre eux soient de ce fait invisibles sur le graphique. On peut y remédier en trichant un peu sur les valeurs des données que l'on a observées. Puisque les résultats au test sont des nombres entiers, on introduira une légère distorsion arbitraire dans chacune des notes qui font partie d'un groupe qui apparaît comme un seul point sur le diagramme. Il faut faire cela pour les valeurs x comme pour les valeurs y. On peut en voir l'effet en comparant les deux diagrammes de la Figure 6.2. La légère altération des valeurs dans l'étendue $- 0.2/0.2$ produit le diagramme de droite. Celui de gauche contient 18 points alors qu'il y a eu 30 observations. Dans celui de droite, toutefois, il y a 30 points, un pour chaque observation, présentés en groupes de deux (x, y).
2. Le nuage de points est de forme elliptique, ce qui est le cas habituellement dans les manuels, mais il se peut qu'il ne soit pas aussi régulier dans la pratique. Si c'est une ellipse, le coefficient de corrélation convient pour résumer la relation des deux variables : plus l'ellipse est mince (plus le nuage ressemble à une droite), plus la corrélation est élevée. Mais si la forme *n'est pas* une ellipse, le coefficient de corrélation peut sous estimer la relation (dans le cas où la corrélation n'est pas linéaire mais curviligne) ou cacher des effets spéciaux tels que des effets plafond ou plancher. Il est en conséquence de bon usage de produire les diagrammes de corrélation avec beaucoup de soin

et d'essayer de comprendre ou d'expliquer pourquoi, lorsque c'est le cas, le nuage de points qui apparaît n'est pas une ellipse.

Ce second point peut être illustré par l'observation de diagrammes de corrélation pour des coefficients de corrélation de 0.90, 0.80 et 0.70 respectivement (Figure 6.3). Bien que le nuage de points correspondant à une corrélation de 0.70 soit encore de forme elliptique, il tend à s'arrondir, et il est évident que la force de la relation s'en ressent. Avec une corrélation inférieure à 0.70, il est clair qu'on aura un nombre considérable de classements erronés lors de la conversion des résultats au test en cours d'étude en résultats au test-critère.

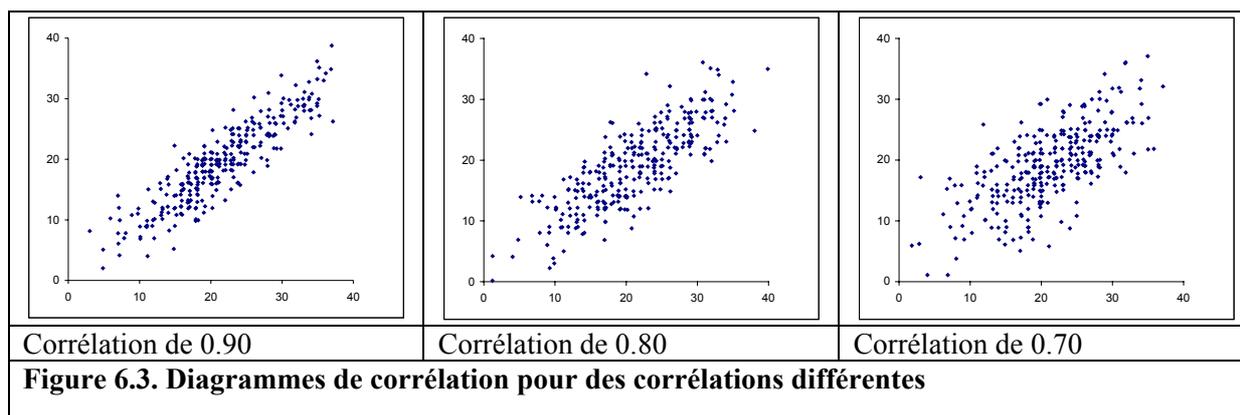


Figure 6.3. Diagrammes de corrélation pour des corrélations différentes

Solidité de l'appariement des classements : Outre la production d'un diagramme de corrélation, il est essentiel de construire un tableau de contingence bivarié montrant le classement conjugué selon les niveaux du CECR sur le test et le critère. Nous intitulerons désormais ce tableau « Table de décision ». Le Tableau 6.1 en donne un exemple. Les cellules du tableau contiennent les effectifs de candidats. Comme on peut le voir dans la première cellule, 68 candidats ont été classés au Niveau 2 par le test et les critères à la fois alors que 7 (première ligne, deuxième colonne) qui étaient classés en A2 par les critères se retrouvaient en B1 selon les résultats du test en cours d'étude.

Tableau 6.1. Table de décision : Classement selon le test et le critère

		Test			
		A2	B1	B2	Total
Critère	A2	68	7	0	75
	B1	20	90	20	130
	B2	2	3	90	95
	Total	90	100	110	300

Plusieurs remarques doivent être faites à propos de l'information qu'apporte ce type de tableau :

1. Dans une étude de validation externe, les rôles du critère et du test ne sont pas symétriques. Le critère a (et doit avoir) la priorité parce qu'il a été sélectionné pour sa relation « avérée » au CECR. Ainsi, au plan conceptuel, si les classements donnés, par le test d'une part, et par le critère d'autre part, sont en contradiction, c'est le test qui a un problème.
2. Il est facile de trouver un indice d'accord entre le classement du test et celui du critère. Les trois cellules de la diagonale du Tableau 6.1 sont celles où le test et le critère débouchent sur le même classement. Les cellules hors diagonale sont celles où le test et le critère débouchent sur des

classements différents. Le nombre de candidats dans les cellules de la diagonale de l'exemple ci-dessus est $68+90+90=248$ ce qui, dans ce cas précis, sur un total de 300, correspond à

82,7%

On doit rendre compte d'un tel pourcentage en même temps que de la « Table de décision » parce qu'il résume de façon claire l'accord des classements fondés sur le test et sur le critère. Les utilisateurs du Manuel seront peut-être étonnés d'apprendre qu'un accord sur le classement de 83% est en fait un résultat tout à fait honorable. Dans la majorité des cas, le niveau d'accord du tableau de contingence à deux variables sera notablement plus faible, même avec une corrélation élevée.

Bien évidemment, on voudra savoir pourquoi ce pourcentage n'est pas plus élevé. Une raison importante peut être que la définition de standard a une validité limitée.

- On trouve une première indication dans la comparaison des marges du tableau. Nous y voyons que, selon le test-critère, 75 candidats sont classés au niveau le plus bas (A2) tandis que, selon la procédure de la définition de standard, 90 candidats sont classés en A2. Il semble donc que le seuil qui sépare A2 de B1 dans la définition de standard soit trop élevé et qu'il pourrait vraisemblablement être abaissé. En suivant un raisonnement semblable, le tableau suggère que le seuil entre B1 et B2 est trop indulgent et classe plus facilement en B2 que le test-critère. On pourrait donc, sur la base de ce résultat, modifier les résultats reposant sur la définition de standard.
 - On doit cependant souligner qu'il faut être prudent si l'on fait cela mécaniquement. Le test-critère n'est pas infaillible même s'il jouit d'une priorité logique dans l'évaluation d'un nouveau test. Un aspect important d'une bonne procédure validation est d'essayer de trouver une explication rationnelle aux différences qui sont apparues. (Par exemple, l'équilibre entre les choix multiples et les questions ouvertes peut être différent dans le test et dans le critère et cela peut expliquer une partie des différences observées). Trouver une telle explication, aussi confortable soit-elle intellectuellement, n'est pas la même chose qu'avoir raison : l'explication peut être au mieux considérée comme une hypothèse qu'il faudra corroborer par une nouvelle preuve empirique. Cela montre que la validation n'est pas une activité exceptionnelle mais un processus permanent d'accumulation de preuves qui constituent la base scientifique et la justification de l'utilisation du test et de la prise des décisions qui sont fondées sur son résultat.
3. Si on amende la définition de standard sur la base d'une « Table de décision » comme le Tableau 6.1, c'est-à-dire si l'on modifie les seuils, la table devra être recalculée. La raison en est qu'un meilleur appariement entre les deux marges du tableau ne garantit pas automatiquement l'augmentation de l'indice d'accord (le % dans la diagonale).
 4. Même si la définition de standard est de qualité, il restera des cellules hors de la diagonale du tableau, et ce pour la même raison que la corrélation entre deux tests excellents ne sera jamais parfaite (1.0) :
 - a. Ni le test ni le critère ne sont parfaitement fidèles, et cela affectera leur corrélation. Voir l'Annexe C pour un exposé détaillé de cette question.
 - b. On traite en fait les tests, ainsi que les résultats critériés, comme des quantités unidimensionnelles (c'est-à-dire que pour les uns comme pour les autres on résume la performance par un seul score, et non par plusieurs scores organisés en profil). Mais en général ce score est hétérogène et ses composants peuvent être différents dans le test et dans le critère. Une analyse factorielle conjointe (qui utiliserait des items ou des épreuves du test et du critère) peut éclairer cette possibilité. Voir l'Annexe F pour un exposé détaillé de l'analyse factorielle.
 5. L'échantillon de l'exemple du Tableau 6.1 est de 300 et nous trouvons un classement correct pour environ 82% des candidats. Mais la visée d'une étude de validité n'est pas un échantillon

particulier mais une généralisation à la population cible. Les deux risques les plus importants qui apparaissent dans toute analyse de données sont pertinents ici aussi :

