

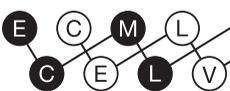


EMPOWERING LANGUAGE PROFESSIONALS
VALORISER LES PROFESSIONNELS EN LANGUES
SPRACHLEHRENDE IN IHRER ROLLE STÄRKEN



Evaluer par les tâches les langues à fins spécifiques à l'université: un guide

Johann Fischer, Catherine Chouissa, Stefania Dugovičová et Anu Virkkunen-Fullenwider



European Centre for Modern Languages
Centre européen pour les langues vivantes
Europäisches Fremdsprachenzentrum



COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

POUR L'EXCELLENCE DANS L'EDUCATION AUX LANGUES

Le Centre européen pour les langues vivantes (CELV) est un « Accord partiel élargi » du Conseil de l'Europe regroupant trente-quatre Etats membres¹. L'institution vise à rapprocher les théories issues des politiques linguistiques et les pratiques d'apprentissage en classe. A une époque de changement et de mobilité sans pareils, le Centre propose des approches concrètes permettant de répondre aux questions et aux défis posés aux sociétés culturellement diverses en Europe.

Le CELV cherche à opérer un changement positif pour les professionnels de l'éducation aux langues en :

- promouvant des approches innovantes ;
- développant la qualité de l'enseignement et de l'apprentissage des langues ;
- soutenant la mise en œuvre de politiques linguistiques éducatives ;
- encourageant le dialogue entre les praticiens et les décideurs dans le domaine de l'éducation aux langues.

Au sein du Conseil de l'Europe, les activités du CELV sont complémentaires à celles de la Division des politiques linguistiques, chargée de développer des politiques et des outils de planification en matière d'éducation aux langues, et à celles du Secrétariat de la Charte européenne des langues régionales ou minoritaires.

La présente série de publications est issue du programme d'activités 2008-2011 du CELV intitulé *Valoriser les professionnels en langues : Compétences – Réseaux – Impact – Qualité*. Le programme a été conduit dans un contexte marqué, dans le domaine de l'éducation, par des développements politiques majeurs sur la scène internationale, qui requièrent de la part des enseignants des qualifications professionnelles de plus en plus poussées. Le corps enseignant est censé contribuer aux processus de réforme de l'éducation nationale et répondre à un grand nombre de défis liés, entre autres, à l'application de normes dans l'enseignement, à l'évaluation axée sur les résultats, à une plus grande autonomie des institutions éducatives et à une hétérogénéité ethnique et culturelle croissante parmi les apprenants.

Les publications reflètent l'implication et l'engagement actif de tous ceux qui ont participé à une série de 24 projets internationaux, en particulier des équipes de coordination.

L'ensemble des publications du CELV et des matériels d'accompagnement est disponible sur le site internet <http://www.ecml.at/publications>.

¹ Les 34 Etats suivants sont membres de l'Accord partiel élargi du CELV : Albanie, Andorre, Arménie, Autriche, Bosnie-Herzégovine, Bulgarie, Croatie, Chypre, République tchèque, Estonie, Finlande, France, Allemagne, Grèce, Hongrie, Islande, Irlande, Lettonie, Liechtenstein, Lituanie, Luxembourg, Malte, Monténégro, Pays-Bas, Norvège, Pologne, Roumanie, République slovaque, Slovénie, Espagne, Suède, Suisse, « L'ex-République yougoslave de Macédoine », Royaume-Uni (situation au 30 juin 2011).

Evaluer par les tâches les langues à fins spécifiques à l'université: un guide

Johann Fischer, Catherine Chouissa,
Stefania Dugovičová et Anu Virkkunen-Fullenwider

avec la contribution spéciale d'Amelia Kreitzer Hope

Centre européen pour les langues vivantes

Editions du Conseil de l'Europe

Edition anglaise:

Guidelines for task-based university language testing

ISBN 978-92-871-7165-8

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

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

Couverture: Gross Werbeagentur, Graz

Mise en page: Stenner Medienproduktion, Graz

Editions du Conseil de l'Europe

F-67075 Strasbourg Cedex

<http://book.coe.int>

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

Nikolaiplatz 4

A-8020 Graz

Autriche

www.ecml.at

ISBN 978-92-871-7164-1

© Conseil de l'Europe, 2012

Imprimé en Autriche

Evaluer par les tâches les langues à fins spécifiques à l'université: un guide

Remerciements

Cette publication est le résultat d'une large collaboration. Elle associe le travail des membres de l'équipe de projet, les idées et les contributions des participants de notre atelier et de notre rencontre de réseau, de professeurs de langues, concepteurs de tests et examinateurs des établissements d'origine des membres de l'équipe de projet, et de plusieurs collègues et partenaires de différents réseaux et groupes professionnels.

Nous aimerions remercier tout d'abord le Comité Scientifique d'UNICert, qui a contribué à l'élaboration de la structure GULT pour des tests de niveau de capacité.

Nous sommes reconnaissants au Centre européen pour les langues vivantes du Conseil de l'Europe (CELV) qui nous a permis de travailler au projet GULT, et au personnel du CELV à Graz qui pendant nos réunions nous a permis de bénéficier de ses conseils d'experts et d'un environnement favorable.

Nous remercions tous les participants de notre atelier pour leurs commentaires et conseils constructifs.

Nous adressons des remerciements particuliers à tous les participants de la rencontre de réseau en janvier 2011 pour leur esprit de coopération, leurs apports critiques et leurs précieuses contributions: Mathilde Anquetil, Andromaqi Haloçi, Ann Elisabeth Laksfoss Hansen, Amelia Kreitzer Hope, Diana Ioniță, Marina Mattheoudakis, Marie-Thérèse Maurer, Cristina Pérez Guillot, Helena Šajgalíková, Joachim Schlabach, Carin Söderberg, Evan Stewart, Andrej Stopar et Rias van den Doel.

Nous aimerions également remercier plus spécialement les quatre responsables de groupes de notre rencontre de réseau GULT pour leurs commentaires avisés. Nous sommes particulièrement reconnaissants à Evan Stewart pour ses suggestions concernant la structure du guide et sa rédaction. Nous devons beaucoup à Amelia Kreitzer Hope pour son analyse critique de publications sur l'évaluation et pour ses apports théoriques; Amelia Hope a effectivement joué un rôle majeur dans la conception des chapitres 4, 7 et 8 du guide GULT, ce dont nous lui sommes extrêmement reconnaissants.

Pour finir, nous tenons à remercier notre consultante Isabel Landsiedler pour son soutien constant tout au long de ce projet.

1. Table des matières

Première partie: Guide GULT	9
1. Finalité et objectifs	11
2. Le contexte du projet GULT	12
3. Le public cible	14
4. Pourquoi évaluer les LSP par les tâches: la suite logique de l'enseignement par tâches en LSP	15
4.1. Evaluer par les tâches dans des disciplines diverses	15
4.2. L'approche actionnelle / par tâches dans l'évaluation en langues	16
4.3. L'évaluation par les tâches et l'approche communicative	20
4.4. Approches par tâches	24
5. Caractéristiques principales de l'évaluation par les tâches	26
5.1. Le construit (« construct »)	26
5.2. L'intégration des quatre compétences	27
5.3. La structure d'un test ou d'un examen par tâches	29
La structure GULT	29
Etape 1: Introduction à la thématique	30
Etape 2: Les compétences réceptives	30
Etape 3: Le travail d'étude de cas / d'analyse de cas	31
Etape 4: Les compétences productives	31
Etape 4a: Ecrire	32
Etape 4b: Parler	32
Durée	33
6. Qui peut tirer bénéfice de l'évaluation par les tâches et en quoi ?	34
7. Défis et limites	38
7.1. Réserves d'ordre psychométrique	38
7.2. Réserves d'ordre pratique	38
7.3. Autres réserves	40

8.	L'évaluation de la performance du candidat	41
9.	Usages possibles des tests / examens par tâches: comment exploiter les tests par tâches	44
9.1.	Tests de connaissances: examens de fin de cours	44
9.2.	Tests de niveau de capacité et « prochievement tests »	45
10.	Comment élaborer un test par tâches	47
10.1.	La conception du test	47
10.1.1.	La problématique	47
10.1.2.	L'authenticité de la tâche	47
10.1.3.	Le dossier	47
10.1.4.	La situation	48
10.1.5.	La tâche	48
10.1.6.	Les tâches intermédiaires pour les différentes compétences	48
10.2.	Etapas suivantes dans l'élaboration du test: essais	49
11.	Bibliographie	50
	Deuxième partie: Ressources documentaires	57
	Document 1: Tests et évaluation	59
	Document 2: Exemples de mise en œuvre de l'évaluation par les tâches dans des systèmes universitaires de certification du niveau de capacité: mise en situation réelle	66
	Document 3: Exemples de tests et d'examens par tâches	72
	Document 4: Exemples de mise en œuvre de l'évaluation par les tâches dans l'enseignement des langues à l'université	75
	Document 5: Grilles d'évaluation	77
	Document 6: Définitions de la « tâche »	99
	Document 7: Définitions de l'« authenticité »	104
	Document 8: Etapes de la conception d'un test par tâches	110
	Document 9: Glossaire	116



La présente publication résulte du projet du Centre européen pour les langues vivantes intitulé « Evaluer par les tâches les langues à fins spécifiques à l'université » (GULT).

De plus amples informations et des matériaux complémentaires sont disponibles sur le site internet <http://gult.ecml.at>.

Première partie: Guide GULT

1. Finalité et objectifs

Avec l'intégration des projets, des études de cas, des simulations globales et des activités de WebQuest, l'approche actionnelle se généralise dans l'enseignement des langues. Trop souvent néanmoins, les examens de fin de cours et les tests de niveau de capacité mesurent encore le niveau atteint par les étudiants en utilisant des tests à items discrets qui ne correspondent pas à ce qui a été fait en cours. C'est pourquoi le projet GULT (Guidelines for Task-Based University Language Testing¹) s'est fixé pour but d'élaborer une structure permettant d'adopter une approche par tâches pour l'évaluation en langues. Il s'agit de mettre en cohérence l'enseignement des langues et l'évaluation et d'encourager une approche actionnelle dans les deux cas, afin de rendre les tests de langues plus pertinents pour les personnes testées et d'évaluer jusqu'à quel point celles-ci sont capables de communiquer de manière appropriée dans une situation de communication particulière à leur domaine d'études ou à leur environnement professionnel futur.

Ce guide concerne l'évaluation en langues à fins spécifiques (LSP: Languages for Specific Purposes) à l'université aux niveaux B2 et C1 du *Cadre européen commun de référence pour les langues*² car, dans la plupart des pays, les LSP sont enseignées à ces niveaux. Néanmoins le modèle décrit ci-dessous peut aisément s'appliquer au niveau C2 et, en adaptant le déroulement, au niveau B1. Le projet GULT cible l'enseignement et l'évaluation des LSP, mais la démarche pourrait être utilisée pour évaluer la langue sur objectifs universitaires (LAP: Languages for Academic Purposes) et appliquée à l'évaluation en langue générale.

Ce guide a pour objet l'évaluation par les tâches. C'est pourquoi les aspects généraux de l'évaluation en langues ne sont pas traités dans le texte principal, mais la ressource documentaire 1 dans la seconde partie de la publication fournit une introduction à l'évaluation en langues en général.

En complément du guide, le site web du projet (<http://gult.ecml.at>) propose des modèles de tests pour les niveaux B2 et C1 en espagnol, français, allemand et anglais.

1 Titre original: « Guidelines for University Language Testing », changé en « Guidelines for Task-Based University Language Testing » afin de le rendre plus clair.

2 Conseil de l'Europe, 2001.

2. Le contexte du projet GULT

L'approche actionnelle a déjà été introduite dans les programmes d'enseignement de nombreux centres de langues universitaires en Europe, mais malheureusement pas encore dans l'évaluation. Aussi l'équipe de projet GULT a-t-elle rassemblé les compétences et les résultats de projets antérieurs, à savoir

- les activités du groupe HERMES³ sur la formation professionnelle pour les enseignants des universités et collèges d'agriculture et de sciences connexes dans le projet AFANet⁴;
- les résultats du projet WebCase⁵ sur l'utilisation des études de cas dans les programmes d'enseignement des agro-industries à l'université;
- le projet LCaS⁶ qui avait pour objet la mise en place de modules de formation des enseignants à une utilisation plus développée d'études de cas pour l'enseignement des langues à l'université;
- le projet EXPLICS⁷ avec pour but la conception d'études de cas en ligne et de simulations globales en onze langues;
- les compétences de l'équipe allemande d'UNICert®⁸ avec son expertise en évaluation en langues et en LSP;
- l'équipe UNICert@LUCE⁹ en République slovaque et en République tchèque;
- l'équipe française du CLES¹⁰ avec son savoir-faire dans l'évaluation actionnelle en langues à l'université; et
- le réseau finlandais des centres de langues universitaires avec son expérience de l'évaluation en langues à fins spécifiques.

Il a ainsi réuni un certain nombre de spécialistes de l'enseignement des langues conçu dans une approche actionnelle à l'université, de l'enseignement des langues à fins

3 Voir site HERMES: www.uni-tuebingen.de/hermes/; Casey et Fischer, 2004, 2005; Fischer, 2003, 2004; Fischer et Casey, 2005.

4 Voir site web AFANet: www.afanet.info.

5 Voir site web WebCase: www.webcase-online.info; Cain et Heath, 2004.

6 Voir site web LCaS: www.ecml.at/mtp2/LCaS; Fischer *et al.*, 2008.

7 Voir site web EXPLICS: www.zess.uni-goettingen.de/explics; Fischer *et al.*, 2009.

8 Voir site web UNICert®: www.unicert-online.org; Eggensperger et Fischer, 1998; Voss, 2010.

9 Voir site web UNICert@LUCE: www.casajc.sk/unicert.htm.

10 Voir site web CLES: www.certification-cles.fr/; Petermann, 2008.

spécifiques et de l'évaluation en langues dans le but de créer un nouveau modèle pour l'évaluation en langues à l'université, fondé sur les principes de l'enseignement et de l'apprentissage par les tâches et adapté aux besoins particuliers des apprenants en langues à fins spécifiques.

3. Le public cible

Ce guide a été conçu

- pour fournir aux responsables de l'enseignement des langues à l'université – directeurs de centres de langues ou de départements de langues universitaires par exemple – un cadre d'évaluation actionnelle en langues et l'outil nécessaire pour mettre en place l'approche GULT dans leurs établissements;
- pour aider les enseignants de langues sur objectifs spécifiques dans l'enseignement supérieur à concevoir et à administrer des tests de langues basés sur des tâches pour évaluer les capacités langagières de leurs étudiants;
- pour promouvoir une approche actionnelle dans l'enseignement et l'évaluation au niveau universitaire en informant les partenaires – à savoir les associations professionnelles, les réseaux régionaux, nationaux et internationaux, les administrations et ministères de l'Education nationaux et régionaux, les organismes de tests universitaires nationaux et internationaux – sur l'enseignement, l'apprentissage et l'évaluation par les tâches.

4. Pourquoi évaluer les LSP par les tâches: la suite logique de l'enseignement par tâches en LSP

L'enseignement des langues à l'université intégrant de plus en plus les modèles actionnels, l'élaboration d'évaluations par les tâches en est la suite logique.

4.1. Evaluer par les tâches dans des disciplines diverses

L'équipe GULT ne prétend pas que l'évaluation par les tâches (TBA: Task-Based Assessment) offre de meilleurs tests, mais que ces tests sont plus proches de la vraie vie, et qu'ils préparent donc les étudiants à leur vie future. Ce lien avec des activités proches de la vraie vie est certainement l'une des raisons expliquant que plusieurs disciplines universitaires, telles que la médecine, le droit, les sciences sociales ou les sciences de la vie, ont introduit l'approche actionnelle à la fois dans l'enseignement et dans l'évaluation.

A l'origine, la priorité de l'équipe GULT n'était pas l'évaluation de la langue générale, pour laquelle quelques modèles d'exams basés sur des tâches ont déjà été mis en pratique, mais l'évaluation des savoir-faire en langues à fins spécifiques, puisque l'évaluation des LSP est une des activités majeures des centres de langues universitaires en Europe. Les modèles GULT peuvent, bien entendu, être adaptés et utilisés pour l'évaluation de la langue générale.

Si l'on utilise une approche par tâches pour concevoir un test de LSP dans une situation réaliste spécifique basée sur des faits et sur des documents authentiques¹¹, les candidats peuvent montrer jusqu'à quel point ils sont capables de communiquer dans une langue étrangère, ce qui rend les résultats plus utiles et plus pertinents pour eux. Ce type de test est également plus transparent pour les candidats, car ils peuvent plus facilement juger par eux-mêmes comment ils ont réussi à gérer linguistiquement une situation de communication réaliste. C'est pourquoi la finalité du projet GULT est de concevoir un guide et des modèles pour une évaluation transparente des LSP dans une approche par tâches à l'université.

11 C'est-à-dire une situation qui correspond à une situation authentique dans la vie réelle et des documents et des faits qui n'ont pas été conçus à des fins d'enseignement.

4.2. L'approche actionnelle / par tâches dans l'évaluation en langues

Définition de la « tâche »¹²

Dans une évaluation « par les tâches » ou « actionnelle », parfois appelée aussi « évaluation de la performance »¹³, l'examineur donne au candidat une tâche à accomplir, tâche qui pousse le candidat à agir. Par « tâche » nous entendons:

A task is an activity which requires learners to use language, with emphasis on meaning, to attain an objective.

(Skehan, 2003: 3)

Traduction française:

Une tâche est une activité qui exige des apprenants qu'ils utilisent la langue, en privilégiant le sens, pour atteindre un objectif.

Tasks [...] are activities which have meaning as their primary focus. Success in tasks is evaluated in terms of achievement of an outcome, and tasks generally bear some resemblance to real-life language use. So task-based instruction takes a fairly strong view of communicative language teaching.

(Skehan, 1996: 20)

Traduction française:

Les tâches [...] sont des activités qui ont le sens pour objet premier. La réussite est évaluée en termes de réalisation d'une production, et les tâches sont en général proches de l'usage de la langue dans la vraie vie. Ainsi l'enseignement par les tâches s'appuie assez fortement sur l'approche communicative dans l'enseignement des langues.

Dans une approche par tâches, une macro-tâche décrit l'activité générale qui doit être effectuée, un résultat bien défini étant attendu à la fin. Afin d'atteindre ce résultat, les candidats doivent élaborer des productions prédéfinies, à savoir un rapport écrit expliquant le problème à résoudre et ses solutions possibles, et une présentation orale de ces solutions à des interlocuteurs donnés dans un contexte donné. Toutes ces activités et productions sont liées entre elles. Pour parvenir à ces résultats, les candidats doivent souvent réaliser des tâches intermédiaires, ou micro-tâches, précises.

12 Voir aussi la ressource documentaire 6 « Définition de la tâche ».

13 Voir par exemple Bachmann, 2007; Wigglesworth, 2008.

Terminologie

Dans la littérature, on trouve deux termes: « task-based » et « action-based » en anglais, ou en français « approche par tâches » et « approche actionnelle », et en allemand « aufgabenbasiert » et « handlungsorientiert ». Ces termes sont fortement reliés entre eux, mettant l'accent dans un cas davantage sur l'enseignant qui donne une tâche au candidat et, dans l'autre cas, sur la personne testée qui doit agir. Dans le contexte anglophone, on utilise certainement plus souvent le terme « task-based », tandis qu'en français c'est l'expression « approche actionnelle » qui prévaut. En allemand l'expression « Aufgabenorientierung » est passablement ambiguë, dans la mesure où « Aufgabe » est utilisé pour tout, du simple exercice à items isolés aux devoirs à la maison et pour une « tâche » plus complexe. C'est pourquoi en allemand il est préférable d'utiliser le terme « Handlungsorientierung » bien que le terme « Aufgabenorientierung » ait été récemment davantage employé.¹⁴

De fait, la tâche que l'examineur donne aux candidats incite ceux-ci à agir. La relation entre les deux termes « task-based » et « action-based » peut donc être illustrée de la manière suivante:

task / tâche / Aufgabe	→	action / action / Handlung
task-based entrée par les tâches aufgabenorientiert	→	action-based perspective actionnelle / approche actionnelle handlungsorientiert
↑ Input		↑ Output
teacher's view perspective de l'enseignant Perspektive der Lehrkraft	→	learner's view perspective de l'apprenant Perspektive des Lerners

14 Voir par exemple Bausch *et al.*, 2006.

Le concept GULT de l'évaluation par les tâches

Pour l'équipe GULT, « évaluer par les tâches » signifie évaluer la compétence langagière sur la base de tâches réalistes: les candidats sont mis dans une situation à laquelle ils pourraient être confrontés dans la vraie vie et ils doivent travailler sur une macro-tâche, en général un problème ou un projet. Des tâches intermédiaires permettent de tester les différents savoir-faire langagiers (de manière séparée ou holistique).

Dans cette approche, la tâche est utilisée en tant que déclencheur ou *input*, stimulant le candidat à interagir avec son ou sa partenaire et / ou l'examinateur, tandis que le produit ou *output* (la production du candidat) est évalué à différents niveaux en tenant compte de la compétence linguistique, mais aussi de la compétence pragmatique et du contenu. Pour évaluer les compétences productives, on a besoin de critères et de grilles transparents et clairement définis.¹⁵

Il est important de se souvenir que, dans le cadre GULT, l'enseignement, l'apprentissage et l'évaluation se situent dans un contexte universitaire, contexte dans lequel toutes les compétences langagières sont habituellement requises et où, du fait de l'utilisation de langues à fins spécifiques, le contenu revêt une importance particulière. Si des tests de langues sans lien avec un programme d'enseignement et d'apprentissage se doivent de restreindre les aspects de contenu pour garantir des chances égales à tous les candidats, le contenu joue par contre un rôle essentiel dans ce contexte.

L'approche GULT tente de combiner l'expérience de différentes approches de l'évaluation en langues en définissant un contexte particulier et en faisant reposer le test sur une macro-tâche, requérant des candidats qu'ils mènent des échanges riches de sens sur une question spécifique où le contenu joue un rôle majeur, et en évaluant les savoir-faire dans les différentes activités langagières.

Des recherches ultérieures seront nécessaires afin d'étudier si l'approche GULT a réussi à élaborer une structure rendant plus valides les tests de langues à l'université, et si cette approche est parvenue à associer les avantages de différentes approches de l'évaluation en langues, ainsi que Bachman l'a suggéré (2007: 70f):

Issues related to language ability and language use contexts and the interaction between these have been addressed, in a dialectic, in language assessment research, and have led to three general approaches to defining the construct, or what we want to assess: (1) ability-focused, (2) task-focused, and (3) interaction-focused. While the different theoretical perspectives that underlie these approaches are not mutually exclusive, they are based on different sets of values and assumptions. [...] Because of these differences, the conundrum of ability and context and how they interact in language use and language

15 Voir chapitre 8 et ressource documentaire 5.

assessment is, in my view, essentially a straw issue, theoretically, and may not be resolvable at that level.

Nevertheless, the theoretical issues raised by these different approaches have important implications and present challenging questions for both empirical research in language testing and for practical test design, development, and use. These theoretical issues also provide valuable insights into how we can enrich the ways in which we conceptualize what we assess and how we go about assessing it. For research, they imply the need for a much broader, more catholic methodological approach, involving both so-called quantitative and qualitative perspectives and methodologies. For practice, they imply that exclusive focus on any one of these approaches (ability, task, interaction), to the exclusion of the others, will lead to potential weaknesses in the assessment itself, or to limitations on the uses for which the assessment is appropriate. This means that we need to address all three in the design, development, and use of language assessments.

Traduction française:

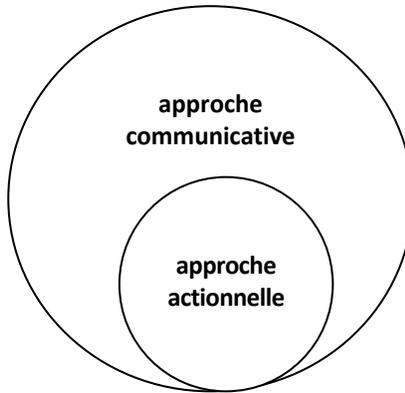
Les questions portant sur la capacité langagière et les contextes d'utilisation de la langue, et leur interaction ont été abordées, dans une démarche dialectique, dans la recherche sur l'évaluation en langues, et ont conduit à trois approches générales pour définir le construit, à savoir ce que nous voulons évaluer: (1) centrée sur la capacité, (2) centrée sur la tâche, et (3) centrée sur l'interaction. Si les différentes perspectives théoriques qui sous-tendent ces approches ne s'excluent pas mutuellement, elles reposent néanmoins sur des valeurs et des hypothèses différentes. [...] A cause de ces différences, l'énigme de la capacité et du contexte, et de leur mode d'interaction dans l'usage de la langue et l'évaluation en langue est, à mon avis, pour l'essentiel une question secondaire, sur le plan théorique, et elle pourrait bien rester sans réponse à ce niveau.

Néanmoins, les questions théoriques soulevées par ces différentes approches ont des conséquences importantes et posent des questions difficiles à la fois pour la recherche empirique sur l'évaluation en langues et pour la conception, l'élaboration et l'utilisation de tests dans la pratique. Ces questions théoriques fournissent également des perspectives très utiles sur la façon dont nous pourrions enrichir nos modes de conceptualisation de ce que nous évaluons et nos manières de l'évaluer. Pour la recherche, elles impliquent qu'il est nécessaire d'avoir une approche méthodologique bien plus large, plus éclectique, utilisant à la fois les perspectives et les méthodologies dites quantitatives et qualitatives. Pour la pratique, elles signifient que cibler uniquement une de ces approches (capacité, tâche, interaction) à l'exclusion des autres conduira à des faiblesses potentielles dans l'évaluation elle-même ou à une limitation des usages auxquels l'évaluation est adaptée. Cela signifie que nous devons prendre en compte les trois dans la conception, l'élaboration et l'utilisation des évaluations en langues.

4.3. L'évaluation par les tâches et l'approche communicative

L'apprentissage et l'évaluation par les tâches peuvent être considérés comme une évolution plus ciblée et plus élaborée de l'approche communicative dans l'enseignement des langues. Les fondements de l'approche communicative impliquent une focalisation sur la transmission du sens, l'utilisation de ressources authentiques, l'intégration de diverses compétences et la centration de l'enseignement sur les étudiants et leurs besoins de communication. Ce que l'approche actionnelle ajoute, c'est l'accent mis sur l'intégration d'actes communicatifs globaux dans un contexte et une situation particuliers avec un objectif spécifique qui reflète les objectifs de communication présents ou futurs de l'apprenant.

La relation entre l'approche communicative et l'approche actionnelle peut être illustrée de la manière suivante:



Cette relation dépend toutefois de la définition de « l'approche communicative », que des chercheurs français, par exemple, conçoivent différemment: Bourguignon (2006: 58) parle d'un passage « de 'l'ère communicative' [...] à 'l'ère communic-active' autour d'une approche qui ne soit plus seulement 'communicative' mais 'communic-actionnelle' ». Bourguignon poursuit:

On se rend compte qu'avec la perspective actionnelle, nous changeons de paradigme. Nous passons du paradigme de la connaissance, de la simplification à celui de la compétence, de la complexité qui ne veut pas disjoindre l'objet et le sujet, la réflexion et l'action, l'apprenant et l'usager mais les conjointre pour que la finalité de l'apprentissage d'une langue ne soit pas la bonne note mais l'utilisation autonome de la langue dans des situations plus ou moins complexes qui peuvent aller de la lecture d'un prospectus à celle d'une œuvre de Shakespeare! [...]

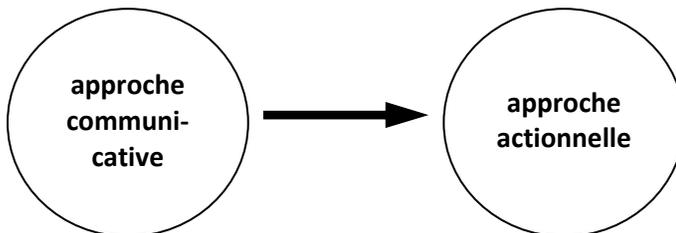
Au niveau de la perspective actionnelle, la communication est au service de l'action. Il s'agit de passer de l'interaction (= parler avec les autres) à ce que C. Puren appelle la « co-action », et que nous avons choisi d'appeler la « communic-action », c'est-à-dire agir avec les autres, communiquer pour agir. [...]

Avant l'approche communicative, l'enseignement / apprentissage de la langue était totalement coupé de son utilisation sociale. Le savoir de référence était le « savoir savant », la linguistique et la langue s'apprenaient à partir d'exercices de grammaire et de vocabulaire. Avec l'approche communicative, on fait « rentrer » le milieu social dans le milieu scolaire. Il s'agit de simuler des situations de communications « empruntées » au milieu social. Se mettent en place des « tâches communicatives » qui visent l'efficacité dans la transmission de l'information. Pour autant, ces tâches communicatives restent des tâches d'apprentissage qui seront évaluées en tant que telles. Il y a donc bien disjonction entre tâche d'apprentissage et tâche sociale.

Avec la perspective actionnelle, tâche d'apprentissage et tâche sociale sont liées. Il ne s'agit plus en classe de simuler des situations d'usage en donnant à l'apprenant le rôle de l'usager. L'apprenant sera acteur d'un projet qui ne visera pas seulement la réutilisation ou l'application de connaissances mais aussi sa construction en fonction des objectifs à atteindre et des aléas de la situation.

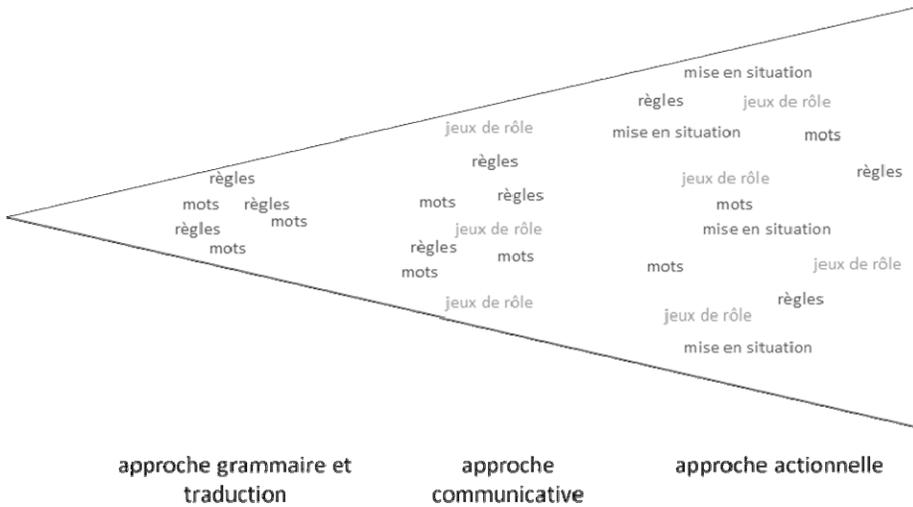
(Bourguignon, 2006: 63-65)

Ceci signifie que la relation entre l'approche communicative et l'approche actionnelle peut être considérée selon la représentation suivante¹⁶:



En conclusion, selon l'équipe GULT la relation entre les différentes approches de l'enseignement et de l'évaluation en langues pourrait être illustrée comme suit, même si cette illustration est très simplifiée:

16 Voir aussi Puren 2006a, où il explique sa distinction entre l'approche communicative et l'approche actionnelle.

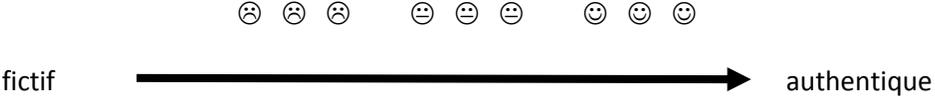


Les activités utilisées dans l’enseignement et l’évaluation par les tâches devraient être aussi proches que possible des tâches de communication auxquelles les apprenants sont ou seront confrontés à l’extérieur de la salle de cours de langue. A titre d’exemple de la différence entre des tests conçus dans une approche communicative et des tests conçus dans une approche actionnelle, une épreuve orale de type communicatif pourrait inclure de comparer deux images ou de donner son opinion sur le végétarisme. Un test par tâches n’utiliserait ces formats que dans la situation peu probable où des descriptions visuelles détaillées ou le fait de défendre son point de vue sur le végétarisme seraient des contextes de communication que les candidats rencontreraient dans leur vie professionnelle ou universitaire hors de la salle de classe.

Pour expliciter plus précisément la différence entre des tests conçus dans une approche communicative et des tests conçus dans une approche actionnelle, considérons l’omniprésent jeu de rôles, pilier de bien des séquences communicatives. Dans un test de type communicatif, on pourra demander aux candidats de jouer les rôles du médecin et du patient et d’improviser un dialogue sur une maladie. Par contre, dans un test par tâches, on n’exigerait jamais des candidats qu’ils fassent semblant d’être des médecins, sauf s’il s’agissait d’étudiants de médecine qui auraient besoin dans l’avenir de communiquer avec des patients dans leur langue seconde. On ne demanderait pas à des étudiants de commerce de tenir le rôle d’un PDG, mais celui d’étudiants en stage, car être directeur général requiert des compétences accumulées au fil de nombreuses années d’activité dans le domaine des affaires.

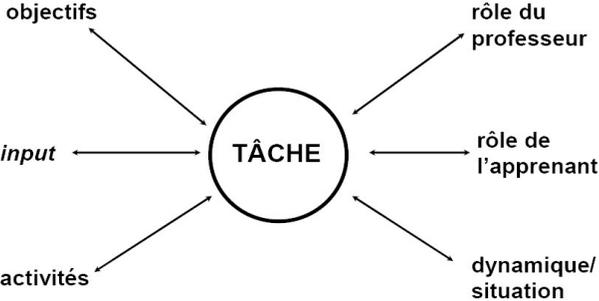
Il faut noter qu’un test de langue ne sera jamais tout à fait authentique, à moins de tester les candidats dans la vraie vie dans un contexte de langue cible (évaluation

directe). Néanmoins, un test conçu dans une approche actionnelle devrait tenter de simuler aussi fidèlement que possible une situation authentique tout en reconnaissant que le scénario se situera quelque part sur un continuum entre le fictif et l'authentique. Plus le cadre est authentique, mieux c'est:



On ne jugera pas seulement si les candidats ont utilisé une formulation linguistiquement correcte, mais en premier lieu s'ils ont réussi à réaliser la tâche, par exemple s'ils ont trouvé l'endroit où ils voulaient se rendre (niveau A) ou si les auditeurs ont compris le contenu de la présentation et si le contenu a eu un impact sur les activités du groupe (niveau C).