- a. Tous les pourcentages ou autres indices calculés dans l'échantillon du Tableau 6.1 sont des estimations de la valeur correspondante dans la population cible. Avec des échantillons de petite dimension, les approximations risquent d'être faibles.
 - b. Si l'on veut généraliser à la population cible, l'échantillon utilisé devra être représentatif de cette population.
6. Dans la construction de l'indice d'accord, toutes les cellules hors diagonale avaient un poids égal (elles étaient seulement décomptées). Mais il peut arriver qu'un classement à un niveau trop élevé ait d'autres conséquences (et plus sérieuses d'un point de vue social) qu'un classement à un niveau trop faible et vice-versa. Si tel était le cas, cela signifierait qu'il faut pondérer les différentes cellules hors diagonale. En conséquence, un examen attentif des cellules hors diagonale peut être important et éclairant, notamment si le tableau n'est pas symétrique (un effectif de candidats différent au dessus et en dessous de la diagonale).

Un dernier point important à souligner concernant les études de validation externe du type de celle que nous examinons ici est qu'elle peut en fait offrir un raccourci pour le processus général de mise en relation.

Eu égard au processus général de mise en relation de la Spécification, de la Standardisation et de la Validation empirique tel que décrit dans ce Manuel, la fonction de la validation externe est essentiellement de valider et, si nécessaire, d'amender la définition de standard telle qu'elle a été mise en œuvre dans l'étape de Standardisation.

On peut toutefois sauter entièrement le processus complet de la définition de standard, comme décrit dans la Partie 5.7 (Évaluer la difficulté des items et en déduire les seuils en agrégeant ces évaluations). Il est tout à fait possible de choisir l'alternative de conduire le processus de définition de standards empiriquement lors de l'étape de validation. Les étapes à suivre seraient :

- Choisir des seuils très provisoires, éventuellement arbitraires ;
- Créer une Table de décision comme le Tableau 6.1 ;
- Ajuster les seuils au test que l'on étudie pour les aligner sur ceux du critère ;
- Calculer une autre Table de décision avec les seuils réajustés ;
- Répéter le processus jusqu'à ce qu'on atteigne un pourcentage élevé de classements corrects.

Bien évidemment, il faut se méfier de ne pas entreprendre mécaniquement la tâche ci-dessus. Premièrement, il est important qu'un test ait été conçu pour évaluer un construit très semblable à celui que teste le critère. Mener à bien le processus de Spécification décrit dans le Chapitre 4 peut aider à déterminer si tel est bien le cas. Deuxièmement, comme mentionné ci-dessus, une procédure de validation de qualité implique d'être en mesure de donner une explication rationnelle pour expliquer tout changement effectué.

6.3.2 Cas pratiques

L'exemple exposé dans la partie qui précède contient l'essentiel de ce que devrait produire, en principe, une étude de validation. Il se peut que la façon de produire le diagramme de corrélation ou des graphiques équivalents et celle de présenter les Tables de décision soient, dans certains cas, plus recherchée que l'introduction ne le laisse entendre. Dans cette partie-ci on examinera un certain nombre de cas différents et on terminera par un exemple pratique démontrant qu'une bonne étude de validation, encore que non définitive, peut être menée à bien avec des ressources limitées.

On examine deux cas de mise en relation au CECR : au moyen d'un test-ancre, et d'un lien direct. On présuppose que cette distinction correspond assez bien à des situations réelles. Dans le premier cas, celui du test-ancre, on a choisi le test-ancre parce qu'il est déjà calibré sur le CECR de façon valide et

documentée. Il sert ainsi de critère dans lequel on peut avoir suffisamment confiance. C'est un exemple typique de ce qu'on appelle quelquefois mise en relation indirecte. On en traitera dans la Partie 6.3.2.1. Dans le cas de mise en relation directe, le but serait de relier le test au CECR sans test-ancre. Néanmoins, le principe est le même que dans l'autre cas : il faut une mesure critériée et, s'il n'y en a pas, il faut la créer. On exposera différentes manières de le faire dans la Partie 6.3.2.2.

6.3.2.1 Mettre en relation à l'aide d'un test d'ancrage

Dans cette procédure, le nouveau test est celui que l'on étudie et un test déjà calibré sert de critère.

Les principales étapes à suivre sont (du moins en principe) les suivantes :

1. Il faut faire passer les deux tests à un échantillon de candidats représentatif de la population ciblée par le nouveau test.
2. Il faut construire le diagramme de corrélation et calculer la corrélation correspondante. Le résultat de cet exercice doit faire l'objet d'une étude et de commentaires approfondis.
3. Si le diagramme de corrélation révèle une relation forte et régulière, on peut passer à la mise en relation des deux tests par une **régression linéaire**. La droite de régression produite par le logiciel est celle qui laisse le plus petit nombre d'erreurs (part de la variance non estimée par la régression). On peut alors utiliser cette droite pour prévoir les scores d'une mesure sur une autre. On trace des lignes droites à 90° de l'axe x (variable prédictive : le test-critère) et de la droite de régression sur l'axe y (variable prédictive : test en cours d'étude) coupé ainsi à 90°. On peut tracer ces lignes à partir des seuils entre les niveaux du CECR sur le test-critère afin d'établir des seuils sur le test en cours d'étude. Il faut garder un certain nombre de questions à l'esprit lorsqu'on suit cette procédure :
 - a. Il n'est pas nécessaire que les deux tests aient le même nombre d'items (et par conséquent le même score maximum) pour ce faire.
 - b. La technique n'est pas symétrique parce qu'elle est fondée sur la régression (linéaire). Elle fonctionne dans **une direction**. Si le Test X est le test d'origine et le Test Y le test cible, on prévoit le résultat du Test Y en utilisant une courbe de régression à partir du résultat du Test X. Supposons un score brut de 28 au Test X à partir de laquelle nous prédisons une note de 22 au Test Y. Cependant, si nous essayons d'inverser et de prévoir la note moyenne au Test X de tous les candidats qui ont obtenu 22 au Test Y, nous ne trouverons pas un résultat de 28 à moins que 28 et 22 soient les scores moyens aux Tests X et Y respectivement. Cette asymétrie dans la méthode de régression n'est pas facile à comprendre. Le Test X est la « variable indépendante » et le Test Y (cible) est la « variable dépendante ». En d'autres termes, le Test X est la variable prédictive et le Test Y est la variable prédite. On ne peut pas renverser l'ordre de la prédiction. Heureusement, dans la situation qui nous concerne, on ne souhaite pas le faire.
4. Enfin, il faut construire la « Table de décision » pour rendre compte du nombre d'erreurs de calcul, calculer le degré d'accord du classement comme exposé dans la Partie 6.3.1 et, le cas échéant, essayer d'adapter les seuils afin d'accroître le pourcentage d'accord.

Les procédures proposées jusqu'ici (fondées sur la régression linéaire) utilisent la Théorie classique des tests. Cependant, l'inconvénient principal de la régression linéaire est que la relation entre deux ensembles de scores peut ne pas être vraiment linéaire – c'est-à-dire mieux représentée par une ligne droite. Si un certain nombre de niveaux sont couverts par les deux tests que l'on apparie, il se peut alors que la fonction qui convienne soit plus proche d'une courbe en S telle que celles produites par l'application d'un modèle probabiliste (TRI) que de la droite produite par cette analyse de régression TCT. Les techniques de la TCT sont plus appropriées pour appairer deux tests de même niveau.