Dans le schéma suivant qui s'appuie sur les travaux de Nunan (Nunan, 1989), Puren (2008: 12) place la tâche au centre du contexte d'apprentissage, car elle est liée aux objectifs, à l'*input* et aux activités d'une part, et au rôle de l'enseignant, au rôle de l'apprenant et à la dynamique ou à la situation d'autre part:



La tâche est donnée et exécutée avec des objectifs bien définis qui nécessitent un *input* particulier, et des activités spécifiques doivent être réalisées. Bien entendu, le rôle de l'enseignant et le rôle de l'apprenant doivent tous deux être revus, car ils sont très différents de situations plus traditionnelles dans la salle de classe.

La citation suivante de Puren (2008: 4) montre l'importance que l'approche actionnelle accorde à la nature holistique des tâches:

La mise en œuvre de la perspective actionnelle suppose logiquement un passage à une « entrée par l'action », l'unité affichée de l'unité didactique (son titre) étant désormais celle d'une action unique.

Puren poursuit (2008: 12):

[...] l'« entrée par l'action », c'est-à-dire un modèle d'unité didactique où toutes les activités dans tous les domaines (CO, CE, PO, PE, lexique, culture, grammaire et graphie-phonie) sont conçues en fonction d'une action unique à partir de laquelle et à propos de laquelle est construite l'unité de l'unité didactique.

Cette conception globale de la tâche signifie que toutes les tâches intermédiaires sont indissociables de la macro-tâche et de l'objectif final. Dans une unité d'enseignement ou un examen de langue conçu selon une approche par tâches, les tâches prises isolément ne sont pas indépendantes les unes des autres; pour réussir à les réaliser, il faut en priorité avoir une vue d'ensemble du projet ou du problème en question. Ainsi les différentes activités langagières sont plus fortement reliées entre elles dans une approche par tâches, où il est nécessaire de comprendre les documents écrits et d'écouter pour agir dans la phase de production de l'épreuve, que dans des tests à items discrets classiques.

4.4. Approches par tâches

L'enseignement des langues dans une approche par tâches peut utiliser les études de cas¹⁷, les projets¹⁸ et les simulations globales¹⁹. Alors que les simulations globales et les projets sont plus créatifs et ouverts et mettent l'accent sur les compétences de production, les études de cas partent d'un problème à résoudre et commencent par une phase de réception importante, qui est essentielle pour mener à bien l'activité; sans une analyse détaillée du problème, l'apprenant ou le candidat ne sera pas à même de concevoir une solution.²⁰

17 Voir aussi Almagro Esteban et Pérez Cañado, 2004; Daly, 2002; Kaiser, 1983; Kiefer, 2004; Uber Grosse, 1988.

18 Voir aussi Ribé et Vidal, 1993; Schart, 2003.

19 Voir aussi Caré et Debyser, 1995; Debyser, 1996; Yaiche, 1996.

20 Voir aussi Fischer *et al.*, 2008; Fischer *et al.*, 2009.

On peut définir ainsi les études de cas:

Darstellung einer konkreten Situation aus der betrieblichen Praxis oder dem Alltagsleben, die anhand bestimmter Tatsachen, Ansichten und Meinungen dargestellt wird, auf deren Grundlagen eine Entscheidung getroffen werden muss.

(Kaiser, 1983: 20)

Traduction française:

Présentation d'une situation concrète tirée de la réalité de l'entreprise ou de la vie quotidienne, qui est représentée à travers certains faits, positions et opinions sur la base desquels une décision doit être prise.

L'équipe LCaS a défini les études de cas de la manière suivante:

analyse d'un problème / dilemme dans une situation donnée pour lequel il n'existe pas qu'une seule solution

(Equipe de projet LCaS)

Les « simulations globales » sont décrites ainsi par Debyser (1996: IV):

Une simulation globale est un protocole ou un scénario cadre qui permet à un groupe d'apprenants pouvant aller jusqu'à une classe entière d'une trentaine d'élèves, de créer un univers de référence – un immeuble, un village, une île, un cirque, un hôtel – de l'animer de personnages en interaction et d'y simuler toutes les fonctions du langage que ce cadre, qui est à la fois un lieu-thème et un univers du discours, est susceptible de requérir.

Dans l'évaluation en langues, les études de cas et les projets conviennent pour des tests de niveau de capacité et pour des examens de fin de formation. Si un cours repose sur une simulation globale, un projet spécifique intégré dans cette simulation peut aussi servir d'examen de fin de cours. Certains domaines de spécialité tels que l'économie, les affaires, le droit, la médecine et les sciences sociales se prêtent davantage aux études de cas, alors que les projets et les simulations sont plus appropriés dans des domaines tels que les sciences.

5. Caractéristiques principales de l'évaluation par les tâches

5.1. Le construit (« construct »)

Le construit général de l'examen GULT suit une perspective interactionniste²¹, c'est-à-dire que l'examen cible l'interaction dans un contexte de communication. Généralement, l'examen n'adopte pas une approche intégrative où les quatre activités langagières sont testées de manière intégrée²²; cependant les différentes tâches évaluant ces quatre 'compétences' (qui peuvent dans certains cas être combinées) s'insèrent toutes dans un contexte unique et sont construites autour d'une macro-tâche. Les tâches, qui devraient être aussi proches que possible d'activités de la vraie vie, servent ici d'input, mais la performance langagière du candidat est évaluée en fonction de critères portant sur le contenu, de critères pragmatiques, (socio-)linguistiques et paralinguistiques, comme indiqué dans les grilles d'évaluation consultables dans la ressource documentaire 5. Ainsi, l'examen GULT veut évaluer la capacité linguistique du candidat dans un contexte donné.

La question de savoir si l'on a à faire à un construit fort ou faible²³ dépend de chaque test et des spécifications élaborées par l'institution concernée. En général, il n'est ni fort ni faible, car l'accent est mis à la fois sur la réalisation de la tâche et sur la langue utilisée.

Pour résumer, le construit d'un test GULT peut être ainsi défini:

- Le test GULT est un test par tâches.
- Le test vise à montrer comment les apprenants peuvent s'exprimer dans une situation de la vraie vie dans le cadre de leurs études ou de leur futur lieu de travail, en utilisant (lorsque nécessaire) la langue adaptée au domaine de spécialité en question, même si les résultats ne permettront pas nécessairement de prédire la performance langagière du candidat dans une situation future de communication sur le lieu de travail.
- Le test GULT prend en considération et évalue les critères de contenu, de compétence pragmatique, de compétence linguistique et paralinguistique. Leur

21 Voir Bachman, 2007.

22 Voir section 5.2.

23 Voir Wigglesworth, 2008: 113; voir également la distinction chez Bachman (2002) entre approche « par le construct » et approche « par les tâches ».

part variera en fonction du/des savoir-faire testé(s), du niveau et du cadre institutionnel du test, et des grilles d'évaluation utilisées.

- Le test GULT est un test critérié et non un test normatif. Les concepteurs du test ou l'institution fixent les critères pour chaque test et ceux-ci devraient être alignés sur les critères / catégories du CECRL.

Etant donné que l'approche GULT fournit un cadre pour l'évaluation et non des spécifications concrètes pour un examen ou test précis, chaque institution doit spécifier ce cadre pour chaque test, c'est-à-dire que chaque établissement ou organisme de tests doit définir son propre construit.

5.2. L'intégration des quatre compétences

Les tests de langues évaluent généralement les quatre (ou cinq²⁴) compétences de manière successive et séparée. Ces tests couvrent jusqu'à six différents domaines: la lecture, l'écoute, la production orale, l'interaction orale, la production écrite et l'usage pertinent de la langue (c'est-à-dire la syntaxe et le lexique des LSP). Chaque partie du test évalue normalement une et une seule compétence; dans certains cas, elle en combine deux (ou, rarement, plus), mais le plus souvent les parties évaluant les différentes activités langagières ne sont pas reliées entre elles à l'intérieur du test et abordent des thèmes différents. Il peut y avoir une exception pour l'évaluation de la syntaxe et du lexique des LSP, qui sont souvent intégrés dans l'évaluation des compétences de production.

Tester séparément chaque compétence permet à l'examineur de se faire une idée précise des points forts et des points faibles d'un apprenant dans chaque compétence. De plus, si l'on teste chaque compétence séparément avec un sujet différent ou nouveau pour chaque partie du test, on ne pénalisera pas les candidats qui sont moins familiers de tel ou tel sujet. Dans la vraie vie, ces compétences se présentent rarement séparément et sont normalement reliées entre elles. Comme dans l'approche communicative, il est possible de déterminer si les candidats sont capables de comprendre un texte écrit ou de faire une présentation et d'en discuter avec le groupe. Mais il n'est pas possible de dire s'ils peuvent restituer les informations clés et les idées

24 On utilisera dans ce qui suit la distinction traditionnelle entre quatre différentes compétences, car le CECRL fait une distinction entre la production orale et l'interaction orale pour la compétence « parler », mais pas pour la compétence « écrire ». Il est également évident qu'il y a de grandes différences entre écrire un texte unique, isolé ou un texte qui s'intègre dans une interaction plus longue entre deux ou plusieurs partenaires de communication nécessitant des compétences de lecture (par exemple un clavardage, une plateforme de discussion, un forum ou même un échange de courriels ou de lettres).

sous-jacentes d'une série de documents écrits et oraux à un groupe, en extraire les aspects importants pour eux et les réutiliser pour élaborer leur propre rapport / leur propre présentation ou solution d'un problème. On peut voir les détails un à un, mais sans avoir la vue d'ensemble que vise une approche actionnelle, en d'autres termes l'arbre cache la forêt.

Dans une approche actionnelle, toutes les activités langagières et les autres aspects de la langue (grammaire, vocabulaire, LSP, style, registre etc.) sont associés à travers une approche intégrative et holistique. Pour des raisons de transparence, de validité apparente et de comparabilité, mais surtout à des fins d'analyse, on teste les quatre activités langagières de manière distincte en donnant aux apprenants des tâches qui ciblent une compétence donnée à la fois. Mais ces tâches sont toutes intégrées dans une macro-tâche qui traite d'une situation spécifique quasi-authentique dans un contexte donné.

Alors que les tâches de production peuvent facilement correspondre à des situations de la vie réelle, ce n'est pas toujours le cas pour les parties évaluant les compétences réceptives. Sur ce point on a encore besoin de progresser pour les rendre plus authentiques dans un environnement d'enseignement et d'apprentissage à l'université.²⁵

Faut-il tester les compétences réceptives?

Pour rendre un test ou un examen aussi authentique que possible, l'idéal est de n'évaluer que les compétences de production. Comme la production de textes, autant parlés qu'écrits, repose naturellement toujours sur l'analyse de textes écrits et oraux, donc sur la réception, les compétences réceptives sont également nécessaires. Si les candidats ne comprennent pas les textes, ils ne peuvent pas proposer une solution appropriée au problème. En effet, dans des situations réelles sur le lieu de travail, il n'y a pas d'examens d'écoute ou de lecture, en dehors, parfois, de résumés du problème à traiter. C'est pourquoi dans une situation d'évaluation aussi authentique que possible, on opterait pour une approche totalement holistique. On ne testerait pas les compétences réceptives séparément, elles seraient intégrées dans les parties portant sur les compétences productives.

Cependant, une situation d'enseignement et d'évaluation est différente: si l'objectif principal est d'évaluer les forces et les faiblesses des apprenants, on devrait également évaluer leurs compétences réceptives. En particulier, des utilisateurs compétents d'une langue étrangère ont tendance à mal réussir dans les activités de réception. Ces

25 On peut obtenir un plus haut niveau d'authenticité en demandant aux apprenants d'utiliser la langue dans une situation de la vie réelle dans le pays où la langue cible est parlée. Ce n'est cependant normalement pas possible dans le cadre d'un examen de langue à l'université.

apprenants survolent simplement les textes (et ont un mauvais résultat dans la partie de l'examen portant sur les activités réceptives) et se concentrent sur la formulation de leurs idées (et réussissent bien dans la partie portant sur les activités de production). En fait, ils pourraient avoir encore besoin d'améliorer leurs compétences de lecture ou d'écoute, c'est pourquoi il est important à la fois de leur faire travailler et de tester leurs compétences réceptives.

Par ailleurs, les examens de langues à l'université sont toujours comparés aux tests standardisés existant sur le marché (international) des tests de langues. La plupart des examens à enjeux élevés évaluent à part chacune des quatre compétences, en attribuant souvent des points ou des notes séparés à chaque compétence. C'est pourquoi l'équipe GULT estime qu'il est important de tester aussi les quatre compétences, ne serait-ce que pour des raisons de validité apparente. Néanmoins l'équipe GULT est à la recherche de voies innovantes qui permettent de rendre l'évaluation des compétences réceptives plus authentique.

Un test ou examen GULT testant séparément les quatre compétences, mais dans le cadre d'une situation unique, est donc un compromis entre une situation langagière authentique dans la vie réelle et les traditions existantes en matière de tests.

5.3. La structure d'un test ou d'un examen par tâches

La structure GULT

En s'appuyant sur sa pratique des examens CLES (Certificat de compétences en langues de l'enseignement supérieur) et UNICert, l'équipe GULT, en coopération avec le Comité Scientifique d'UNICert, a élaboré la structure suivante pour un examen par les tâches:

- introduction à la thématique et à la (macro-)tâche,
- compréhension de l'oral et de l'écrit²⁶,
- travail d'étude de cas / d'analyse de cas (en binômes ou en petits groupes),
- production écrite,
- production orale.

26 Si un établissement souhaite rendre son examen encore plus authentique, il pourra opter pour une approche complètement intégrative et ne pas tester les capacités réceptives séparément mais à travers les productions écrites et orales, comme décrit plus avant.

Il est cependant important de remarquer que des champs thématiques différents peuvent nécessiter des structures de tests différentes et différentes tâches ou types de tâches. On peut facilement adopter une approche basée sur un problème dans le contexte des affaires où les hommes d'affaires élaborent des solutions à des problèmes existants. Ceci pourrait ne pas être toujours le cas en sciences et en ingénierie, où il peut plutôt s'agir de réaliser des projets ou d'élaborer de nouvelles machines ou de nouveaux produits.

Les paragraphes suivants traitent spécifiquement des examens UNiCert en tant qu'illustration de la manière de mettre en œuvre un examen reposant sur une tâche. Dans la ressource documentaire 2, vous trouverez également le modèle CLES, un autre exemple de mise en œuvre de l'approche par tâches dans les épreuves de langues à l'université.

Etape 1: Introduction à la thématique

Il est important de décrire le cadre général et la macro-tâche aux apprenants, car cela leur permettra de concentrer leur attention par la suite sur les informations pertinentes. Dans les examens « traditionnels », les candidats recherchent surtout les informations particulières demandées dans chaque item pris isolément, mais le sens général du texte pris dans son ensemble leur échappe très souvent. Dans une approche par tâches par contre, les candidats recherchent les informations dont ils ont besoin pour résoudre le problème posé. Comme dans la vraie vie, l'attention se focalise beaucoup plus sur les éléments pertinents. C'est la raison pour laquelle l'introduction au problème / projet est cruciale pour la compréhension de l'étudiant.

Etape 2: Les compétences réceptives

L'écoute et la lecture ont une nouvelle fonction. On ne demande plus aux candidats de « comprendre pour comprendre », c'est-à-dire que la consigne n'est plus « Répondez aux questions ». Les candidats doivent maintenant comprendre le texte en ayant à l'esprit une intention précise, cela signifie qu'ils doivent gérer la (macro-)tâche et qu'ils auront besoin d'utiliser les informations fournies par les textes dans la partie de l'épreuve qui teste les compétences productives.

Il importe que les textes écrits et d'écoute couvrent des aspects différents, mais ils doivent être en rapport avec la thématique générale, afin d'éviter que les candidats utilisent les informations tirées des documents d'écoute pour répondre aux questions dans la partie lecture et vice versa.

L'équipe GULT a examiné différentes options concernant l'ordre dans lequel il faudrait tester les compétences réceptives. Comme les candidats trouvent souvent l'écoute plus

difficile, on a pensé qu'il serait bon de commencer par la lecture (afin que les candidats aient une idée générale de la thématique), puis d'interrompre l'activité et de passer à l'écoute avant de revenir pour terminer à la lecture. Pour des raisons pratiques cependant, cela ne semble pas faisable, car il serait sans doute quelque peu dérangent pour les étudiants de commencer la partie de compréhension de l'écrit, puis de l'interrompre. C'est pourquoi, en pratique, le test commence toujours par l'écoute, mais on insiste sur la présentation de la tâche pour s'assurer que les candidats sont bien au courant de la macro-tâche et de son contexte avant de débiter l'écoute.

C'est pourquoi on a décidé de donner en même temps aux candidats les deux parties de l'examen testant les compétences réceptives (lire et écouter). Une fois qu'on a fait passer les enregistrements, chaque candidat peut décider s'il veut passer plus de temps sur les tâches d'écoute ou de lecture. Il n'a pas besoin d'attendre que les autres candidats aient fini les tâches d'écoute avant de passer à la lecture. A nouveau, cela correspond à une situation de la vraie vie où les gens ont une certaine flexibilité dans l'organisation de leur travail.

Etape 3: Le travail d'étude de cas / d'analyse de cas

L'élément nouveau du test GULT est l'analyse du dossier en autonomie par les candidats. Les étudiants sont mis en binômes ou petits groupes qui élaboreront ensemble une solution ou une proposition qu'ils présenteront ensemble dans la partie orale de l'examen. En fonction des caractéristiques locales, les candidats peuvent avoir une totale liberté et utiliser toutes les ressources disponibles, par exemple la bibliothèque, les possibilités des centres de ressources ou l'internet.

Une différence importante entre les tests par tâches d'un côté et de nombreux tests standardisés et examens universitaires de l'autre réside dans l'utilisation d'un dossier très complet de ressources authentiques. L'objectif, ce faisant, est de préparer les apprenants et candidats à des situations réalistes dans leur vie professionnelle future, en entraînant et en testant leur capacité à exploiter des ressources – une tâche qu'ils auront à accomplir par la suite dans leur carrière.

Etape 4: Les compétences productives

Dans les parties écrite et orale de l'examen, les candidats doivent présenter une production particulière dans un contexte concret, ce qui signifie que la macro-tâche est inscrite dans une situation réaliste concrète et un cadre qui correspondent à des activités réelles. A cette fin, ils doivent aussi utiliser et traiter les informations obtenues dans les autres parties de l'épreuve, y compris les parties d'écoute et de lecture.

Etape 4a: Ecrire

Pour la partie de production écrite de l'examen, les apprenants doivent à nouveau travailler seuls, car on veut évaluer la performance de chaque candidat individuellement à l'écrit. Ils sont néanmoins autorisés à utiliser le dossier et les notes prises lors de la phase d'analyse de cas.

A la fin de cette partie, on photocopie les productions écrites des copies, puisqu'elles seront à nouveau nécessaires pour l'épreuve orale.

Etape 4b: Parler

L'épreuve orale n'a pas lieu immédiatement après les autres parties de l'examen, car les apprenants pourraient être déjà assez fatigués et cela désavantagerait ceux qui passeraient en premier. C'est pourquoi il vaut mieux faire passer l'épreuve orale quelques jours plus tard.

Pour la partie orale, les candidats qui ont collaboré lors de la phase d'analyse de cas se retrouvent en binômes ou en petits groupes. L'épreuve comporte les parties suivantes:

- préparation: avant l'épreuve elle-même, on donne aux candidats une copie de leur production écrite et on leur demande de préparer une présentation conjointe de leur solution (présentation à deux ou en groupe);
- épreuve – première partie: présentation;
- épreuve – deuxième partie: discussion, à savoir clarification de certains aspects, réponse à des questions.

Une telle situation correspond à une situation réelle: on demande à une équipe de lancer un projet ou de résoudre un problème; elle doit analyser la situation, préparer un rapport écrit, qu'elle présente ensuite à l'équipe ou au chef de service. Cela rend la situation d'examen bien plus authentique et les candidats investissent davantage d'énergie dans leurs propositions, car ils ressentent le besoin de montrer leurs compétences (en langue et dans leur domaine de spécialité) et d'essayer de convaincre leurs interlocuteurs.

La longueur de la partie orale était un souci majeur chez les professeurs de langues qui ont assisté aux ateliers GULT ou à des présentations. Certaines universités ont rapporté qu'elles avaient complètement supprimé la partie orale, car elles n'étaient pas capables de gérer le grand nombre de candidats. L'équipe GULT estime néanmoins que la production orale devrait être un élément clé de tout test de langue et qu'il est plus important de la tester que de tester les compétences réceptives ou la grammaire et le lexique. Il ne peut y avoir de communication que si les gens peuvent s'exprimer, à

l'écrit et à l'oral. Toutefois, les établissements peuvent opter pour une durée moins longue de la partie orale de l'examen. L'équipe pense que les jugements des enseignants seront d'autant plus réalistes que l'on aura consacré plus de temps à la partie orale, mais on comprend que dans bien des cas seul le strict minimum peut être mis en place.

Durée

En ce qui concerne la durée, à titre d'exemple, les durées suivantes ont été expérimentées avec succès pour un examen UNICert III au niveau C1 du CECRL²⁷, bien que ceci dépende évidemment de la longueur des textes et du nombre et de la complexité des questions ou des tâches:

- compréhension de l'oral et compréhension de l'écrit: 90 minutes pour réaliser les tâches d'écoute et de lecture après deux auditions des enregistrements;
- travail d'étude de cas / analyse de cas (en binômes ou en petits groupes): 90 minutes, incluant une pause (possible) de 30 minutes;
- production écrite: 90 minutes;
- oral: 30 minutes par candidat (60 minutes par binôme).

Bien sûr, ceci n'est qu'un exemple dans un contexte donné, les durées peuvent être modifiées en fonction des besoins et des ressources de l'établissement.

La durée de l'examen sera plus courte au niveau B2 du CECRL.

27 Cette structure a été élaborée par le Comité Scientifique d'UNICert lors d'une rencontre de travail à Dötlingen, Allemagne. Elle est connue depuis sous le nom de « Modèle de Dötlingen » (*Dötlinger Modell*). Vous trouverez les durées de l'examen CLES dans la ressource documentaire 2.

6. Qui peut tirer bénéfice de l'évaluation par les tâches et en quoi ?

En général, lorsque l'évaluation en langues est conçue dans une approche actionnelle, l'épreuve a un sens et les tâches intermédiaires ont une fonction précise, puisqu'elles s'insèrent dans le cadre de la macro-tâche de l'examen.

On peut considérer qu'avec une approche par tâches, les tests de langues sont authentiques, motivants, centrés sur l'étudiant et flexibles:

- L'évaluation est **authentique**, car les documents utilisés dans les tests sont authentiques et les tâches sont inspirées de celles auxquelles les étudiants seront sans doute confrontés dans un futur proche. En d'autres termes, dans un test d'anglais économique par exemple, on ne demande pas aux étudiants de se mettre à la place d'un directeur mais plutôt d'un stagiaire ou d'un assistant.
- Elle est **motivante**, parce que l'ensemble du processus d'évaluation est clairement en lien avec le domaine d'études de l'étudiant. On peut dire que les tests basés sur une tâche vont au-delà des tests traditionnels en langues à fins spécifiques, qui partent de documents spécialisés dans un domaine donné, mais proposent des items traditionnels plutôt que des tâches visant à atteindre un objectif. L'épreuve a un sens, parce que les tâches intermédiaires ont une fonction précise, dans le cadre de la macro-tâche.
- L'évaluation est **centrée sur l'étudiant**, car 1) l'épreuve porte entièrement sur les capacités langagières et les ressources nécessaires dans le domaine d'études de l'étudiant, et 2) les productions réalisées lors de l'évaluation sont totalement conçues par l'étudiant, même si les ressources et les supports d'évaluation ont été fournis par les examinateurs.
- Elle est **flexible**, parce que l'approche par tâches convient à de nombreuses situations d'évaluation, des tests de connaissances aux tests de niveau de capacité. Elle peut être utilisée en contrôle continu ou pour des évaluations de fin de cours et cibler une ou plusieurs compétences. Dans des tests de niveau de capacité reposant sur des tâches, on évalue les quatre compétences.

Lors de la passation d'un examen conçu dans une approche actionnelle, on a fait plusieurs observations au sujet des étudiants:

- Ils paraissent moins angoissés que lors d'examens traditionnels, ce qui peut être dû à la composante sociale du test, à savoir le travail en binôme ou en groupe.

- Ils semblent se concentrer (davantage) sur la tâche à cause de sa structure et de son importance pour la production finale; en d'autres termes, le test n'est pas composé d'items sans rapport les uns avec les autres, mais d'une tâche unique.
- Les étudiants s'impliquent davantage dans le sujet / le contenu à cause de son authenticité et de sa pertinence pour la vraie vie.
- Les étudiants parviennent à intégrer les informations recueillies dans les parties de lecture et d'écoute dans les parties de production écrite et orale de l'examen. En d'autres termes, les compétences réceptives ne sont pas évaluées hors de tout contexte, elles sont au contraire pertinentes pour l'évaluation des compétences de production.
- Ils ont tendance à obtenir de meilleurs résultats globaux. Chaque étudiant peut apporter quelque chose lors de l'épreuve, ce qui semble motiver les étudiants à donner le meilleur d'eux-mêmes.
- Ils sont davantage aptes à montrer l'étendue de leurs savoir-faire en lecture et en écriture, en particulier les étudiants très compétents en langues. Dans les tests de langues traditionnels, on attend de tous les étudiants qu'ils répondent de la même manière aux mêmes questions. Dans les tests reposant sur une tâche, les étudiants décident eux-mêmes de la longueur et du contenu de leurs textes parlés ou écrits, dans lesquels ils peuvent faire la preuve de leurs connaissances et capacités particulières, en exploitant par exemple les informations implicites dans la tâche de lecture.
- Les étudiants sont capables de dépasser leur réaction initiale éventuellement négative; ils semblent plus satisfaits après l'examen. Certains étudiants peuvent dans un premier temps réagir négativement à un examen par tâches, mais ils ont tous tendance à être plus satisfaits après un examen de ce type qu'après un examen traditionnel. Une raison peut en être que, après avoir réalisé que l'implication de l'étudiant dans un examen reposant sur une tâche est d'un tout autre ordre que de répondre à des questions ou de résoudre des items dans un examen de langue traditionnel, les étudiants prennent conscience de la liberté qu'ils ont dans l'élaboration de leurs réponses.
- Il semble qu'il y ait moins de stress, peut-être à cause du travail de groupe pendant la phase de préparation.
- On observe aussi moins de stress pendant l'examen, parce que les candidats travaillent sur un problème authentique (il est visible qu'ils s'impliquent fortement dans la thématique).
- Ils sont étonnés de la variété des solutions correctes possibles.

- Ils engagent et mènent des discussions très intéressantes pendant l'épreuve orale.

Quelques observations ont également été faites au sujet des examinateurs:

- Il semble que leur implication et leur concentration sur la thématique de la tâche lors de l'examen soient plus fortes, c'est-à-dire que le test actuel suscite nettement plus d'intérêt.
- Les examinateurs ont un rôle plus actif dans la discussion, le second évaluateur y prenant également part. Ceci conduit à davantage d'interaction avec les candidats lors de l'évaluation des savoir-faire oraux.
- Le climat de la situation d'évaluation paraît plus détendu: « On ne ressent pas le stress des étudiants », selon le commentaire d'un examinateur.
- Ils ont tendance à oublier qu'ils sont en situation d'examen.
- Ils sont surpris de voir combien les discussions entre les étudiants sont animées.
- Les examinateurs sont également surpris de devoir arrêter des candidats à la fin du test: « Malheureusement l'épreuve est terminée et nous devons nous arrêter ».
- Ils peuvent vraiment voir la capacité des étudiants à gérer linguistiquement une situation.

De l'analyse des commentaires des candidats et des évaluateurs, il est ressorti que les principaux avantages mentionnés étaient la plus grande motivation des candidats et des examinateurs, ce qui a conduit à de meilleurs résultats et à un taux de satisfaction plus élevé.

L'un dans l'autre, on peut conclure qu'évaluer par les tâches

- augmente la motivation et l'implication (des candidats et des examinateurs);
- a une meilleure validité apparente; et
- peut-être aussi une validité de contenu, une validité simultanée et une validité consécutive plus élevées.

Les approches par tâches stimulent la compétence langagière dans les quatre habiletés mais aident également à développer les savoir-faire nécessaires pour réussir dans la vie professionnelle, et pas seulement dans les études de langues. Ces savoir-faire comprennent:

- l'aptitude au travail de groupe,
- l'aptitude à résoudre des problèmes,
- la compétence de présentation,
- la compétence de discussion,
- la compétence de négociation,
- l'aptitude au compromis,
- la compétence interculturelle dans un environnement international,
- les savoir-faire universitaires / « compétences transversales »,
- la motivation de l'apprenant.

Idéalement, dans un contexte universitaire, toute situation d'évaluation devrait aussi être une situation d'apprentissage, et les expériences passées nous montrent que, dans l'évaluation par les tâches, chaque situation d'évaluation devient une situation d'apprentissage, pour le candidat et pour l'examineur.

Bien sûr introduire l'approche actionnelle pour l'évaluation dans les cours de langues va conduire à un enseignement qui prépare au format du test. L'équipe GULT considère qu'il s'agit pour l'essentiel d'un « effet en retour »²⁸ positif, car cela conduira à introduire l'approche actionnelle dans l'enseignement et l'apprentissage, et les apprenants cesseront de concentrer leur attention sur des aspects de la langue tels que la grammaire et le vocabulaire hors contexte et de travailler sur des parties de tests à items discrets. De plus, les apprenants seront encouragés à travailler par eux-mêmes sur des sujets d'examens par tâches donnés antérieurement lorsqu'ils se prépareront au test à venir.

En effet, avec l'approche par tâches il n'est pas possible de s'entraîner pour le test sans améliorer considérablement ses compétences langagières, c'est-à-dire qu'il n'y a pas de raccourcis menant au test, comme on pourrait le voir avec certains tests standardisés où des tâches ou activités standardisées sont répétées dans chaque test sous une forme ou une autre.

C'est pourquoi chaque enseignant doit s'assurer que l'approche par tâches ne conduit pas à un effet en retour négatif, ce qui serait le cas si les thématiques et les tâches intermédiaires se répétaient et restreignaient l'étendue des sujets à traiter et des types de tâches. Cela signifie que les candidats ne peuvent pas se préparer à l'avance à certaines parties du test, par exemple en apprenant certaines phrases ou parties par cœur.

28 Voir également Norris, 2002: 341.

7. Défis et limites

Les limites des tests par tâches relèvent généralement de deux catégories: des questions d'ordre psychométrique et des réserves concrètes, liées à la mise en œuvre pratique.

7.1. Réserves d'ordre psychométrique

Au nombre des problèmes liés aux tests par tâches on compte la difficulté à concevoir un test véritablement authentique. De par sa nature même, un test n'est pas une situation « authentique ». Même si l'on utilise des documents authentiques, on sera amené à demander aux candidats de les exploiter selon des modalités qui ne reflètent pas leur utilisation dans la vraie vie. Comme traité précédemment, la meilleure représentation que l'on puisse se faire de l'authenticité est peut-être le point ultime d'un continuum, que les concepteurs de tests devraient tenter d'approcher au plus près tout en reconnaissant qu'il est inaccessible.

Une autre critique faite à l'évaluation par les tâches est son manque de généralisabilité. Bachman, par exemple, a souligné que la complexité d'une tâche signifie que la performance fournie par un candidat réalisant une tâche lors d'un test ne peut pas être utilisée pour prédire sa performance dans une situation réelle ultérieure. En d'autres termes, il est en pratique impossible de généraliser d'une tâche d'évaluation à une autre, ou d'extrapoler à partir d'une tâche vers d'autres tâches ou vers la vraie vie²⁹. Même si cela est vrai, l'équipe GULT estime cependant que ceci est également vrai (ou même encore davantage) pour tous les autres tests de langues et que cela n'est par conséquent pas une caractéristique spécifique de l'évaluation par les tâches. En fait, au vu d'analyses de besoins préliminaires – qui devraient être effectuées pour toute structure de test – les tests par tâches correspondent davantage à des situations réelles dans le contexte du candidat et sont susceptibles de donner une meilleure image de la compétence langagière de celui-ci.

7.2. Réserves d'ordre pratique

Les défis liés à la mise en œuvre de l'évaluation par les tâches sont nombreux, à chaque niveau: aux niveaux individuel, local, et même national.

29 Bachman, 2002: 461f.

Les enseignants d'université, par exemple, sont avec raison inquiets au niveau individuel de l'augmentation de la charge de travail liée à l'évaluation par les tâches. Dans de nombreux établissements d'enseignement supérieur, l'évaluation n'a pas fait l'objet d'une prise en compte spécifique dans les charges de travail des universitaires. Il est clair que, pour les enseignants, évaluer selon une approche actionnelle signifie

- au début plus de travail pour concevoir les tests, et également
- au début plus de travail de correction.

Il faut du temps pour prendre l'habitude de rassembler, pour un test par tâches, un ensemble représentatif de documents. Habituellement, de par la thématique d'actualité des documents et pour des raisons de sécurité des épreuves, les documents ne peuvent être utilisés qu'une seule fois à des fins d'évaluation, même s'ils peuvent ensuite être réutilisés en cours. Il faut du temps aussi pour apprendre à imaginer les bons types de tâches, mais une fois que c'est acquis, la même structure peut être utilisée dans les tests par tâches ultérieurs.

Comme, dans l'évaluation par les tâches, il n'y a pas de réponses justes ou fausses aux tâches à réaliser, les examinateurs doivent être particulièrement attentifs à la fidélité inter- et intra-individuelle. De plus, il est nécessaire de réunir les examinateurs après chaque test par tâches d'importance majeure pour qu'ils s'accordent sur les critères d'évaluation, en particulier quand on certifie le niveau de capacité.

Au niveau des établissements, les coûts peuvent être un obstacle à la mise en place de l'évaluation par les tâches. L'augmentation de la charge de travail des enseignants liée à la conception et à la notation de tests par tâches se traduit souvent par des rémunérations plus élevées. De plus, les évaluations par tâches peuvent nécessiter davantage de logistique, par exemple l'accès à l'internet dans les salles d'examens et / ou plus de photocopies.