Il y a aussi d'autres avantages à utiliser des techniques basées sur la TRI. Avec une analyse TRI, les paramètres du modèle TRI sont évalués sur la base des réponses aux items (Théorie de la réponse à

l'item). L'un des paramètres est l'indice de difficulté des nouveaux items. C'est la raison pour laquelle la phase durant laquelle on fait une évaluation statistique des paramètres est appelée **calibrage**. La propriété particulièrement intéressante de l'utilisation des modèles TRI est que le résultat d'un test est comparable aux résultats de tout autre test auquel il est relié par des items-ancres. Tous les résultats rendent compte d'un même construct abstrait appelé variable latente. Cette variable latente est généralement symbolisée par la lettre grecque thêta. Une fois le calibrage effectué, on peut comparer chaque candidat à tous les autres candidats en estimant son indice thêta.

L'étude d'un lien basé sur un système d'ancrage avec une analyse de type TRI implique les étapes suivantes :

1. Comme précédemment, il faut faire passer les deux tests à un échantillon de candidats représentatifs de la population cible du nouveau test. Toutefois, l'un des plus gros avantages pratiques du recours à la TRI est le fait que l'on ait des résultats utiles même quand les données sont recueillies dans un « plan expérimental incomplet ». Si l'on doit corrélérer deux tests relativement longs, faire passer les deux tests en entier à chaque candidat de l'échantillon peut se révéler un véritable problème. Avec un plan expérimental adéquat, on peut obtenir de bons résultats en faisant passer deux moitiés de tests à chaque candidat. Dans l'Annexe G, on examine des plans expérimentaux de passation adéquats. On fait passer le test en respectant un plan expérimental incomplet. Tous les candidats répondront à des items du test-critère et à des items du nouveau test mais aucun candidat ne sera obligatoirement tenu de répondre à tous les items des deux tests.
2. Il faut de nouveau construire le diagramme de corrélation et calculer la corrélation correspondante ainsi qu'étudier les résultats. Cette fois, cependant, pour construire un diagramme de corrélation, au lieu de scores brutes (comme dans la partie précédente), on utilise les deux estimations de l'indice de capacité latent de chaque candidat (sur l'échelle thêta). Une estimation est basée sur le pattern de réponses des items du test critère et l'autre est basée sur celui des items du test en cours d'étude. Ainsi, pour chaque candidat, on dispose d'une estimation sur la même échelle de mesure (thêta), indépendante de l'échantillon d'items à partir duquel elle a été obtenue.

Construire le diagramme de corrélation et en rendre compte apporte des informations mais a des limites :

- Les estimations de chaque personne ne sont fondées que sur la moitié d'un test. Cela donnera des corrélations plus faibles qu'avec deux tests complets. (voir l'exposé de la formule de Spearman Brown dans l'Annexe C).
 - Les estimations jumelées ne montreront pas de différences systématiques en ce qui concerne leur valeur moyenne, même si les deux tests ne sont pas de difficulté égales.
3. Puis il faudra achever l'analyse TRI et l'estimations des paramètres pour les candidats et pour les items. Dans ce processus, il y a une possibilité de **tester statistiquement** que les deux tests traitent vraiment du même construct (voir Annexe G). Les résultats de ces tests statistiques sont pertinents pour la validation de la mise en relation. On peut affiner le construct en supprimant des candidats ou peut-être des items qui causent des distorsions. Alternativement, si les tests statistiques montrent qu'il est approprié de le faire, on peut diviser les tests en séries différentes (par exemple 2 ou 3) qui évaluent des groupes particuliers de compétences. Il y a aussi la possibilité d'examiner ce que l'on appelle « l'item à fonctionnement différentiel » en conduisant des analyses secondaires pour des sous-groupes de la population testée afin de voir si les résultats (quant aux items et aux personnes) sont différents de manière significative – c'est-à-dire indiquant un degré de biais en faveur ou au détriment de certains sous-groupes.
 4. L'utilisation de la TRI permet aussi de rendre compte de façon sensiblement plus fine de la structure des classements corrects ou erronés. Supposons le seuil (sur la variable latente, pas en termes de score brut) entre A2 et B1 à valeur de 1.3 sur l'échelle thêta. Un candidat ayant une

valeur latente exactement de ce niveau (une personne vraiment limite A2-B1) a une probabilité de 50% d'être classée en A2 quel que soit le test utilisé. Toutefois, pour des candidats plus éloignés du seuil, la probabilité d'une décision correcte s'accroîtra. La qualité discriminative d'un test s'exprime par le taux auquel la décision correcte augmente au fur et à mesure que la valeur latente s'éloigne du seuil. La Figure 6.4 illustre cela.

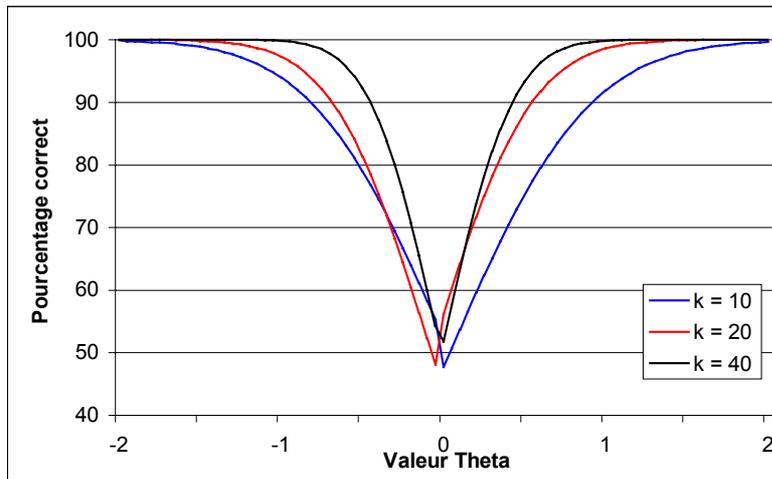


Figure 6.4 Puissance de décision

- Trois courbes sont représentées correspondant à des tests qui contiennent respectivement 10, 20 et 40 items et qui répondent tous aux attentes du modèle de Rasch. La courbe inférieure représente le test le plus court, la courbe supérieure le test le plus long. Dans l'exemple, le seuil entre les deux niveaux qui se suivent a été représenté arbitrairement à zéro pour la variable latente. Ce pourrait être, en fait, n'importe quelle valeur thêta.
- La note minimale pour être classé sur le niveau le plus élevé pour chacun des trois tests est de 6, 11 et 22 items corrects respectivement. Ces chiffres ne sont **pas précisément** les scores attendus au seuil thêta mais les **nombre entiers** les plus proches du résultat attendu. Cela explique pourquoi les minima des trois courbes ne coïncident pas avec précision et pourquoi le pourcentage de réussite au minimum n'est pas précisément 50%.
- Les trois courbes sont toutes un peu plus escarpées à gauche du seuil qu'à droite. Cela s'explique par le fait que (pour l'exemple) les paramètres de difficulté étaient concentrés principalement à gauche du seuil.

Il est indispensable qu'un utilisateur de la TRI ait une idée claire de ce qu'est la TRI et ne se laisse pas impressionner par toutes les arcanes techniques qu'on a tendance à associer aux applications de cette théorie. L'Annexe G décrit de manière détaillée les modèles TRI et on conseille aux utilisateurs de ce Manuel de la lire avant de s'engager dans la technicité des applications concrètes. Cette partie conclura donc sur quelques éclairages de points expliqués dans l'Annexe G. On a fait référence au modèle de Rasch mais la plupart des remarques suivantes s'appliquent à la plupart des modèles TRI utilisés de nos jours :

- **Il n'est pas vrai** que la seule utilisation du modèle de Rasch garantira automatiquement un bon test (fidèle, de bonne qualité).
- **Il n'est pas vrai** que l'apport de preuves écrasantes que le modèle de Rasch s'ajuste bien aux données garantira que l'on a produit un test fidèle et de qualité.

- **Il n'est pas vrai** qu'un tel ajustement statistique garantit que l'estimation des paramètres des items est stable.
- **Il n'est pas vrai** que les bons tests doivent être unidimensionnels.
- **Il n'est pas vrai** que les tests qui s'ajuste à un modèle TRI unidimensionnel sont pauvres en ce sens qu'ils ne mesurent qu'une seule dimension d'une seule compétence dans un seul domaine.
- **Il n'est pas vrai** que tous les modèles TRI sont unidimensionnels. Les modèles multidimensionnels sont moins répandus et plus difficiles à utiliser mais ils existent.

6.3.2.2 Mettre en relation par les évaluations des enseignants

Toute validation empirique suppose qu'il y ait comparaison entre deux (ou éventuellement plus) séries de résultats observés. Dans le contexte actuel, il s'agit d'une comparaison entre un test-critère, déjà relié au CECR de manière exemplaire, et le test en cours d'étude. S'il manque un critère parce qu'on ne dispose pas de test calibré, ce qui peut fort bien arriver dans cette période initiale de la mise en œuvre du CECR, il faudra construire une mesure critériée. Cela ne signifie pas que la mesure critériée doit avoir le même format que le test à valider. Si le test est un test papier/crayon donnant des notes dans la gamme, disons, de zéro à quarante, il n'est pas nécessaire que le test-critère soit un test papier/crayon ou ait la même étendue de scores observables.

Au contraire, une bonne application pourrait consister à utiliser les jugements des enseignants (par exemple) sur les capacités des candidats, le résultat étant formulé sous forme d'un classement direct sur l'un des niveaux du CECR.

Dans un tel cas, si les résultats de la mesure critériée sont formulés selon les niveaux du CECR (catégories) tandis que les résultats du test sont des scores bruts (quantitatifs), alors la construction d'un diagramme de corrélation n'apporte pas beaucoup d'informations, notamment si l'on n'a pas de gros échantillons.

Les **Graphiques de type boîte à moustaches** sont un outil d'information pour ce type de cas, comme l'illustre la Figure 6.5

- a. La Figure 6.5 est basée sur un recueil de données artificiel qui contient, en tout, les résultats de 750 candidats à un test composé de 50 items. On connaît le niveau de classement de chaque candidat par une évaluation externe servant de critère, comme l'évaluation par des enseignants par exemple, sur chacun des trois niveaux contigus. On a appelé ces niveaux 1, 2 et 3 sur l'axe x du schéma mais ils pourraient correspondre à A2, B1 et B2, par exemple.
- b. L'axe y représente les scores au test.
- c. Dans chacun des trois groupes, la distribution des notes au test est visualisée de manière graphique pratique. La ligne centrale horizontale qui traverse le point correspond à la médiane. Le bord supérieur montre le troisième quartile (75%) et le bord inférieur représente le premier quartile (25%). La boîte grisée indique ainsi où se situe les 50% au centre de la distribution. Au-dessus et au-dessous de la boîte se trouvent ce que l'on appelle les « moustaches ». Dans ce cas, le bord supérieur des moustaches correspond à la note la plus élevée (le maximum) et la partie inférieure à la note la plus faible (le minimum). Quelquefois le 10^{ème} et le 90^{ème} centiles apparaissent dans la position exacte des valeurs aberrantes.
- d. Le graphique est ici très clair : le test est tout à fait adapté pour discriminer entre les trois niveaux bien qu'il reste des chevauchements importants dans les scores obtenus par les candidats classés à différents niveaux. Ce n'est pas inhabituel.

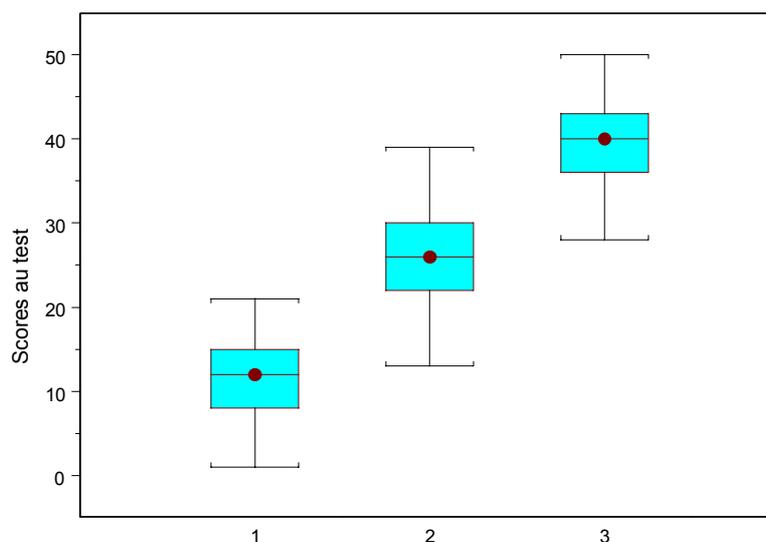


Figure 6.5 Boîtes à moustaches

Ces graphiques sont notamment utiles lorsque le critère externe est un classement par niveaux – comme c’est le cas avec l’évaluation par les enseignants ou autres évaluateurs.

En fait, ces évaluations pourraient prendre plusieurs formes, chacune ayant des avantages et des inconvénients :

1. **Echelle globale** : une seule évaluation globale qui réponde à la question « A quel niveau est cette personne ? » en utilisant les échelles et sous échelles du CECR les plus appropriées. Un « test » à un seul item.
2. **Grille analytique** : un certain nombre d’évaluations en rapport avec les échelles du CECR pour des activités différentes de communication langagière et des aspects différents de compétence langagière de communication. On peut sélectionner les échelles elles-mêmes ou la grille analytique des descripteurs qui en est dérivée telle que les Tableaux 8 (Oral) ou 9 (Ecrit) de ce Manuel et/ou le Tableau 2 du CECR. Cela donnerait au moins un « test » de 5 ou 6 items.

Si l’on utilisait une Analyse théorique de réponse aux items on pourrait profiter de ce que les descripteurs ont été calibrés à l’aide du modèle de Rasch. Les seuils des différents niveaux du CECR ont été établis sur le continuum latent défini lors du calibrage initial. En conséquence, les résultats obtenus pourraient être transformés en une estimation sur le même continuum latent en ancrant les niveaux de l’échelle aux niveaux seuils, comme le fait l’analyse du Rasch Rating Scale Model. Ces seuils apparaissent dans le Tableau 6.2.

3. **Liste de contrôle** : alternativement, on pourrait demander des évaluations de listes de contrôle de 30 à 50 descripteurs. Des listes de ce type ont été utilisées dans le projet suisse de recherche qui a calibré les descripteurs et élaboré les niveaux du CECR. Cela pourrait livrer des « tests » de 30 à 50 items.

Tableau 6.2 : Seuils de l'échelle du CECR

Niveaux communs de référence	Niveaux affinés (Suisse)	Seuil
C2		3.90
C1		2.80
	B2+	1.74
B2		0.72
	B1+	-0.26
B1		-1.23
	A2+	-2.21
A2		-3.23
A1		-4.29

Les listes de contrôle des Portfolios européens des langues qui adoptent les descripteurs suisses (et, en conséquence, les descripteurs du CECR) ou les listes de contrôle de DIALANG dans l'Annexe C du CECR pourraient être la source d'« items calibrés » dans ce but. Demeure bien évidemment la question de savoir combien, parmi les tâches représentées par l'ensemble des descripteurs d'un niveau donné, l'apprenant devrait être capable d'en effectuer afin que l'on considère qu'il a atteint le niveau en question. Il n'y a pas de réponse simple à cette question mais, à condition que la sélection des descripteurs soit relativement représentative (comme c'est le cas dans une liste de contrôle d'un Portfolio européen des langues qui suit le modèle suisse), si un apprenant « est capable » de faire ce qui est énoncé dans les deux-tiers des descripteurs, on peut considérer qu'il est au niveau en question. S'il atteint 85%, il est probablement à la limite (ou près) du niveau suivant.