Au niveau national, des réserves ont été exprimées au sujet de la compatibilité de l'évaluation par les tâches avec les cultures de l'évaluation et de l'enseignement en langues existantes. Souvent, l'évaluation par les tâches entre en conflit avec des traditions et conceptions de l'éducation ancrées de longue date. Des experts de l'évaluation représentant différents pays européens et leurs universités ont formulé les commentaires suivants:

- On s'est interrogé sur la pertinence / la compatibilité de l'évaluation par les tâches (TBA) avec des identités locales ou nationales variées.
- Il est difficile de modifier une culture de l'évaluation à l'échelle nationale, cela prend beaucoup de temps.

- Les aspects pragmatiques de la mise en place de l'évaluation par les tâches devraient être analysés au niveau local, et les dispositions nécessaires devraient être discutées et adoptées à l'unanimité.

7.3. Autres réserves

Les experts en évaluation consultés lors de l'élaboration du projet GULT ont soulevé plusieurs autres points:

- **Contrôle qualité:** comment peut-on contrôler la qualité des tests par tâches ?
- **Applicabilité des grilles et des critères d'évaluation:** quels grilles et critères d'évaluation sont-ils adaptés à une évaluation par les tâches ?
- **Lien entre l'évaluation par les tâches et le CECRL:** comment les tests par tâches peuvent-ils être reliés au Cadre européen commun de référence pour les langues, dans lequel l'évaluation par les tâches n'est même pas mentionnée à titre d'exemple ?
- **Absence d'auto-évaluation / de réflexion (de l'étudiant):** comment les étudiants peuvent-ils être impliqués dans l'évaluation de tests par tâches; devrait-il y avoir une « sous-tâche de réflexion » supplémentaire ?
- **Autonomie des étudiants:** les étudiants devraient-ils être plus impliqués dans la recherche de sources documentaires pour les tests par tâches ? Dans le modèle GULT, ce sont les concepteurs des sujets qui fournissent actuellement les dossiers documentaires, mais en fonction du contexte, des traditions et des équipements locaux, on peut aussi envisager de donner accès à diverses ressources.

8. L'évaluation de la performance du candidat

Imaginer une méthode pour recueillir la performance langagière d'un candidat est une chose; c'en est une autre, tout aussi importante, de décider comment évaluer l'échantillon langagier obtenu. En effet, les résultats des tests ne sont utiles et utilisables que si les procédures d'évaluation sont appropriées. Certaines épreuves par tâches incluent des sous-parties de compréhension de l'écrit ou de l'oral et, dans ces cas, le mode d'évaluation est généralement tout à fait classique. Pour ce qui est de la macro-tâche et des parties de production de l'épreuve, par contre, l'évaluation de la performance implique normalement un jugement subjectif, ce qui en fait une entreprise plus compliquée. Comme décrit au chapitre 7, on devrait prêter une attention particulière aux grilles d'évaluation et à la fiabilité des examinateurs.

Les compétences réceptives

Dans le cadre GULT comme dans d'autres modèles de tests par tâches, des parties de l'épreuve portent sur la compréhension de l'écrit et de l'oral. Bien que ces activités portant sur les capacités réceptives soient conçues comme des tâches intermédiaires aidant le candidat à réaliser la macro-tâche, elles sont en général notées en tant que parties à part. En ce qui concerne leur structure, les activités GULT portant sur les capacités réceptives ne sont pas très différentes de celles qu'on trouve dans les tests existants: les candidats écoutent et lisent un ou plusieurs textes authentiques, puis répondent à des questions vérifiant la compréhension globale et la compréhension de détails particuliers du / des texte(s). On devrait utiliser une variété de formats pour que se révèlent les forces et les faiblesses individuelles des candidats. Comme pour tout autre test, il faut préparer le corrigé à l'avance pour des raisons de transparence et d'objectivité, ce qui facilite également la correction des tests.

Les compétences productives

Pour évaluer la performance de production, il faut élaborer des échelles de notations / des grilles d'évaluation reflétant le construit qui sous-tend la conception du test. Des contraintes d'ordre pratique limitent habituellement le nombre d'aspects de la performance qui peuvent être évalués par les correcteurs. Pour décider quels aspects sont les plus importants, les concepteurs de tests pourront se reporter aux différentes compétences énumérées dans le CECRL (Conseil de l'Europe, 2001) et dans le *Manuel pour relier les examens de langues au CECRL* (Conseil de l'Europe, 2009); ils pourront suivre les suggestions du *Manuel*, à savoir déterminer les domaines de compétence sur lesquels le test devrait porter et aussi les différents aspects dans chacun de ces domaines, en consultant les descripteurs pertinents pour chaque domaine. La culture ou

L'histoire de l'établissement peuvent aussi imposer d'inclure certains aspects, tels que la grammaire, dans la grille de notation.

Le poids accordé à une réalisation réussie de la tâche est également une décision majeure dans l'élaboration de cette échelle. Dans certains modèles d'évaluation par tâches, réussir à mener à bien la tâche est l'unique critère utilisé pour juger la performance du candidat. Dans la plupart des cas cependant, les échelles d'évaluation incluent des critères liés à la fois à la compétence langagière et à la réalisation de la tâche. De nombreuses grilles d'évaluation prennent en compte pour la production écrite le contenu, les compétences pragmatiques et les compétences linguistiques, et pour la production orale le contenu, les compétences pragmatiques, linguistiques et paralinguistiques. Elles pourront également inclure des aspects tels que la complexité, la précision et la fluidité.

Après avoir arrêté le choix des critères à utiliser pour juger la performance du candidat, l'établissement doit déterminer le poids de chacun de ces éléments dans la note finale. Les critères linguistiques seront sans doute plus importants et on leur attribuera un pourcentage plus élevé à un niveau plus avancé, car la correction linguistique prend de plus en plus d'importance à des niveaux de compétence plus élevés. A des niveaux plus bas, nous attendons seulement des apprenants qu'ils soient capables de transmettre leur message (de là l'importance du contenu et des aspects pragmatiques). A des niveaux plus élevés, nous attendons d'eux qu'ils expriment leurs idées de manière linguistiquement correcte. Les grilles utilisées devraient garantir que le nombre de faux positifs et de faux négatifs³⁰ sera aussi réduit que possible.

Dans les cas où les résultats serviront à prendre des décisions importantes telles que l'exclusion ou l'admission dans une formation ou un programme d'études, les auteurs du test devront également fixer des normes ou des scores de césure pour ces décisions cruciales. Ce faisant, l'établissement doit prendre en considération la difficulté globale de l'épreuve et la difficulté des différentes tâches, ainsi que le cadre institutionnel général. Dans de nombreux cas, ces aspects sont fixés par des règlements de l'université. Ceci signifie que l'établissement doit définir quelles sont les notes minimales pour qu'un résultat soit acceptable, et il doit attribuer les notes possibles en fonction des différentes tâches, en conformité avec les règlements de l'université. Ceci signifie en bref qu'on applique les mêmes normes et procédures que pour tout autre test de langue élaboré dans les règles.

30 On utilise les concepts de « faux négatifs » et de « faux positifs » en référence à des étudiants qui ont échoué alors qu'ils auraient dû réussir au test (faux négatifs) ou qui ont réussi alors qu'ils auraient dû échouer (faux positifs).

Pour garantir la comparabilité des tests par tâches avec d'autres tests de langues, ainsi que la comparabilité des résultats³¹, il est important de s'assurer que les examinateurs ont la même interprétation des descripteurs du CECRL. Ceci nécessite d'organiser des sessions de formation et des rencontres régulières des équipes, et d'utiliser consciencieusement le *Manuel pour relier les examens de langues au CECRL* (Conseil de l'Europe, 2009).

Des tâches et des échelles de notation élaborées avec soin ne pourront produire de résultats utilisables que si les examinateurs sont bien formés à l'utilisation de ces échelles. Il est important pour garantir la fidélité inter-examinateurs, c'est-à-dire la cohérence entre les différents examinateurs, d'organiser des sessions de formation, afin de s'assurer que tous les examinateurs comprennent et appliquent les grilles d'évaluation de la même manière. Ces sessions sont d'autant plus importantes que l'évaluation n'occupe habituellement pas une place importante dans la formation des enseignants de langues et ne fait pas partie des qualifications requises lors du recrutement d'enseignants de langues à l'université (l'importance de l'évaluation en langues est en effet souvent négligée par l'institution dans son ensemble et les compétences des enseignants dans ce domaine sont considérées comme allant de soi). De plus, ces sessions de formation amélioreront la fiabilité des jugements des enseignants sur la performance des étudiants.

Si on peut se contenter d'un évaluateur pour un examen à enjeux limités (par exemple un examen de fin de cours), des examens à enjeux élevés (le plus souvent des tests de niveau de capacité ou des « prochievement tests ») nécessitent deux évaluateurs. Dans ce cas, la répartition des rôles doit être définie au préalable, afin que chaque examinateur sache comment se comporter pendant l'épreuve.

31 Bachman, 2002: 455, 461.

9. Usages possibles des tests / examens par tâches: comment exploiter les tests par tâches

Comme évoqué au début, l'enseignement par les tâches va de pair avec les tests et l'évaluation par les tâches. Bien que les tests par tâches requièrent un certain niveau d'authenticité et nécessitent d'aborder des événements d'actualité, rien n'oblige à les jeter une fois qu'ils ont rempli leur fonction d'évaluation. Ils peuvent être à nouveau utilisés et exploités. Les tests par tâches peuvent aisément être utilisés en cours et dans des programmes d'auto-formation à d'autres fins que l'évaluation, par exemple:

- comme matériel de cours,
- comme ressources d'apprentissage dans des programmes d'auto-formation,
- comme matériel d'entraînement au test,
- comme tests d'auto-évaluation,
- comme modèles de tests.

9.1. Tests de connaissances: examens de fin de cours

Les tests par tâches peuvent être utilisés en tant qu'évaluations de fin de cours. Dans ce cas, pour des raisons de durée, il ne sera le plus souvent pas possible de suivre l'approche intégrée mentionnée précédemment, qui évalue les quatre compétences dans une épreuve unique. C'est pourquoi un établissement ou un enseignant de langues pourra utiliser une étude de cas pour évaluer uniquement la production orale, uniquement la production écrite, ou les deux conjointement.

Il est bien clair que les tâches doivent correspondre à l'enseignement dispensé pendant la formation, c'est-à-dire que les apprenants doivent avoir été habitués aux études de cas en en réalisant une ou plusieurs en cours s'ils doivent travailler sur une étude de cas lors de l'examen terminal.

Ces dernières années, des établissements ont adopté dans leurs référentiels de cours une approche « portfolio » pour l'enseignement et l'évaluation en langues. Dans ce cas, les apprenants travaillent sur un ou plusieurs projets, une simulation globale et / ou des études de cas tout au long de la formation, et les résultats de ces activités composeront la note finale. Certains aspects ou certains types de textes peuvent être évalués pendant la formation; d'autres seront évalués lors de l'examen terminal. Si, par exemple, un enseignant utilise une simulation globale en cours, il pourra demander aux apprenants pendant le cours de rédiger une lettre commerciale en relation avec leur projet de

simulation globale; et à l'examen terminal, il leur demandera de rédiger un compte rendu sur une étude de cas en lien avec leur projet. De cette manière, l'enseignant peut tester la capacité des candidats à rédiger différents types de textes dont ils pourraient avoir besoin dans leur vie professionnelle future.

9.2. Tests de niveau de capacité et « prochievement tests »

L'approche par tâches peut également être utilisée pour tester le niveau de capacité (ou pour des « prochievement tests »). L'examen CLES et l'examen UNiCert sont tous deux des modèles en usage de tests de niveau de capacité / de « prochievement tests » construits sur des tâches. Ils ne sont pas directement en lien avec une formation, mais évaluent la compétence langagière globale du candidat à la fin d'un processus d'enseignement et / ou d'apprentissage.

L'expression « prochievement tests » définit une combinaison des tests de niveau de capacité et des tests de connaissances, ce qui signifie qu'ils sont reliés à un référentiel particulier, mais vont au-delà, puisque l'évaluation est critérielle:

In general, the profession defines achievement tests as those limited to a particular body of material just covered in class(es) and proficiency instruments as those testing the total range of skills and contexts a learner may be able to handle – regardless of where and when they may have been learned – and testing them through actual interaction in realistic situations. *Prochievement tests* are a combination of the preceding two types, testing students' ability to perform in only the contexts and situations that have been practiced in class.

(Gonzalez Pino, 1998: 120)

Traduction française

En général, la profession définit comme tests de connaissances les tests limités à un contenu particulier étudié en cours et comme instruments de mesure du niveau de capacité ceux qui portent sur l'éventail complet des savoir-faire et des contextes qu'un apprenant devrait être capable de gérer – peu importe où et quand ils ont été acquis – et qui les évaluent par une interaction réelle dans des situations réalistes. Les « *prochievement tests* » sont une combinaison des deux types précédents, ils évaluent la capacité des étudiants à réussir dans les seuls contextes et situations qui ont été travaillés en cours.

Dans le contexte d'UNiCert, les apprenants doivent d'abord suivre au moins 112 heures d'enseignement en présentiel (plus une part importante de travail personnel) et réussir leurs examens de fin de cours, avant d'être autorisés à s'inscrire à un examen UNiCert. C'est pourquoi on peut le considérer comme un « prochievement test ». Le CLES, de son côté, n'exige pas nécessairement d'avoir suivi au préalable une formation ou un

programme spécifique, même si les étudiants le feront en général avant de s'inscrire à la certification. C'est donc un test de niveau de capacité évaluant la compétence linguistique dans le cadre d'une situation spécifique.

Si on compare les examens UNICert et CLES, UNICert attribue des notes pour chaque examen puisqu'il est totalement intégré dans le système d'enseignement et d'évaluation de l'université (les notes sont celles qui sont utilisées dans chaque établissement), alors que CLES n'attribue pas de notes mais fonctionne selon le principe « réussite / échec ». Le CLES peut néanmoins être utilisé comme une évaluation de fin de cours (par exemple à l'Université de Strasbourg), auquel cas il y a des notes. Ceci démontre que les tests et examens GULT peuvent opter pour un système « réussite / échec » ou pour un barème de notation détaillé.

Pour obtenir davantage de détails sur les différentes utilisations d'UNICert et de CLES, reportez-vous à la ressource documentaire 2.

10. Comment élaborer un test par tâches

10.1. La conception du test

10.1.1. La problématique

Lorsqu'il élabore un sujet d'examen reposant sur une tâche, le concepteur doit tout d'abord rechercher un problème ou un projet authentique. Cela doit être un problème (ou projet) que les candidats ne connaissent pas ou seulement partiellement, car dans le cas contraire ceux qui maîtrisent le domaine sont avantagés par rapport à ceux qui le découvrent. Le problème ou projet doit être authentique et correspondre au domaine de recherche ou aux centres d'intérêt des candidats, c'est-à-dire qu'il doit être pertinent pour eux.

10.1.2. L'authenticité de la tâche

Le concepteur doit ensuite réfléchir à une macro-tâche compatible avec la situation des candidats. Ceci inclut l'authenticité de leur rôle. Cela signifie qu'on ne confiera pas à des étudiants à l'université le rôle du PDG d'une société multinationale, car ils ne commenceront pas leur carrière professionnelle une fois diplômés comme PDG ou ne seront pas tout de suite des managers, mais débiteront à un niveau bien inférieur, et devront grimper l'échelle hiérarchique. Si le rôle et la tâche sont authentiques, les propositions élaborées par les candidats seront elles aussi bien plus authentiques.

10.1.3. Le dossier

A l'étape suivante, le concepteur du sujet doit rassembler des documents appropriés, oraux et écrits, qui traitent du problème ou projet en question. Ce matériel doit ensuite être trié:

- Quel(s) fichier(s) audio(s) / vidéo(s) peut / peuvent être utilisé(s) pour la partie de compréhension de l'oral de l'examen ?
- Quel(s) texte(s) écrit(s) est / sont le mieux adapté(s) pour la partie de compréhension de l'écrit ?
- Les textes écrits / d'écoute sont-ils complémentaires (ou le candidat peut-il utiliser les informations tirées des documents d'écoute pour répondre aux questions dans la partie de lecture) ?
- Quels documents sont utiles pour le dossier ? Dans quel ordre les présenter ? Quels documents peut-on laisser de côté ?

10.1.4. La situation

Maintenant le concepteur doit décrire le contexte de la tâche écrite. Cette description devrait être assez détaillée, afin de garantir que les candidats ont tout à fait compris la situation.

10.1.5. La tâche

A l'étape suivante, le concepteur doit élaborer de manière plus détaillée une macro-tâche réaliste. Cette tâche doit préciser la nature générale de l'activité, à savoir une étude de cas ou un projet. En d'autres termes on dit aux candidats: « Voici la situation, voici votre rôle, voici le problème / le projet, et maintenant vous devez élaborer une solution / une proposition ».

10.1.6. Les tâches intermédiaires pour les différentes compétences

A l'étape finale de la phase de conception, on rédige les différentes tâches intermédiaires pour chacune des quatre compétences. Elles doivent toutes être intégrées dans la macro-tâche et se compléter.

Pour chaque partie de l'examen, une fois qu'il a préparé la tâche, le concepteur doit retourner vers la macro-tâche et vérifier si chaque tâche prise isolément est bien un élément de l'ensemble, si elle aide à atteindre l'objectif global, c'est-à-dire la solution du problème ou l'élaboration d'un projet, et si elle est pertinente pour la macro-tâche. Dans une approche holistique, le concepteur assemble les différentes tâches pour en faire « un tout ».