Il faut ici un avertissement en ce sens que l'affirmation ci-dessus s'applique aux « niveaux critériés » de A2, B1 et B2 représentés par les descripteurs des sous échelles du CECR qui n'apparaissent *pas* au-dessus d'une ligne horizontale tracée d'un côté à l'autre de l'entrée de ce niveau. Les descripteurs qui apparaissent au-dessus de cette ligne ont été calibrés au niveau « plus » suisse entre deux « niveaux critériés » pertinents. Pour A2, par exemple, dans la logique de ce qui vient d'être dit, un apprenant qui exécuterait environ 85% du contenu des descripteurs en B2 (au-dessous de la ligne) serait sur la limite de A2+ (=A2.2). Un apprenant qui exécuterait environ 85% du contenu des descripteurs calibrés en A2+ (=A2.2) serait vraisemblablement sur la limite de B1. La même chose s'applique en B1+ et B2+.

Si l'on applique la TRI pour analyser les données des listes de contrôle, on peut éviter l'incertitude mentionnée ci-dessus. On pourrait concevoir les descripteurs comme des items de test d'une échelle de cotation puisqu'ils ont été évalués à l'origine sur une échelle à quatre points dans le projet de recherche. On pourrait considérer l'évaluation de l'expert de chaque descripteur en relation à un candidat donné comme le résultat observé du candidat à cet item. Comme l'indice de difficulté de la plus grande partie des items du

Portfolio suisse et du CECR utilisés comme items est publié³⁷ et que les niveaux du CECR ont été définis sur la même échelle avec les seuils définis décrits ci-dessus, on pourrait transformer directement le résultat du test à l'étude en niveaux du CECR. Le gros avantage de cette approche est que les descripteurs réellement sélectionnés pour procéder au jugement peuvent être des sous-ensembles arbitraires (du moins en principe) des originaux plutôt qu'ils ne nécessitent (du moins en principe) d'être l'échantillon important et représentatif de ce qui serait exigé avec l'approche critériée de 65% décrite ci-dessus.

Quelle que soit l'approche ci-dessus choisie, ou toute autre approche, pour juger d'une cotation, il est de bon usage de considérer une telle procédure d'évaluation comme une procédure ordinaire de test. Cela signifie qu'il faut faire attention à sa validité interne en tant que test. Quelques questions stimulantes et pertinentes à cet égard sont énumérées ci-dessous :

1. Il faut former les évaluateurs par les processus de Familiarisation et de Standardisation décrits dans les Chapitres 3 et 5 de ce Manuel. Ils doivent connaître le standard.
2. Si le jugement est un seul jugement global, comment peut-on alors évaluer sa fidélité ? D'un point de vue psychométrique, cela revient à utiliser un test à un seul item ; ainsi il n'y a pas lieu de calculer les indices de fidélité interne. En ce cas, il faudrait imaginer une démarche test-retest et examiner toutes sortes de problèmes sur la possibilité d'évaluations répétées.
3. Les évaluations avec des listes de contrôle de ce qu'un candidat est capable ou pas de faire exigent que l'évaluateur ait une bonne connaissance du candidat. De fait, un examinateur ne pourra évaluer qu'un nombre limité de candidats (ses propres étudiants). Un problème s'ajoute lorsque des enseignants évaluent leurs propres apprenants : il se peut qu'ils exagèrent la différence entre les apprenants les plus faibles et les plus forts. Il y a de bonnes raisons de penser que de nombreux enseignants suisses impliqués dans le projet de calibrage sont tombés dans ce travers (North 2000 ; Schneider et North 2000).
4. Quelle que soit celle des trois options que l'on adoptera, on sait que la sévérité des évaluateurs est variable. Comment peut-on alors être sûr que les différences d'indulgence et de sévérité ne sont pas à l'origine des aspects non souhaités de la variance ?

Si l'on utilise plusieurs juges qui évaluent des échantillons des comportements les plus facilement observés (par exemple, des passages de textes écrits) comme une façon d'éviter les deux problèmes décrits ci-dessus, on arrive à une situation d'élaboration de test aussi complexe que le test en cours d'étude. Néanmoins, à condition que les juges soient indépendants du processus de la définition de standard (extérieurs à la subjectivité de ce processus), formés convenablement et qu'ils disposent des outils appropriés (par exemple les Tableaux 8 et 9), cette option mérite d'être poursuivie comme cela a été fait avec succès en Finlande. En fait, le Tableau 5.8 vient d'une étude de ce type.

On peut alors porter son attention sur la variance de l'évaluateur et un vrai bon usage exigerait une Etude G (voir l'Annexe E). Si l'on utilisait une Analyse TRI, on pourrait se servir du modèle multi-facettes de Rasch tel qu'il a été opérationnalisé dans le programme FACETS. Ce modèle tient compte d'une troisième facette (l'évaluateur), évalue sa sévérité et en tient

³⁷ Les paramètres des items pour les descripteurs sont publiés dans North 2000 et dans Schneider et North 2000. Toutefois, cinq échelles de descripteurs du CECR ont été produites en combinant des descripteurs **calibrés** ou en utilisant des données partielles. Il n'existe pas de paramètres des items pour les descripteurs des échelles pour : Coopération à visée fonctionnelle ; Discussions et réunions formelles ; Maîtrise du système phonologique ; Maîtrise de l'orthographe ; Correction sociolinguistique ; Prendre des notes (conférences, séminaires, etc.) et certains de descripteurs socio linguistiques ont été **calibrés** dans un projet ultérieur (North, B. 2002 « A CEF-bases Self-assessment Tool for University entrance » in Alderson, J.C.A. (sous la direction de) *Case Studies in applying the Common European Framework*, Strasbourg, Council of Europe.

compte au moment des évaluations de la compétence des candidats. Cette méthode d'ajustement de la subjectivité a été utilisée pour étalonner l'échelle des performances du CECR enregistrées en vidéo au cours du projet de recherche suisse (voir North 2000) et disponibles pour accompagner la version expérimentale de ce Manuel.

Ces points illustrent l'éventuelle complexité d'une bonne étude de validation. Le défi principal n'est pas la sophistication technique, la taille de l'échantillon ou autres exigences psychométriques. Ce qui compte essentiellement pour une vraie validation c'est la disponibilité d'un test-critère auquel on peut se fier comme indicateur valable des niveaux du CECR.

On peut objecter qu'il est presque impossible d'élaborer un test-critère de ce type et que, en conséquence, on perd son temps à faire des efforts sur la validation empirique. Mais l'imperfection des critères est la règle générale et, en conséquence, la validation n'est pas une question de tout ou rien. La validation est un processus cumulatif d'apport de preuves et une preuve imparfaite est meilleure que pas de preuve du tout. L'essentiel est que le processus soit bien justifiés (notamment documenté) et que l'on se rende compte de ses limites.

6.3.2.3 Un exemple de fonctionnement avec un petit budget

Ce chapitre conclura sur un type d'exemple différent et qui utilise l'évaluation des enseignants. North (2000) rend compte de l'utilisation de l'évaluation des enseignants praticiens pour référencer des banques d'items en anglais, allemand, espagnol et français sur l'échelle des Eurocentres. Cet exemple montre comment l'étalonnage de principes peut s'appliquer avec efficacité et économie même dans un contexte « faiblement développé ». On a utilisé les évaluations des enseignants pour vérifier, par une validation externe indépendante, la définition de standard dans la banque d'items. On avait fixé des seuils provisoires entre les niveaux sur la base de la difficulté prévue des items, pondérée par les données de l'analyse. La Figure 6.6, d'après North (2000)³⁸, montre les résultats de la banque d'items provisoire (test en cours d'étude) sur l'axe Y et les évaluations des enseignants (critère) sur l'axe X.

Figure 6.6 Référencer une banque d'items sur l'évaluation des enseignants

Banque d'items											r=0.93 n=68
	9									2	
	8									1	
	7							8	3		28 sujets seulement sur 68 (41%) ont exactement la même note
	6					2			8		
	5				1	8	2	1			
	4			4		4					Relation régulière
	3		2	6	5						
	2	1	1	5							8 apprenants sont placés au Niveau 7 par les enseignants et au Niveau 6 par le programme
	1	4									
		1	2	3	4	5	6	7	8	9	
Evaluation de l'enseignant : Sprachkenntnisse											

³⁸ North, B. : Linking Language Assessments : an example in a low stake context. System 28, 555-577.

Les deux ensembles de résultats sont présentés en termes de niveaux sur l'échelle en question. Cela signifie que le test et les enseignants utilisent tous les deux la même mesure. Si l'on voulait ajuster les seuils de la banque par rapport à l'évaluation des enseignants, il faudrait confronter les résultats thêta de la banque d'items et les évaluations des enseignants à l'aide d'un graphique de type boîte à moustaches (voir la Figure 6.5 ci-dessus).