Lorsqu'il prépare les différentes tâches pour chaque partie de l'examen, le concepteur doit penser à la manière d'intégrer « l'usage pertinent de la langue » (syntaxe et lexique) dans les tâches successives, c'est-à-dire comment faire pour que l'« usage pertinent de la langue » soit (ou devienne) inhérent aux tâches de réception (ou de production).

En général, il peut être très utile de répéter la macro-tâche dans chaque partie du sujet et d'expliquer de manière très détaillée les tâches et productions que les candidats ont à élaborer, car les répétitions aident à expliciter la situation, les rôles et les tâches.

10.2. Etapes suivantes dans l'élaboration du test: essais

Comme tout autre test de langue respectant des mesures d'assurance-qualité, les tests par tâches doivent être pilotés, pré-testés et / ou mis à l'essai, ainsi que décrit dans le *Manual for Language Test Development and Examining* (2011). Dans un contexte universitaire, en particulier pour un test GULT basé sur des thématiques d'actualité, ceci peut signifier:

Piloting involves asking a small number of people to complete items as if in a test. This may be quite informal and could, for example, involve work colleagues if no one else can be found. Their RESPONSES are analysed and, together with their comments [...], may be used to improve items further.

(Manual for language test development and examining, 2011: 30)

Traduction française:

Piloter implique de demander à un nombre réduit de personnes de répondre à des items dans les conditions d'un test. Ceci peut être très informel et pourrait par exemple impliquer des collègues de travail si personne d'autre n'est disponible. Leurs REPONSES sont analysées et peuvent être utilisées, avec leurs commentaires [...], pour améliorer les items.

De la même façon, on peut organiser un pré-test avec un nombre limité de candidats potentiels pour éviter que les candidats fournissent des réponses non prévues lors de la session réelle. Mettre à l'essai, c'est-à-dire se livrer à une analyse statistique à petite échelle de tâches corrigées subjectivement, ne sera sans doute possible que dans des établissements d'une certaine taille ou dans des réseaux d'établissements.

11. Bibliographie³²

Abendroth-Timmer, D. et Breidbach, S. (éd.) (2000). *Handlungsorientierung und Mehrsprachigkeit*. Frankfurt am Main: Peter Lang.

AFANet, site internet <http://www.afanet.info>.

Almagro Esteban, A. et Pérez Cañado, M. L. (2004). « Making the case method work in teaching business English: A case study ». In: *English for Specific Purposes* 23, 137-161.

ALTE (1998). *Multilingual glossary of language testing terms: Studies in language testing* 6. Cambridge: Cambridge University Press.

*Bachman, L. (2002). « Some reflections on task-based language performance assessment ». In: *Language Testing* 19(4), 453-476.

*Bachman, L. (2007). « What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment ». In: Fox, J., Wesche, M., Bayliss, D., Cheng, L., Turner, C.E. et Doe, C. (éd.). *Language testing reconsidered*. Ottawa: University of Ottawa Press, 41-71.

Bausch, K.-R., Burwitz-Melzer, E., Königs, F. G. et Krumm, H.-J. (éd.) (2006). *Aufgabenorientierung als Aufgabe*. Tübingen: Narr.

Bourguignon, C. (2006). « De l'approche communicative à l'approche communicationnelle: une rupture épistémologique en didactique des langues-cultures ». In: *Synergies Europe* 1, 58-73.

Bourguignon, C. (2007). « Apprendre et enseigner les langues dans la perspective actionnelle: le scénario d'apprentissage-action ». Communication faite le 7 mars 2007, et publication en ligne dans: *Les Langues Modernes*, disponible sur le site: <http://www.aplv-languesmodernes.org/spip.php?article865>.

*Bourguignon, C., Delahaye, P. et Puren, C. (2007). *Évaluer dans une perspective actionnelle: L'exemple du « Diplôme de Compétence en Langue »* [disponible auprès de C. Bourguignon].

Brown, J. D. (1998). *New ways of classroom assessment*. Alexandria, VA: Teachers of English to Speakers of Other Languages.

32 Les publications traitant d'aspects spécifiques de l'évaluation par tâches (ou évaluation de la performance) dont l'équipe GULT recommande la lecture sont précédées d'un astérisque: *.

*Brown, J. D., Hudson, T., Norris, J. et Bonk, W.J. (2002). *An investigation of second language task-based performance assessments*. Technical report No. 24, Honolulu: University of Hawaii Press.

Cain, P. et Heath, S. (éd.) (2004). *Handbook for lecturers: Design, development and delivery of web-based case study resources Supporting student centred learning*. Newcastle: University of Newcastle.

Caré, J.-M. et Debyser, F. (1995). *Simulations globales*. Sèvres: CIEP.

Casey, E. et Fischer, J. (2004). « Improving interaction between subject area specialists and language teachers at UCARS (Universities and Colleges of agricultural sciences) ». In: Satchell, R. et Chenik, N. (éd.). *University language centres: Forging the learning environments of the future. Papers from 7th CercleS Conference. Paris, 19-21 September 2002*. Paris / Dublin: CercleS, 431-443.

Casey, E. et Fischer, J. (2005). « The use of case studies in language teaching ». In: Moravčíková, A., Taylor Torsello, C. et Vogel, T. (éd.). *University language centres: Broadening horizons, expanding networks. Proceedings of the 8th CercleS Conference. Bratislava, 9-11 September 2004*. Bratislava: CercleS, 163-175.

*Chalhoub-Deville, M. (2001). « Task based assessments: Characteristics and validity evidence ». In: Bygate, M., Skehan, P. et Swain M. (éd.). *Researching pedagogic tasks: Second language learning, teaching and testing*. Harlow: Longman.

Clapham, C. et Corson, D. (éd.). *Encyclopedia of language and education*. Vol. 7: Language testing and assessment. Dordrecht: Kluwer Academic.

CLES, site internet <http://www.certification-cles.fr/>.

Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Strasbourg: Conseil de l'Europe / Paris: Didier.

Council of Europe (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). A manual*. Strasbourg: Council of Europe, Language Policy Division; disponible sur le site internet: http://www.coe.int/t/dg4/linguistic/Manuel1_EN.asp.

Daly, P. (2002). « Methodology for using case studies in the business English language classroom ». In: *The Internet TESL Journal – For teachers of English as a second language*, Vol. VIII, No. 11. <http://iteslj.org/Techniques/Daly-CaseStudies/> [publication en ligne].

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. et McNamara, T. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.

Debyser, F. (1996). *L'immeuble*. Paris: Hachette.

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.

Eggensperger, K.-H. et Fischer, J. (éd.) (1998). *Handbuch UNICERT®*. Bochum: AKS-Verlag.

Elder, C. (1993). «How do subject specialists construe classroom language proficiency ». In: *Language Testing* 10(3), 235-254.

*Elder, C. et Brown, A. (1997). « Performance testing for the professions: Language proficiency or strategic competence? ». In: *Melbourne Papers in Language Testing* 6(1), 68-78.

Elder, C., Iwashita, N. et McNamara, T. (2002). « Estimating the difficulty of oral proficiency tasks: What does the test-taker have to offer? ». In: *Language Testing* 19(4), 47-368.

*Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

EXPLICS, site internet: www.zess.uni-goettingen.de/explics.

*Finch, A. (2002). « Authentic assessment: Implications for EFL performance testing in Korea ». In: *Secondary Education Research* 49, 89-122. Consultable sur le site: http://www.finchpark.com/arts/Authentic_Assessment_Implications.pdf.

Fischer, J. (2003). « AFANet – HERMES: A bridge between scientists and language teachers: Goals and perspectives ». In: *Języki obce: narzędzie komunikacji naukowej. Ogólnopolska Konferencja Naukowa z okazji 50-lecia Studium Języków Obcych Akademii Rolniczej im. Augusta Cieszkowskiego w Poznaniu*. Poznań: Wydawnictwo Akademii Rolniczej, 16-30.

Fischer, J. (2004). « HERMES – The AFANet Language Policy Group: Strengthening language teaching for students of agriculture and related sciences ». In: Heath, S. (éd.). *Capitalising on innovation in the curriculum in European higher education. Practice and promise in the disciplines of agriculture, forestry, aquaculture and environmental sciences. Proceedings of the AFANet – ICA Conference held at the University of Gent, Belgium, 8 to 10 January 2004*. Nancy: Institut national polytechnique de Lorraine (INPL), 113-131.

Fischer, J. et Casey, E. (2005). « Der Einsatz von Fallstudien im Fremdsprachenunterricht an Hochschulen ». In: Gebert, D. (éd.). *Innovation aus Tradition. Dokumentation der 23. Arbeitstagung 2004*. Bochum: AKS-Verlag, 175-184.

Fischer, J., Casey, E., Abrantes, A. M., Gigl, E. et Lešnik, M. (éd.) (2008). *LCaS: Language Case Studies. Teacher training modules on the use of case studies in*

language teaching at secondary and university level. A handbook. Strasbourg: Council of Europe.

Fischer, J., Musacchio, M. T. et Standing, A. (éd.) (2009). *EXPLICS – Exploiting Internet case studies and simulation templates for language teaching and learning. A handbook*. Göttingen: Cuvillier Verlag.

Foster, P. et Skehan, P. (1996). « The influence of planning and task type on second language performance ». In: *Studies in Second Language Acquisition* 18, 299-323.

*Foster, P. et Skehan, P. (1999). « The influence of source of planning and focus of planning on task-based performance ». In: *Language Teaching Research* 3, 299-324.

Gonzalez Pino, B. (1998). « Prochievement testing of speaking: Matching instructor expectations, learner proficiency level, and task type ». In: *Texas Papers in Foreign Language Education* 3.3, 119-133.

HERMES, site internet: www.uni-tuebingen.de/hermes/.

Hill, B. C. et Ruptic, C. (1994). *Practical aspects of authentic assessment: Putting the pieces together*. Norwood, MA: Christopher-Gordon Publishers.

Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.

Kaiser, F. J. (éd.) (1983). *Die Fallstudien. Theorie und Praxis der Fallstudiendidaktik*. Bad Heilbrunn: Klinkhardt.

Kiefer, K.-H. (2004). « Fallstudien – zum Umgang mit einer erfolgreichen Lernmethode im Fachsprachenunterricht Wirtschaftsdeutsch ». In: *Info DaF* 1, 68-98.

Kohonen, V. (1997). « Authentic assessment as an integration of language learning, teaching, evaluation and the teacher's professional growth ». In: Huhta, A., Kohonen, V., Kurki-Suonio, L. et Luoma, S. (éd.), *Current developments and alternatives in language assessment: Proceedings of LTRC 1996*. Jyväskylä: University of Jyväskylä, 7-22.

Kohonen, V. (1999). « Authentic assessment in affective foreign language education ». In: Arnold, J. (éd.) *Affect in language learning*. Cambridge: Cambridge University Press.

LCaS, site internet: <http://www.ecml.at/mtp2/LCaS>.

Lewkowicz, J. (2000). « Authenticity in language testing ». In: *Language Testing* 17(1), 43-64.

Manual for language test development and examining. For use with the CEFR. Produced by the Members of ALTE on behalf of the Language Policy Division, Council of Europe. (2011). Strasbourg: Council of Europe (Language Policy Division),

disponible sur le site: http://www.coe.int/t/dg4/linguistic/ManualLanguageTest-Alte2011_EN.pdf

McNamara, T. (1996). *Measuring second language performance*. London: Longman.

McNamara, T. (2000). *Language testing*. Oxford et al.: Oxford University Press.

*Mislevy, R. J., Steinberg, L. S. et Almond, R. G. (2002). « Design and analysis in task-based language assessment ». In: *Language Testing* 19(4), 477-496.

Morton, J., Wigglesworth, G. et Williams, D. (1997). « Approaches to validation: Evaluating interviewer performance in oral interaction tests ». In: Brindley, G. et Wigglesworth, G. (éd.). *Access: Issues in language test design and delivery*. Sydney: NCELTR, 175-196.

*Norris, J. (2002). « Interpretations, intended uses and designs in task-based language assessment ». In: *Language Testing* 19 (4), 337-346.

Norris, J. M., Brown, J. D., Hudson, T. et Yoshioka, J. (1998). *Designing second language performance assessments*. Honolulu: University of Hawaii Press.

Norris, J. M., Brown, T. D. et Bonk, W. (2002). « Examinee abilities and task difficulty in task-based second language performance assessment ». In: *Language Testing* 19(4), 395-418.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.

O'Malley, J. M. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. Reading, Mass.: Addison-Wesley Pub. Co.

*Petermann, A. (2008). « CLES – Certificat de compétences en langues de l'enseignement supérieur. Zertifizierung der Sprachkompetenzen im französischen Hochschulbereich ». In: *Fremdsprachen und Hochschule* 79/80, 33-54.

Puren C. (2002). « Perspectives actionnelles et perspectives culturelles en didactique des langues: vers une perspective co-actionnelle co-culturelle ». In: *Les langues modernes* 3, 55-71.

Puren C. (2006a). « De l'approche communicative à la perspective actionnelle ». In: *Le Français dans le Monde* 347, 37-40.

Puren C. (2006b). « Les tâches dans la logique actionnelle ». In: *Le Français dans le Monde* 347, 80-81.

Puren, C. (2008). « Formes pratiques de combinaison entre perspective actionnelle et approche communicative: analyse comparative de trois manuels ». In: *Les Langues Modernes*, 5 janvier 2008, 1-19, consultable sur le site: <http://www.aplv-languesmodernes.org/spip.php?article1409#>.

- Ribé, R. et Vidal, N. (1993). *Project work*. Oxford: Heinemann.
- Schart, M. (2003). *Projektunterricht – subjektiv betrachtet. Eine qualitative Studie mit Lehrenden für Deutsch als Fremdsprache*. Hohengehren: Schneider Verlag.
- Skehan, P. (1996). « Second language acquisition research and task-based instruction ». In: Willis, J. et Willis, D. (éd.). *Challenge and change in language teaching*, Oxford: Heinemann, 17-30.
- Skehan, P. (2001). « Tasks and language performance ». In: Bygate, M., Skehan, P. et Swain, M. (éd.). *Researching pedagogic tasks: Second language learning, teaching and testing*. Harlow: Longman, 167-187.
- Skehan, P. (2003). « Task-based instruction ». In: *Language Teaching* 36, 1-14.
- Spence-Brown, R. (2001). « The eye of the beholder: Authenticity in an embedded assessment task ». In: *Language Testing* 18(4), 463-481.
- *Standring, A. (2009). « Language assessment ». In: Fischer, J., Musacchio, M. T. et Standring, A. (éd.). *EXPLICS – Exploiting Internet case studies and simulation templates for language teaching and learning. A Handbook*. Göttingen: Cuvillier Verlag, 45-65.
- Teasdale, A. (1996). « Authenticity, validity, and task design for tests of well-defined LSP domains ». In: *Thames Valley University Working Papers in English Language Teaching* 2, 137-150.
- Uber Grosse, C. (1988). « The case study approach to teaching business English ». In: *English for Specific Purposes* 7, 131-136.
- UNICert[®], site internet: www.unicert-online.org.
- UNICert[®] LUCE, site internet: <http://www.casajc.sk/unicert.htm>.
- Van den Branden, K., Bygate, M. et Norris, J. M. (éd.) (2009). *Task-based language teaching. A reader*. Amsterdam / Philadelphia: Benjamins.
- Vaughn, C. (1991). « Holistic assessment: What goes on in the raters' minds? » In: Hamp-Lyons, L. (éd.). *Assessing second language writing in academic contexts*. Norwood, NJ: Ablex, 111-126.
- Voss, B. (2010). *Handbuch UNICert 2*. Bochum: AKS-Verlag.
- WebCase, site internet: <http://webcase-online.info>.
- Weideman, A. (2006). « Assessing academic literacy in a task-based approach ». In: *Language Matters* 37 (1), 81-101.
- Weir, C. J. (1990). *Communicative language testing*. Hemel Hempstead: Prentice Hall.

Weir, C. J. (2005). *Language testing and evaluation*. Basingstoke: Palgrave Macmillan.

Wesche, M.B. et Skehan, P. (2002). « Communicative, task-based, and content-based language instruction ». In: Kaplan, R. B. (éd.). *The Oxford handbook of applied linguistics*. Oxford: Oxford University Press, 207-228.

*Wigglesworth, G. (2001). « Influences on performance in task-based oral assessments ». In: Bygate, M., Skehan, P. et Swain, M. (éd.). *Researching pedagogic tasks: Second language learning, teaching and testing*. Harlow: Longman, 186-209.

*Wigglesworth, G. (2008). « Task and performance based assessment ». In: Shohamy, E. et Hornberger, N. H. (éd.). *Encyclopedia of language and education, 2nd Edition, Volume 7: Language Testing and Assessment*. 111-122.

*Wu, W. et Stansfield, C. (2001). « Toward authenticity of task in test development ». In: *Language Testing* 18(2), 187-206.

Yaiche, F. (1996). *Les simulations globales – mode d’emploi*. Paris: Hachette.

Pour des lectures complémentaires sur l’évaluation en langues en général, on pourra consulter également la liste détaillée disponible sous:

<http://languagetesting.info/video/reading.html>.

Deuxième partie: Ressources documentaires

Document 1: Tests et évaluation

Le texte qui suit apporte des informations générales sur l'évaluation en langues. Bien qu'il ne porte pas spécifiquement sur l'évaluation conçue dans une approche actionnelle, il fournit un contexte et des informations de base pour les lecteurs peu familiers de la littérature sur l'évaluation.

L'enseignement des langues étrangères en amont de l'université vise à fournir aux apprenants les outils nécessaires pour communiquer efficacement dans la langue cible dans la vie de tous les jours. Il s'agit de la langue générale (LG). Très peu de programmes d'enseignement des langues étrangères dépassent le niveau de difficulté de la langue utilisée dans la presse écrite. En d'autres termes, avant l'université les apprenants en langue devraient être capables de discuter de sujets abordés dans la presse au niveau de la langue utilisée par celle-ci. Cependant ce niveau de maîtrise linguistique n'est pas suffisant pour faire face aux exigences universitaires en langues.

Les universités et autres institutions académiques mettent souvent en avant la responsabilité qui leur incombe de développer les compétences linguistiques de leurs étudiants afin qu'ils puissent poursuivre leurs études dans une langue étrangère et, une fois diplômés, exercer une activité professionnelle à l'échelle internationale, en utilisant la ou les langue/s étrangère/s dans leur emploi futur. De plus en plus souvent les étudiants valident une partie de leur cursus dans une langue étrangère, soit dans le cadre d'échanges internationaux, soit dans leurs institutions nationales, en suivant des enseignements dispensés dans une langue étrangère. Ceci s'ajoute au fait que la plupart des universités exigent de valider des langues étrangères dans les cursus, que leurs étudiants aient à utiliser des ressources en langues étrangères dans leurs études ou pas.

L'enseignement des langues étrangères à l'université diffère des programmes scolaires en ce qu'il met l'accent ou se concentre sur les langues à fins universitaires (*LAP: Language for Academic Purposes*) ou les langues à fins spécifiques (*LSP: Language for Specific Purposes*), appelées aussi « langues professionnelles ». Les enseignements de LAP couvrent les contenus en langue étrangère dont les étudiants ont besoin s'ils étudient dans une université / suivent un cursus utilisant cette langue étrangère. Ainsi, en complément d'outils linguistiques spécifiques à un domaine donné, les cours couvrent, par exemple, les fonctions de langage que les étudiants doivent maîtriser dans les relations avec une administration universitaire.

Les langues à fins spécifiques ou sur objectifs spécifiques diffèrent à la fois de la langue générale et de la LAP à plusieurs titres. En premier lieu, les utilisateurs de LSP peuvent être clairement définis en fonction de leurs activités, professions ou spécialisations universitaires. Les LSP elles-mêmes ont leurs propres caractéristiques lexicales, syntaxiques et stylistiques. Par exemple, des termes rares par ailleurs peuvent être utilisés fréquemment (par exemple « chaton », « irruption »), la voix passive peut

être beaucoup plus utilisée qu'en LG ou en LAP, et de longs groupes composés, par ailleurs stylistiquement maladroits, sont nécessaires pour atteindre une plus grande précision (« la recherche sur le vaccin contre le cancer de la peau »).

Il est prouvé que l'apprentissage d'une LSP est très motivant pour les étudiants à l'université, nombre d'entre eux envisageant de travailler dans un contexte international. De plus, les possibilités d'échanges réussis avec des universités étrangères augmentent lorsque les étudiants connaissent la langue utilisée par les professionnels étrangers dans leur domaine d'études.

L'enseignement des LSP dans les universités repose sur des recherches et analyses de besoins préalables. Les langues sur objectifs spécifiques étaient un domaine de recherche très apprécié dans les années 1980, et les LSP les plus populaires en Europe étaient alors, et sont peut-être encore aujourd'hui, l'anglais, le français et l'allemand. L'anglais des affaires, le français pour les étudiants en médecine, l'allemand pour les étudiants en biologie sont des enseignements typiques de LSP.

Lorsqu'on parle de l'importance d'un test de langue, on utilise les expressions « tests à enjeux élevés » et « tests à enjeux limités ». Un test à enjeux élevés est une épreuve importante dont le résultat entraîne des décisions majeures telles que l'admission dans une université. En général, les tests à enjeux élevés ne peuvent également être passés qu'un nombre limité de fois et ils peuvent être payants. Les tests à enjeux limités sont des épreuves mineures qui peuvent être repassées, ou s'il y a une série de tests, un résultat faible dans un test peut être compensé par de bons résultats dans un autre test. Les quizzes ou les petits tests de langue sont des exemples typiques de tests à enjeux limités.

Chaque test peut être situé dans un continuum allant des « tests de vitesse » (à une extrémité) à ce que l'on appelle des « tests de niveau » (à l'autre extrémité). Dans les tests de vitesse, c'est la rapidité avec laquelle le candidat est capable de faire le test qui est décisive. Dans un vrai test de niveau, les candidats peuvent disposer de tout le temps qui leur est nécessaire pour passer le test, et donc pour montrer leur niveau dans le domaine testé. Evidemment, l'enseignement des langues étrangères à l'université ne recourt à aucun de ces deux types à l'état pur, mais lorsqu'ils projettent des tests, les concepteurs doivent avoir une idée très claire du temps qui sera accordé pour chaque item et pour l'ensemble du test. Il faudra se baser sur une durée estimée moyenne: jamais tous les étudiants n'auront besoin de la même durée pour effectuer un test.

On utilise les expressions « tests de connaissances » et « tests de performance » en relation avec le contenu à tester. Les tests de connaissances montrent ce que les étudiants ont acquis des contenus linguistiques qui leur ont été enseignés. Pour ce faire, les tests de connaissances doivent porter presque exclusivement sur les contenus de cours et ce sont typiquement les tests terminaux utilisés dans les cours de langues. Au contraire, les tests de performance ne portent pas sur un corpus linguistique prédéfini

mais montrent la capacité du candidat à agir efficacement dans certaines situations. Les étudiants savent quoi étudier pour un test de connaissances, alors que les tests de performance peuvent couvrir toutes sortes de compétences, sans prendre en compte les études de langues antérieures. De nombreux tests de langues commerciaux à diffusion internationale sont des tests de performance.

Les épreuves de langues peuvent être utilisées pour prévoir la réussite future d'un étudiant dans des études de langues, et sont dans ce cas des « tests pronostiques ». Les résultats d'un « test diagnostique » montrent l'état des compétences linguistiques de l'étudiant au moment de l'épreuve. A l'université, le plus répandu de ces deux types est sans aucun doute le test diagnostique.

Dans la phase de conception d'une épreuve de langue, une des décisions importantes à prendre est de déterminer, entre autres, jusqu'à quel point le test doit mesurer les compétences langagières des personnes testées (les autres décisions portant sur la durée accordée pour la passation du test, la sélection des supports du test et les types d'items ou de tâches utilisés dans le test). Les « tests formatifs » se concentrent sur certains aspects de la langue étrangère tels que l'usage des articles, tandis que les « tests sommatifs » se composent d'une ou de plusieurs tâches permettant d'observer les compétences langagières dans leur globalité, par exemple une rédaction. On utilise également le concept de « test formatif » pour décrire un test qui n'est pas noté. Si le test est noté ou si le résultat est « réussite / échec », on communique les notes ou les niveaux aux étudiants à titre d'information, afin qu'ils sachent ce qu'ils doivent améliorer. Dans l'évaluation formative, les tests formatifs sont des outils pour l'enseignement.

Les « tests objectifs » et les « tests subjectifs » devraient en réalité être appelés « tests à barème objectif » et « tests à barème subjectif ». Il n'existe pas de tests de langues vraiment objectifs, car chaque épreuve repose sur la conception que l'auteur du test a de ce qui est important dans une langue et des compétences requises pour réussir une épreuve de langue. Les tests à barème objectif sont typiquement les tests à choix multiples ou les tests vrai-faux avec une grille de correction que chaque évaluateur utilise de la même manière. La condition préalable est qu'il n'y ait qu'une réponse juste par item. La personnalité de l'évaluateur ainsi que ses préférences personnelles jouent un rôle plus important dans les tests à barème subjectif parce qu'il n'y a pas de corrigé, et les consignes écrites peuvent souvent être interprétées de plusieurs manières. C'est pourquoi dans des épreuves à enjeux décisifs, les examinateurs suivent des sessions de formation avant de commencer à noter des tests sommatifs.

Etablir le barème signifie, avant tout, fixer le nombre de points par item, par tâche et par partie, et décider ce qui est exigé pour obtenir la note maximale. Le point le plus important dans un barème est le « point de césure », en d'autres termes, le point où passe la ligne entre ceux qui réussissent le test et ceux qui échouent. Le point de césure optimal ne repose pas seulement sur l'intuition du testeur mais aussi sur la recherche.

A l'université, de nombreux tests sont non pas notés sur une échelle, mais validés ou non validés (« tests réussite / échec »). Dans une épreuve de langue, le point de césure et les limites entre les notes peuvent être décidés sur la base d'une « évaluation normative » ou d'une « évaluation critériée ». Dans une évaluation normative, les notes sont déterminées par les résultats antérieurs de la population testée et sont attribuées aux candidats suivant le percentile dans lequel ils se situent. Par exemple, nous savons déjà avant même de noter le premier test quel pourcentage des candidats obtiendra la meilleure note ou échouera. Cela signifie que le score brut utilisé comme point de césure peut différer d'un test à l'autre, mais pas au point de rendre le test non fiable. Les tests normatifs ne sont pas appropriés pour évaluer des petits effectifs.

L'évaluation (à référence) critériée signifie que le candidat doit satisfaire à certains critères afin de réussir le test ou d'obtenir une certaine note. La plupart des tests de langue à l'université sont critériels, ce qui signifie que l'on sait au préalable ce qui est nécessaire pour obtenir le score maximum, mais on ne sait pas combien d'étudiants obtiendront telle ou telle note ou échoueront au test.

Dans le cas des examens de langues, on peut dire si le test est un « test à items discrets » (« test à éléments isolés ») ou un test « intégré ». Les tests à items discrets sont composés de parties qui ciblent des phénomènes linguistiques clairement identifiables et qui sont totalement séparées les unes des autres, avec leurs propres totaux de points, et le score est la somme des points donnés pour les différents items. Un test intégré peut tester plusieurs compétences à la fois ou la langue en général sans mettre l'accent sur, par exemple, la grammaire ou le vocabulaire. Un test de closure est un test intégré car, afin de compléter un blanc, le candidat doit comprendre les indices aussi bien contextuels que syntaxiques.

L'« évaluation directe » signifie qu'on testera par exemple l'oral en faisant parler les candidats, puis en évaluant leur production. L'« évaluation indirecte », *stricto sensu*, couvre toutes les formes d'évaluation des compétences réceptives, car on ne peut pas tester la compréhension de l'écrit ou de l'oral en tant que telles – dans un test de lecture ou d'écoute, l'enseignant doit recourir à la forme orale ou écrite pour établir ce qui a été compris. Un exemple plus concret d'évaluation indirecte: on donne la moitié d'un dialogue sous forme écrite au candidat et on lui demande de rédiger l'autre moitié. En d'autres termes, la compétence de production orale est testée ici à travers la production écrite.

Pendant près de mille ans, la principale méthode d'enseignement des langues étrangères a été la méthode dite « grammaire-traduction ». Evaluer signifiait principalement traduire des textes de et vers la langue cible. Au milieu du siècle dernier, la « méthode psychométrique-structuraliste » a commencé à gagner du terrain, et les « tests à choix multiples » se sont répandus. Cette méthode et ces types de tests ont été néanmoins sévèrement critiqués, car on sentait qu'ils étaient très loin de la réalité de l'utilisation des langues; on pensait que cocher des choix n'avait pas grand

chose à voir avec un usage véritable de la langue étrangère à des fins de communication. Une vingtaine d'années plus tard, l'ère de la « méthode communicative » marqua l'apparition des « tests communicatifs », dans lesquels les candidats communiquaient vraiment dans la langue cible, en transmettant et échangeant des informations pertinentes. La forme extrême de l'évaluation communicative réside dans les « tests authentiques », qui consisteraient à évaluer en tant que tests de performance des situations authentiques d'utilisation de la langue étrangère. Pour un tel test, on pourrait envoyer des étudiants dans une bibliothèque où ils doivent se procurer un certain livre en utilisant la langue cible; l'évaluateur filmerait les échanges entre les étudiants et le bibliothécaire, et évaluerait ensuite la production orale des étudiants.

Une langue est une entité si complexe qu'un test en langue étrangère ne peut porter que sur une partie de ce tout. Lorsqu'on teste les compétences langagières, on les répartit traditionnellement en deux domaines: les compétences « de réception » (écouter et lire) et les compétences « de production » (parler et écrire). Le terme « modalités », souvent utilisé dans la littérature sur l'évaluation, se réfère également à l'écoute, la parole, la lecture et l'écriture. *Le Cadre européen commun de référence pour les langues* divise les langues en 57 « compétences », qui, toutefois, ne doivent pas être testées une par une. Bien davantage, dans un test portant sur l'écrit par exemple, le testeur doit avoir à l'esprit quelles compétences font effectivement l'objet du test.

Tout test de langue devrait être « valide » et « fidèle ». On peut démontrer qu'un test est valide par une « validation a priori » ou une « validation a posteriori ». La validation a priori, qui a lieu après que le test a été conçu mais avant qu'il soit utilisé, repose sur l'étude et la pondération de théories antérieures sur l'évaluation, et sur l'utilisation de recherches antérieures portant sur des tests et des situations de tests similaires. La validation a posteriori, qui a lieu après que le test a été administré, est établie en utilisant toutes sortes de mesures et d'analyses.

Un test est valide s'il mesure ce qu'il est supposé mesurer et qu'il en convainc les parties prenantes. Il y a plusieurs sortes de validité: la « validité externe » dépend de la représentativité du test et de la possibilité d'en généraliser les résultats. La « validité interne » porte sur le contenu du test et, en particulier, sur le fait que le test lui-même ne contient pas d'éléments non pertinents qui dérangeraient le candidat. La « validité prédictive » signifie que le test peut être utilisé pour prédire la réussite future du candidat dans l'apprentissage de cette langue ou d'une autre. La « validité de contenu » montre si le test a mesuré ce qu'il était supposé mesurer, par exemple que le contenu d'un test de connaissances ne va pas au-delà de ce qui a été enseigné avant le test. On calcule la « validité concourante » lorsque le test a été comparé à d'autres tests à peu près du même type ou passés à la même période. La « validité de construit » (ou « validité conceptuelle ») est liée au cadre théorique qui sous-tend le test, c'est-à-dire si le test évalue vraiment un construit tel que la compréhension de l'écrit en LSP. La « validité apparente » (en anglais: *face validity*), appelée parfois en plaisantant en

anglais « *faith validity* » (« validité de la foi »), concerne l'acceptation du test par les utilisateurs comme objet de mesure de ce qu'il dit mesurer. Par exemple, certains étudiants estiment que la validité apparente des QCM pour évaluer la compréhension de l'écrit est quelque peu discutable parce qu'ils trouvent que les choix proposés sont trop restrictifs ou peu crédibles.

La « fidélité » en matière de tests signifie que si un test est administré à des populations similaires dans des situations d'évaluation similaires, il produira des résultats similaires. La « fidélité intra-examineurs » renvoie à l'aptitude d'un examinateur à noter les tests de manière systématique. On calcule la fidélité intra-examineurs dans le cas où le jugement de l'examineur peut avoir été affecté par la maladie ou l'épuisement. Même l'heure de la journée ou la quantité de travail fourni avant de noter peuvent influencer sur les résultats. On calcule la « fidélité inter-examineurs » quand plusieurs examinateurs notent le même test, les candidats ont en effet besoin de savoir que tous les examinateurs évaluent les tests de la même façon. Tout étudiant a le droit de savoir que la notation est juste.

Une fois le test administré et noté, il est temps de le réexaminer avec attention. Tout enseignant-testeur devrait procéder à une simple « analyse des items », qui consiste à calculer les « indices de difficulté et de discrimination ». La difficulté (symbolisée par la lettre « p » dans les tableaux) se calcule en divisant le nombre de réponses correctes par le nombre de candidats. L'indice de difficulté varie de 0 à 1. Plus la valeur est élevée, plus l'item est facile. Il est néanmoins usuel d'accepter des indices de difficulté plus élevés pour les premiers et les derniers items, afin que les candidats ne soient pas effrayés par le test au début et qu'ils le terminent en ayant le sentiment qu'ils étaient à la hauteur. Cependant, les concepteurs de tests doivent au préalable connaître assez bien la population cible pour avoir idée de ce que sera la difficulté des items du test.

Même si les items du test donnent lieu au nombre attendu de réponses correctes, les concepteurs du test doivent établir s'ils discriminent vraiment les bons des mauvais apprenants. En d'autres termes, si l'indice de difficulté prévu est 0,5 (= la moitié de la population testée a fourni la bonne réponse à l'item), l'item est de mauvaise qualité si ceux qui ne sont pas bons dans la langue ont la réponse juste tandis que les bons apprenants n'ont pas fourni la réponse correcte à cet item. On peut calculer l'indice de discrimination (« D » dans les tableaux) en utilisant une formule simple: la population testée est répartie en une bonne moitié et une moitié faible en fonction des « scores bruts » totaux (le total des points avant toute manipulation ou ajustement). Puis on soustrait le nombre de réponses correctes dans la mauvaise moitié du nombre de réponses correctes dans la bonne moitié. On divise le nombre de réponses correctes restant par le total de toutes les réponses correctes. L'indice de discrimination varie entre -1 (toutes les réponses correctes sont dans la moitié faible) et 1 (toutes les réponses correctes sont dans la bonne moitié).