Toutefois, la Figure 6.6 montre clairement la relation précise entre le classement de la banque d'items et celui des enseignants. La relation paraît régulière, avec une corrélation de 0.93. Néanmoins, seulement 28 des 68 sujets (41%) ont réellement reçu exactement le même niveau malgré la corrélation élevée. Il y a 8 apprenants placés au Niveau 7 par les enseignants et au Niveau 6 par le logiciel. Ceci provient d'un enseignant indulgent. Pourtant, même si ces 8 candidats étaient « à la bonne place » sur le tableau il y aurait encore seulement 50% des élèves qui auraient reçu exactement le même niveau de l'enseignant et du test.

C'est là le problème fondamental du calibrage des échelles pour établir des équivalences : des corrélations élevées masquent des différences importantes dans les niveaux attribués. Une partie de la raison pour laquelle si peu d'apprenants ont toutefois reçu le même classement de la banque d'items et des enseignants provient du degré de finesse de l'échelle. L'échelle des Eurocentres divise les niveaux du CECR en deux (à l'exception du Niveau A1). Si l'on crée une Table de décision qui n'utilise que les niveaux du CECR, comme dans le Tableau 6.2, la proportion de classements corrects augmente considérablement (de 41% à 73%) puisque 50 des 68 apprenants se trouvent maintenant au même niveau selon la banque d'items et les enseignants³⁹.

Tableau 6.3 : Figure 6.6 sous forme de Table de décision

Test (Banque d'items)

	A1 (1)	A2 (2 & 3)	B1 (4 & 5)	B2 (6 & 7)	C1 (8 & 9)	Total
A1 (1)	4	1				5
A2 (2 & 3)		14	4			18
B1 (4 & 5)		5	13	2		20
B2 (6 & 7)			3	16		19
C1 (8 & 9)				3	3	6
Total	4	20	20	21	3	68

On peut tirer trois conclusions de l'exemple ci-dessus.

- Premièrement, il s'agissait d'une étude très simple, supposant peu d'organisation logistique et virtuellement pas de dépense. Et pourtant, le « bénéfice » en termes de confirmation de la définition du standard adopté dans la banque d'items a été considérable.

³⁹ L'expression des niveaux des Eurocentres dans les termes du CECR se justifie par le fait qu'un nombre considérable de descripteurs du CECR viennent de l'échelle des Eurocentres. C'est parce que les descripteurs des Eurocentres ont mieux survécu au processus de validation qualitative que la plupart des autres venus d'autres sources car la formulation des Eurocentres se voulait concrète et positive. La corrélation des classement de 73 descripteurs d'interaction et de production est de 0.88. Le classement partagé sur une Table de décision est de 70% (voir North 2000 : 337).

- Deuxièmement, le calcul d'un indice du coefficient de corrélation peut se faire sur Microsoft Excel qui propose deux formules dans ce but, CORREL et PEARSON. Les techniques de mises sous forme graphiques utilisées dans les deux figures qui affichent les résultats peuvent se faire, comme ci-dessus, avec un tableau de Microsoft Word. Ainsi, puisqu'on n'exige rien de plus compliqué que Microsoft, ce type de validation externe devrait être à la portée de ces praticiens des langues étrangères qui ne se considèrent pas comme particulièrement matheux.
- Troisièmement, lorsqu'on met sous forme graphique une concordance de décision critériée dans un but de validation externe dans le cadre de ce manuel, il ne faut utiliser que les niveaux complets du CECR. L'utilisation de niveaux plus fins a un sens au plan pédagogique comme on l'a montré dans le Chapitre 3 du CECR mais cela exige un degré plus élevé de puissance de décision. C'est pourquoi les évaluations de « haut niveau » (examens) tendent à utiliser des niveaux larges afin de réduire le nombre d'erreurs de classement et c'est la raison pour laquelle le CECR et le Portfolio européen des langues, en tant qu'outils officiels de compte rendu, n'ont que six niveaux critériés.

6.4 Conclusion

Dans la partie qui précède, on a assez largement examiné la validation externe pour compléter l'exposé de la validation interne. Comme avec la validation interne, on a présenté les concepts de base et étudié les procédures fondamentales. Le but principal était de faire prendre conscience aux utilisateurs de ce Manuel quelles sont les hypothèses sous jacentes à différents concepts et approches de la validation externe. Au cours de ce processus, on a essayé d'éclairer leurs fonctions et de montrer leurs limites. Les auteurs de ce Manuel espèrent qu'en utilisant ces informations, les professionnels de l'enseignement des langues qui se sont lancés dans des projets pour relier les examens de langue au CECR peuvent les poursuivre méthodiquement et se trouvent en meilleure position pour prendre des décisions réfléchies sur la façon de relier les examens au CECR et de vérifier la validité de cette mise en relation.

Chapitre 7

Grandes lignes du rapport

7.1. Introduction

Le but de ce chapitre est de donner aux utilisateurs de ce Manuel des lignes directrices pour rendre compte du contenu de l'examen, des démarches suivies pour le relier aux niveaux du CECR et pour présenter schématiquement le niveau du CECR auquel l'examen se situe.

Ces lignes directrices varieront en fonction des destinataires du rapport. On peut en distinguer de deux sortes :

- Le grand public : candidats, utilisateurs du test (avec des intérêts différents, par exemple les parents, les écoles de langues ou les universités, les employeurs, ...)
- Un public plus spécialisé : praticiens de l'enseignement et de l'évaluation.

Un rapport remplit différentes fonctions qui varieront selon le destinataire :

En ce qui concerne le grand public :

- Un **candidat** sera intéressé par un *court rapport* qui le renseigne sur la qualité d'un examen et lui donne une vue d'ensemble des démarches suivies pour relier cet examen au CECR. Cela peut l'aider à choisir l'examen qui répondra à ses besoins.
- Un **utilisateur du test** peut être intéressé par la même chose. Les écoles et les universités seront intéressées par un *rapport court* qui leur donne assez de renseignements pour choisir l'examen qui convient comme, par exemple, le niveau d'entrée exigé ou le niveau de fin d'études donné comme objectif. Les employeurs accorderont de l'importance à une information juste sur le niveau de compétence langagière de leur employé(e) par rapport à la compétence langagière attendue dans le poste de travail.

En ce qui concerne le public plus spécialisé, on peut distinguer deux grandes finalités :

- **Finalité interne** : à l'intérieur de l'organisme d'évaluation qui utilise ce Manuel, on peut faire usage d'un *rapport plus détaillé* comme outil de réflexion. Un tel rapport peut aider à découvrir les forces et les faiblesses de l'examen et celles de la mise en relation avec le CECR. Il peut contribuer au processus continu de l'amélioration permanente de la qualité de l'examen en question (par exemple, construction du test, élaboration du test, feedback pour les rédacteurs d'items et les évaluateurs, etc.). Il peut aussi contribuer à l'amélioration des différentes démarches mises en œuvre dans la mise en relation (par exemple, la définition de seuils).
- **Finalité externe** : les praticiens de l'enseignement et/ou ceux de l'évaluation seront peut-être intéressés par un *rapport plus détaillé* qui les renseigne sur le contenu et la qualité d'un examen et sur les procédures suivies pour relier l'examen et le CECR. Un rapport de ce type apporterait aussi des informations sur la qualité du processus de mise en relation au domaine de l'évaluation en langues – en ajoutant l'expérience de l'utilisation des différentes techniques mises en œuvre. Il aiderait ainsi les établissements de langues à décider quel examen est le plus approprié pour leurs apprenants qui veulent être capables de bien fonctionner dans un certain domaine social. Il aiderait aussi d'autres organismes d'évaluation en contribuant au savoir et à l'expérience relatifs à la mise en relation des examens et du CECR. Il peut aussi enrichir l'inventaire de techniques suggérées qui proposent des idées que l'on pourrait inclure dans une édition future de ce Manuel.

Il faudrait que les producteurs d'examen tiennent compte de cette situation et acceptent de rédiger les deux types de rapports. On recommande aussi vivement de constituer un dossier ou portfolio du processus de mise en relation qui contienne les matériels utilisés, les données, les comptes rendus, etc. De la sorte, les utilisateurs du Manuel seront mieux à même de répondre aux questions à venir des collègues praticiens qui veulent essayer de faire le même choix qu'eux et/ou d'utiliser leur travail.

1. Rapport pour le grand public

Rapport aussi bref et visuel que possible et qui donne clairement au grand public les informations suivantes :

- Les caractéristiques principales de l'examen, son contenu, sa forme,
 - Son niveau par rapport au CECR,
 - La démarche suivie pour le mettre en relation avec le CECR.
- +adresse (ou téléphone ou site) pour plus d'information

On trouvera en annexe les formulaires standards pour ce rapport. Les producteurs d'examen peuvent décider de compléter ce bref rapport schématique par d'autres informations qui montrent la qualité de l'examen.

2. Rapport plus détaillé pour les praticiens de l'enseignement et de l'évaluation

Ce rapport va contenir plus d'informations ; il inclut en conséquence :

- 1) Des schémas sur
 - La relation de l'examen et des niveaux du CECR (Fiches A8 et A23 : Chapitre 4)
 - La corrélation entre le test et le CECR (Graphique de corrélation : Chapitre 6)
 - La qualité des notes seuils (Tableau de décision : Chapitre 6)
 - 2) Un exposé des justifications théoriques sur lesquelles le test est fondé, les objectifs du test, les utilisateurs visés, etc. On fournira en annexe les Fiches A1 à A7.
 - 3) Une description des aspects spécifiques de l'examen que l'on déclare avoir mis en relation avec le CECR en termes de **Spécification** (contenu). On fournira en annexe les copies des fiches appropriées du Chapitre 4 (Fiches A9 à A22) accompagnées des documents auxquels il est fait référence dans les fiches.
 - 4) Une description de ce que l'examen déclare avoir mis en relation avec le CECR en fonction de chacune des **Étapes de la Standardisation** par lesquelles on est passé. Pour des stages importants, on fournira en annexe des copies de la Fiche B1 du Chapitre 5, accompagnées de documents qui apportent la preuve du processus décrit.
 - 5) Une description de la façon dont l'examen prétend être en relation avec le CECR et à quel degré d'adéquation, en fonction des deux étapes de validation empirique (interne et externe) par lesquelles on est passé. Cela inclura les approches mises en œuvre, le nombre de sujets impliqués et les techniques utilisées. On fournira de nouveau en annexe les documents qui apportent la preuve du processus décrit.
 - 6) Une vue d'ensemble schématisée des étapes suivies dans la mise en relation avec le CECR, indiquant si l'utilisateur considère chaque point comme
 - Bien traité
 - Traitement convenable
 - A améliorer
- On trouvera à la fin de ce chapitre une fiche – élaborée en se fondant sur le panorama présenté au Chapitre 1.
- 7) Si possible, une présentation graphique de
 - Pouvoir discriminant de l'item
 - La fidélité des correcteurs individuellement et entre eux.

On encourage les producteurs d'examen à rédiger ces rapports :

- Pour accroître la transparence du contenu de l'examen (justification théorique, objectifs de l'examen, etc.)
- Pour accroître la transparence du niveau visé par l'examen
- Pour donner aux candidats, aux utilisateurs du test et aux praticiens de l'enseignement et de l'évaluation une occasion d'analyser la qualité d'un examen et de ses relations hypothétiques avec le CECR.
- Pour justifier pourquoi certaines procédures conseillées n'ont pas été suivies.

7.2. Démarches

La personne qui assure la coordination et rédige le rapport doit être formée à l'évaluation et avoir été étroitement associée au processus de mise en relation ou avoir facilement accès à toute la documentation recueillie.

On conseille de rédiger d'abord le rapport le plus détaillé et d'en extraire ensuite le rapport plus court destiné au grand public.

On trouvera ci-après une façon possible d'aborder la rédaction du rapport :

- Recueillir toute l'information nécessaire :
 - Information documentaire sur l'examen ;
 - Description de l'examen ;
 - Information sur le processus de familiarisation et documents d'accompagnement ;
 - Fiches remplies du processus de spécification ;
 - Toute l'information disponible sur les étapes du processus de standardisation ;
 - Toute l'information disponible sur les étapes du processus de validation empirique (y compris les données empiriques et/ou les rapports à ce sujet).
- Relater les événements pour chacune des quatre grandes parties du manuel. Décrire ce qui a été fait, quelles sont les recommandations qui ont été suivies et lesquelles ne l'ont pas été. Justifier pourquoi certaines des procédures recommandées n'ont pas été suivies, par exemple :
 - Manque de ressources, organisation réduite ;
 - On a établi un lien avec un cadre de référence national ou local (dans ces cas de « mise en relation indirecte » il faut apporter la preuve du lien entre le cadre de référence local et national et le CECR et du lien entre l'examen et le cadre de référence national ou local) ;
 - Opposition théorique entre la logique de l'examen et la logique du CECR.
- Fournir les documents appropriés, les rapports, l'information de façon bien structurée pour chacune des étapes et chacun des chapitres.
- Présenter graphiquement l'hypothèse d'une relation revendiquée avec le CECR faite sur la base d'une analyse de contenu en utilisant la Fiche A23.
- Présenter graphiquement l'hypothèse d'une relation revendiquée avec le CECR faite sur la base d'une analyse empirique de données en apportant :
 - Les données sur la discrimination des items (graphiques sur imprimantes)
 - La corrélation entre le test et le CECR (graphique de corrélation)
 - La fidélité individuelle et entre évaluateurs (corrélations/tableaux)
 - La qualité des notes seuils (tableau décisionnel)
- Présenter graphiquement le processus de mise en relation d'un examen avec le CECR en utilisant le tableau de la page suivante.

REPRESENTATION GRAPHIQUE DES PROCEDURES PERMETTANT DE RELIER LES EXAMENS AU CECR

"ARGUMENTER"

JUSTIFICATION THEORIQUE

- Veillez utiliser les colonnes pour indiquer jusqu'à quel point le processus a été complété :
1. Meilleure pratique conformément aux procédures présentées dans le Manuel
 2. Bonne pratique conformément aux procédures semblables à celles présentées dans le Manuel
 3. Abordé, mais peut être amélioré
- Ne pas compléter de colonne si le domaine n'a pas été abordé.

SPECIFICATION DU CONTENU DE L'EXAMEN		STANDARDISATION DES EVALUATIONS		VALIDATION EMPIRIQUE PAR L'ANALYSE DES DONNEES DU TEST	
1 2 3					
Validité interne : description et analyse		Familiarisation		Validation interne :	
° du contenu général de l'examen ;		Formation avec des échantillons normalisés de compétence de production		Analyse théorique classique des tests	
° du processus d'élaboration du test ;		Formation avec des échantillons normalisés de compétence linguistique et de réception		Méthodes d'analyse qualitative	
° de la correction, de la notation, des résultats ;				Théorie de la généralisabilité	
° de l'analyse du test et de la révision après passation		Etalonnage d'échantillons de performances locales		Analyse factorielle	
Validité externe : mettre en relation				Analyse théorique de réponse aux items	
° la description générale de l'examen avec l'échelle du CECR ;				Validation externe :	
° la description des activités de communication testées avec les échelles du CECR ;				Corrélation des résultats à des résultats de tests déjà étalonnés sur le CECR	
° la description des aspects de la compétence de communication langagière testée avec les échelles du CECR.		Définition de normes		Corrélation des corrections et des descripteurs du CECR	
		Diffusion et mise en oeuvre		Ancrage du test à un test déjà étalonné sur le CECR	
				Ancrage direct à la banque de descripteurs d'items qui sous tend le CECR	

AFFIRMATION de mise en relation avec le CECR (en se fondant sur la **spécification**)

AFFIRMATION (renforcée en se fondant sur la **spécification** et la **standardisation**)

AFFIRMATION (confirmation en se fondant sur la **validation empirique**)

Ces fiches sont également disponibles sur le site Internet www.coe.int/lang/fr

INTRODUCTION

Faire une brève introduction avec un court raisonnement théorique, l'exposé des buts généraux et des objectifs spécifiques de l'examen et sa relation au CECR. Inclure une brève vue d'ensemble des étapes suivies pour apporter la preuve de la qualité de la mise en relation avec le CECR.

DESCRIPTION GENERALE DE L'EXAMEN		
Information générale Intitulé de l'examen Langue testée Organisme d'évaluation Date de cette version Type d'examen But Population cible Nombre de candidats par an	<input type="checkbox"/> international <input type="checkbox"/> national <input type="checkbox"/> régional <input type="checkbox"/> institutionnel <input type="checkbox"/> 1 ^{er} cycle du sec. <input type="checkbox"/> 2 ^{ème} cycle du sec. <input type="checkbox"/> enseignant sup. <input type="checkbox"/> adultes	
Domaine(s) principal(aux)	<input type="checkbox"/> public <input type="checkbox"/> personnel <input type="checkbox"/> professionnel <input type="checkbox"/> éducatif	
Activités de communication testées	<input type="checkbox"/> 1 Compréhension de l'oral <input type="checkbox"/> 2 Compréhension de l'écrit <input type="checkbox"/> 3 Interaction orale <input type="checkbox"/> 4 Interaction écrite <input type="checkbox"/> 5 Production orale <input type="checkbox"/> 6 Production écrite <input type="checkbox"/> 7 Compétences intégrées <input type="checkbox"/> 9 Médiation orale d'un texte <input type="checkbox"/> 10 Médiation écrite d'un texte <input type="checkbox"/> 11 Langue (cad Grammaire, Vocabulaire, Cohérence) <input type="checkbox"/> 12 Autre (préciser) :	Intitulé de l'épreuve
Intitulé et durée des épreuves (se reporter à la colonne de droite ci-dessus)	Intitulé de l'épreuve 1 2 3 4 5 6	Durée de l'épreuve
Types de tâches utilisées	<input type="checkbox"/> choix multiple <input type="checkbox"/> vrai/faux <input type="checkbox"/> appariement <input type="checkbox"/> classer <input type="checkbox"/> phrase à trou(s) <input type="checkbox"/> phrase à finir <input type="checkbox"/> texte à trou fermé (réponse à choisir) <input type="checkbox"/> texte à trou (closure) <input type="checkbox"/> courte réponse à question ouverte <input type="checkbox"/> réponse élaborée (texte/monologue) <input type="checkbox"/> interaction avec l'examineur <input type="checkbox"/> interaction paritaire	Dans quelles épreuves (indiquez le n° ci-dessus) <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> etc. <input type="checkbox"/>

Information pour les candidats et les enseignants	<input type="checkbox"/> finalité <input type="checkbox"/> principaux domaines <input type="checkbox"/> sous tests du test <input type="checkbox"/> tâches <input type="checkbox"/> échantillons du test <input type="checkbox"/> vidéo d'un oral banalisé	<input type="checkbox"/> échantillons de réponses <input type="checkbox"/> critères de correction <input type="checkbox"/> barème <input type="checkbox"/> échantillons de performances standards montrant le niveau de réussite <input type="checkbox"/> échantillon de diplôme/certificat
Compte rendu	<input type="checkbox"/> note globale <input type="checkbox"/> note par sous test	<input type="checkbox"/> note globale plus profil visualisé <input type="checkbox"/> profil visualisé par sous test

NIVEAU DU CECR		DESCRIPTEUR GLOBAL DU CECR
UTILISATEUR EXPERIMENTE	C2	Est capable de comprendre facilement tout ce qu'il/elle lit ou entend. Peut résumer des informations de source orale et écrite différentes, retrouver un raisonnement et en rendre compte dans une présentation cohérente. Peut s'exprimer spontanément avec aisance et précision et discriminer des nuances fines de sens même en situation complexe.
	C1	Est capable de comprendre une large gamme de textes longs et exigeants et de reconnaître l'implicite du discours. Peut s'exprimer spontanément avec aisance sans d'évidentes recherches pour s'exprimer. Peut utiliser la langue de manière souple et efficace que le but soit social, académique ou professionnel. Peut produire des textes clairs, bien construits et détaillés sur des sujets complexes en manifestant un usage maîtrisé des modèles de construction, des connecteurs et des articulateurs
UTILISATEUR INDEPENDANT	B2	Est capable de comprendre les idées principales d'un texte complexe sur des sujets abstraits autant que concrets y compris des discussions techniques dans son domaine de spécialisation. Peut communiquer à un niveau d'aisance et de spontanéité qui rend possible une interaction normale avec un locuteur natif sans contrainte d'un côté comme de l'autre. Peut produire un texte clair et détaillé sur une large gamme de sujets et expliquer un point de vue sur une question en donnant les avantages et les inconvénients des différentes options.
	B1	Est capable de comprendre les points principaux d'une intervention claire standard sur des sujets familiers rencontrés régulièrement au travail, à l'école, lors des loisirs, etc. Peut se débrouiller dans la plupart des situations qu'on est susceptible de rencontrer dans une région où la langue est parlée. peut produire des textes simplement articulés sur des sujets courants ou d'intérêt personnel. Peut décrire des expériences et des événements, ses rêves, ses espoirs et ses ambitions et justifier et expliquer brièvement opinions et projets.
UTILISATEUR ELEMENTAIRE	A2	Est capable de comprendre des phrases et des expressions utilisées couramment relatives aux domaines les plus immédiatement significatifs (par exemple, information personnelle et familiale très élémentaire, faire des courses, la géographie locale, l'emploi). Peut communiquer lors de tâches courantes et simples demandant un échange simple et direct d'information sur des sujets courants et familiers. Peut décrire en termes simples des données sur lui/elle-même, l'environnement immédiat et des sujets dans des domaines relatifs aux besoins immédiats.
	A1	Est capable de comprendre et d'utiliser des expressions quotidiennes courantes et des expressions très élémentaires visant à répondre à la satisfaction de besoins concrets. Peut se présenter et présenter les autres et peut répondre à des questions personnelles et en poser telles que le domicile, les gens qu'on connaît et les choses qu'on a. Peut communiquer de façon simple à condition que les interlocuteurs parlent lentement et clairement et soient prêts à aider.

PROFIL DE LA RELATION AUX NIVEAUX DU CECR								
C2								
C1								
B2.2								
B2								
B1.2								
B1								
A2.2								
A2								
A1								
Panorama

VUE D'ENSEMBLE DES PROCEDURES POUR RELIER LES EXAMENS AUX NIVEAUX DU CECR

A : SPECIFICATION

Mise en relation avec le CECR par la spécification du contenu de l'examen

Familiarisation interne avec le CECR	- individuelle	<input type="checkbox"/>
	- entre collègues	<input type="checkbox"/>
Analyse sur la base de la description générale de l'examen et de l'échelle globale du CECR		<input type="checkbox"/>
Analyse sur la base d'une description d'activités de communication différentes testées (par exemple, compréhension de l'écrit, interaction de l'oral, etc.) et des échelles pertinentes du CECR		<input type="checkbox"/>
Analyse sur la base d'une description des aspects spécifiques des compétences de communication langagière testées (par exemple, étendue du vocabulaire, aisance à l'oral, etc.) et des échelles pertinentes du CECR		<input type="checkbox"/>

B : STANDARDISATION

Mise en relation avec le CECR par la formation à l'aide de tâches normalisées et d'échantillons de performance

Formation à la familiarisation		<input type="checkbox"/>
Formation à l'évaluation des performances en relation aux niveaux du CECR en utilisant des échantillons normalisés		<input type="checkbox"/>
Etalonnage des échantillons locaux de performances sur les niveaux du CECR		<input type="checkbox"/>
Formation à l'évaluation du niveau des items de tests du CECR en utilisant des échantillons normalisés		<input type="checkbox"/>
Etalonnage des échantillons locaux de tests sur les niveaux du CECR		<input type="checkbox"/>
Organisation de réunions régulières de standardisation		<input type="checkbox"/>

C : VALIDATION EMPIRIQUE

Mise en relation avec le CECR par la vérification empirique des normes définies en A et B

Analyse quantitative de la qualité du test		<input type="checkbox"/>
Analyse qualitative de la qualité du test		<input type="checkbox"/>
Validation sur un test utilisé comme critère externe		<input type="checkbox"/>
Validation sur un ensemble de jugements de l'évaluation d'une performance utilisée comme critère externe		<input type="checkbox"/>
Validation sur un ensemble d'évaluation d'enseignants utilisé comme critère externe		<input type="checkbox"/>