La « transparence » en matière d'évaluation signifie que les tests sont décrits suffisamment en détail pour que tous ceux qui en utilisent les résultats sachent ce qui a été testé et comment, et ce que signifie le résultat, c'est-à-dire ce que le candidat ayant obtenu une certaine note ou un certain score est capable de faire dans la langue cible. C'est pourquoi les universités européennes se donnent beaucoup de mal pour harmoniser la terminologie utilisée dans les descripteurs et les critères de leurs tests, l'outil le plus important dorénavant étant le *Cadre européen commun de référence pour les langues*.

Document 2: Exemples de mise en œuvre de l'évaluation par les tâches dans des systèmes universitaires de certification du niveau de capacité: mise en situation réelle

La certification CLES

La certification française CLES (Certificat de compétences en Langues de l'Enseignement Supérieur) a été créée en 2000 par décret du Ministre de l'Education Nationale et de l'Enseignement Supérieur. Un groupe de travail composé d'enseignants du secteur LANSAD (LANGues pour Spécialistes d'Autres Disciplines) originaires de plusieurs universités a défini le format exact des épreuves pour les niveaux B1 (CLES 1), B2 (CLES 2) et C1 (CLES 3) du CECRL au terme d'une phase d'expérimentation. Ce groupe coordonne la conception des sujets³³ et l'administration des sessions organisées dans les différents centres de langues universitaires, et fournit les grilles d'évaluation.

Dans le contexte universitaire, une certification en langue doit permettre d'attester que les candidats disposent des compétences langagières et opérationnelles nécessaires pour:

- traiter de l'information: sélectionner, trier, hiérarchiser des données, identifier des points de vue et les croiser;
- s'approprier cette information: restituer des arguments, les mettre en relation, présenter une opinion, débattre, proposer une solution, parvenir à un consensus.

Pour vérifier la maîtrise de savoir-faire de cette nature, il faut placer l'étudiant dans une situation directement inspirée de la « vraie vie » et lui demander de réaliser une tâche complexe, nécessitant de mobiliser des compétences et savoir-faire multiples (langagiers, sociolinguistiques, pragmatiques etc.) en les combinant. Il est normal pour un étudiant de lire et / ou d'écouter pour accéder au savoir, de s'approprier des informations et de les réinvestir à l'écrit ou oralement. Dans le droit fil de la perspective actionnelle du CECRL, le choix d'organiser le test autour d'une macro-tâche unique s'est donc imposé.

Les sujets sont des scénarios d'évaluation: on fournit une mise en situation, un dossier documentaire, on définit des tâches de production à réaliser dans le cadre de ce

33 Les concepteurs de sujets, enseignants dans les centres de langues universitaires, et travaillant normalement en binômes, doivent respecter un cahier des charges révisé annuellement. Les sujets conçus dans un centre sont relus et testés dans deux autres centres avant d'être validés et mis à disposition dans une banque de sujets nationale.

scénario, les tâches réceptives intermédiaires permettent d'alimenter les productions écrite et orale.

CLES 2, niveau B2: thématique et langue générales

Les épreuves correspondent à la structure GULT décrite au paragraphe 5.3., puisque les quatre compétences sont évaluées. La phase d'étude de cas en binôme ou en groupe n'est cependant pas prévue. Le cahier des charges fixe les formats:

- écouter: deux à trois documents sonores authentiques d'une durée totale de 5 minutes, avec trois passages;
- lire: un dossier composé de trois à cinq textes authentiques d'une longueur totale de 9 000 signes environ;
- écrire: un compte rendu, un rapport (plus précisément décrit dans la mise en situation) de 250 à 300 mots.

Ces trois premières parties font l'objet d'une épreuve unique de 2 heures 45; les candidats gèrent librement leur temps une fois passées les 30 minutes réservées à la compréhension de l'oral. La dernière partie suit immédiatement:

- Interaction orale: entre deux ou trois étudiants, d'une durée de 10 à 15 minutes. Dans la continuité de la mise en situation initiale, les étudiants endossent les rôles qui leur sont attribués et tentent d'arriver à un compromis en réinvestissant les informations obtenues précédemment.

CLES 3, niveau C1: thématique liée au domaine d'études, langue à fins spécifiques

Ce niveau de capacité devrait concerner majoritairement des étudiants déjà avancés dans leur cursus universitaire (master et au-delà) qui devront montrer qu'ils sont capables de communiquer sur des sujets en rapport avec leur domaine d'études ou de recherche. C'est pourquoi il a paru plus réaliste de ne pas dissocier les activités réceptives et les activités productives, ainsi que cela a été évoqué au paragraphe 5.2. Seules les compétences de production sont donc explicitement évaluées. Cependant la tâche ne peut pas être réalisée si les compétences réceptives des candidats ne sont pas d'un niveau suffisant, implicitement elles sont donc également prises en compte.

Le dossier documentaire à exploiter se compose de:

- un à deux documents d'écoute authentiques d'une durée totale de 10 minutes environ;
- des documents écrits d'une longueur de 15 000 signes environ, pouvant inclure aussi des documents iconographiques, schémas, statistiques etc. issus en partie de publications à caractère scientifique.

Les candidats disposent de trois heures pour s'approprier les informations et préparer un support visuel (PowerPoint ou transparents) qui va accompagner leur présentation orale. Ils travaillent seuls et gèrent librement leur temps. Ils présentent ensuite leur sujet:

- production orale: 10 minutes de production orale en continu (le candidat soumet ses conclusions au jury), puis 10 minutes d'interaction avec le jury;
- production écrite: après son oral, le candidat dispose de 60 minutes pour rédiger une note de synthèse (ou un article, un abstract etc. selon la mise en situation) de 600 mots environ qui intègre également les points discutés avec le jury pendant la partie orale.

Le cahier des charges prévoit qu'un spécialiste du domaine valide le choix de la thématique et des documents et fasse partie du jury.

Evaluation

Ainsi que le démontrent les grilles reproduites dans le document 5, la compétence linguistique (lexicale et grammaticale) n'est pas la seule évaluée. L'accomplissement de la tâche et la capacité à communiquer sont des critères d'évaluation tout aussi importants que la correction de la langue.

Par ailleurs, pour passer l'examen et obtenir le certificat, il faut que chacune des compétences évaluées soit validée. Il n'y a pas de notes, le principe retenu est celui du validé / non validé. Pour valider les compétences réceptives (CLES 2), il faut avoir 60% au moins d'items justes. Pour valider les compétences de production, il faut valider TOUS les critères retenus dans les grilles. On considère en effet que tous ces critères sont également constitutifs du niveau de capacité B2 ou C1.

Utilisations de CLES

En tant que certification du niveau de capacité, le CLES est indépendant des enseignements assurés dans les centres de langue et pourrait en théorie être passé sans avoir au préalable suivi de cours. De même, il n'y a pas de cours de préparation au CLES: compte tenu du format du test, tout travail sur les quatre activités langagières visant à comprendre et à produire une langue authentique prépare aussi au CLES.

Cependant, le CLES peut aussi être utilisé comme évaluation de fin de cours ou de fin de cursus. A l'Université de Strasbourg par exemple, l'objectif fixé en fin de Licence aux étudiants de disciplines scientifiques est le niveau B2, c'est pourquoi tous ces étudiants passent le CLES 2 comme examen de fin de semestre 5. Cependant, il n'est pas obligatoire de réussir la certification pour obtenir le diplôme de Licence. Lorsque le CLES est utilisé comme examen, chaque partie est notée selon des modalités fixées localement, et on calcule la moyenne de ces notes, donc il peut y avoir compensation entre les compétences. Ceci permet d'obtenir une note satisfaisante à l'examen sans pour autant obtenir la certification (voir ci-dessus).

L'examen UNICert

UNICert en général³⁴

UNICert est un système de cours et d'examens de langues mis sur pied par des membres de l'association allemande des centres de langues universitaires (Arbeitskreis der Sprachenzentren – AKS) au début des années 1990. Ce système, initié par Bernd Voss et Christine Klein-Braley puis adopté par l'AKS, vise à promouvoir et faire évoluer l'enseignement / apprentissage des langues à l'université à tous les niveaux. C'est un cadre qui permet aux établissements membres d'élaborer leurs propres cours et examens en suivant des règlements et un cahier des charges communs. Les programmes d'enseignement et d'examens sont normalement accrédités par le Comité Scientifique pour une période de trois ans (ou cinq ans dans le cas d'une évaluation *in situ* complète du centre de langues).

UNICert distingue quatre niveaux UNICert: UNICert I est axé sur le niveau B1 du CECRL, UNICert II sur le niveau B2 du CECRL, UNICert III sur le niveau C1 du CECRL et UNICert IV sur le niveau C2 du CECRL. Un niveau d'entrée « UNICert

34 Pour des informations supplémentaires sur UNICert voir: B. Voss (2010). *Handbuch UNICert 2*. Bochum: AKS-Verlag; et le site UNICert: www.unicert-online.org.

Basis », introduit en 2011, en particulier pour des langues à la structure plus complexe, est axé sur le niveau A2 du CECRL. UNICert ciblant les compétences langagières spécifiques à l'université, le système se concentre cependant sur les niveaux de langues supérieurs, où il est plus probable d'être confronté à la langue académique.

Les établissements membres d'UNICert mettent sur pied pour leurs centres un programme de cours correspondant au règlement d'UNICert. Ces cours peuvent être optionnels dans les cursus de Licence et de Master ou faire partie intégrante des études, selon le contexte de l'établissement. Ces programmes d'enseignement et d'évaluation sont ensuite contrôlés par les bureaux UNICert de l'Université technique de Dresde et de l'Université de la Ruhr à Bochum, et par des membres du Comité Scientifique, qui accrédite ensuite ces programmes.

En 2011, plus de 50 établissements ont proposé des programmes d'enseignement et des examens UNICert dans environ 25 langues. Le nombre de certificats le plus élevé est décerné en anglais (en particulier aux niveaux II et III), suivi par l'espagnol (principalement aux niveaux I et II) et le français (niveaux I à III), même si ces dernières années le suédois d'une part et les langues asiatiques ainsi que l'arabe et le turc d'autre part ont enregistré une augmentation importante.

En 2003 un système de franchise a été mis en place en République slovaque et en République tchèque sous le nom d'UNICert[®]LUCE; ses bureaux se trouvent à l'Université d'économie de Bratislava. Depuis, plus de 60 programmes d'enseignement et d'examens ont été accrédités dans onze établissements dans les deux pays.³⁵

Le Comité d'UNICert a récemment adopté l'approche actionnelle dans ses examens, ainsi que décrite dans le guide GULT, et il encourage ses membres à introduire des approches par tâches dans l'enseignement et l'évaluation, en particulier aux niveaux supérieurs.

UNICert à l'Université de Göttingen

L'approche actionnelle a été introduite au Centre de langues et de compétences transversales (Zentrum für Sprachen und Schlüsselqualifikationen – ZESS) de l'Université de Göttingen pour l'examen UNICert III (correspondant au niveau C1 du CECRL), qui est une certification du niveau de capacité (à enjeux élevés) à la fin d'une série de cours conduisant les apprenants au niveau C1. Certes il n'est pas possible de faire passer dans chaque cours un examen aussi complexe et long, mais on évalue chaque compétence séparément pendant et à la fin de la formation. Si les compétences

35 Informations complémentaires sur UNICert[®]LUCE sur le site Internet: <http://www.casajc.sk/unicert.htm>.

d'écoute et de lecture sont testées de manière « traditionnelle » lors d'un examen, pour les compétences de production on demande aux apprenants de réaliser des tâches le plus authentiques possible inspirées de la vraie vie, et de constituer une sorte de portfolio, en rédigeant par exemple un procès-verbal, en négociant un contrat, en écrivant un compte rendu, en soumettant une solution dans une étude de cas, sous forme orale et / ou écrite. Ainsi les apprenants se familiarisent avec l'approche par tâches de l'évaluation en langues et ont déjà acquis une bonne compétence dans le travail sur projet et les études de cas quand ils passent l'examen UNICert III.

En d'autres termes, lorsqu'un établissement introduit l'approche par tâches dans son évaluation en langues, il n'est pas obligé de faire « tout tout de suite » en mettant en place un examen basé sur une tâche qui porte sur les quatre compétences, ainsi que décrit plus haut; au contraire, il peut introduire certains éléments à différentes étapes du programme d'enseignement / apprentissage et de l'évaluation.

Pour ce qui est de l'examen UNICert, il porte sur les quatre compétences. La structure d'un examen UNICert, en l'occurrence un examen UNICert III (correspondant au niveau C1 du CECRL) à l'Université de Göttingen, est décrite ci-dessus au paragraphe 5.3. « Durée ».

Document 3: Exemples de tests et d'examens par tâches³⁶

Test 1: CLES 2 Espagnol – Niveau B2

Comme le montre l'exemple disponible sur le site web (<http://gult.ecml.at>), les tâches intermédiaires dans le CLES fournissent des informations qui doivent être réutilisées pour réaliser les macro-tâches, à savoir la production écrite et la production orale qui repose sur une interaction entre deux candidats.

De ce point de vue et de par l'utilisation de scénarios qui cherchent à reproduire des activités de la vraie vie, le CLES est fondé sur une approche visant à évaluer la performance par les tâches, dans laquelle « la simulation d'activités de la vie réelle [sont] associées à des caractéristiques situationnelles et interactionnelles, où la communication joue un rôle central ». [Original en anglais: « the simulation of real-world tasks, and associated situational and interactional characteristics, wherein communication plays a central role (see Bachman & Palmer, 1996) »] (Brown *et al.*, 2002: 10).

Comme évoqué précédemment, les compétences réceptives ne sont pas évaluées à part, car CLES 3 suit une approche holistique de l'évaluation: comme dans des situations de la vraie vie, les candidats doivent transmettre, commenter et évaluer les informations pertinentes tirées des documents fournis. Les compétences réceptives sont donc nécessaires pour être capable d'accomplir les tâches ciblant les compétences productives.

Test 2: CLES 3 Allemand – Niveau C1

Voir ci-dessus la description du test: Test 1: « CLES 2 Espagnol – Niveau B2 ».

Test 3: UNICert III Anglais des affaires – Niveau C1

L'examen UNICert présenté ici a été conçu à l'Université de Göttingen et donné à des étudiants qui avaient suivi un programme d'enseignement d'environ 120 heures en présentiel (plus 240 heures d'apprentissage en dehors du cours) en anglais des affaires. Dans cet examen, les candidats sont mis dans une situation authentique qu'ils

36 Des échantillons d'épreuves sont disponibles sur le site internet: <http://gult.ecml.at>.

pourraient rencontrer dans la vraie vie (un stage dans une entreprise internationale) et doivent élaborer une solution à un problème en petits groupes après avoir analysé en détail une quantité importante de documents. Puis ils doivent soumettre un document écrit (un compte rendu) et présenter leur cas à l'oral (présentation de leur solution, suivie d'une discussion détaillée de leur approche).

Chacune des quatre compétences est ensuite testée et évaluée séparément. Pour l'évaluation des compétences productives, on utilise les grilles UNICert disponibles dans le Document 5.

Test 4: UNICert III Français – Niveau C1

Cet examen UNICert de Français, conçu par Claudie Bréhinier au ZESS de l'Université de Göttingen, suit la même approche que le test 3 (anglais des affaires), mais il a été donné à la fin d'un programme d'enseignement UNICert III en français universitaire à des étudiants de différentes disciplines universitaires.

Test 5: Examen GULT Anglais – Niveau B2

Cet exemple d'examen a été conçu par le groupe de travail en langue anglaise lors de la réunion de réseau GULT, qui a eu lieu en janvier 2011 au Centre européen pour les langues vivantes à Graz, et il s'appuie sur une première ébauche élaborée pendant l'atelier GULT qui a eu lieu au CELV en mai 2010. Il s'adresse à des étudiants de niveau B2 du *Cadre européen commun de référence pour les langues* (CECRL). Lors de la réunion de réseau, les différentes équipes ont élaboré des variantes de cet examen en ciblant soit des étudiants en sciences environnementales, soit des étudiants issus de diverses disciplines universitaires.

Test 6: Examen GULT Français – Niveau B2

Cet exemple d'examen a été conçu par le groupe de travail en langue française lors de la réunion de réseau GULT, qui a eu lieu en janvier 2011 au Centre européen pour les langues vivantes à Graz. Il est destiné à des étudiants en ingénierie de niveau B2 du CECRL.

Références

Bachman, L. et Palmer, A. (1996): *Language testing in practice*. Oxford: Oxford University Press.

Brown, J.D., Hudson, T, Norris, J. et Bonk, W. J. (2002): *An investigation of second language task-based performance assessments*. Honolulu: University of Hawaii Press.

Document 4: Exemples de mise en œuvre de l'évaluation par les tâches dans l'enseignement des langues à l'université

Un exemple d'apprentissage et d'évaluation en langue en Master à l'Université de Strasbourg

Plusieurs facteurs ont conduit les enseignants à repenser depuis quelques années les modalités de travail et d'évaluation de leurs étudiants inscrits en Master:

- le volume horaire réduit consacré à l'encadrement de ces groupes (16 heures par semestre);
- la volonté d'armer ces étudiants pour leur insertion professionnelle après le Master;
- le souhait de travailler en collaboration avec les enseignants de spécialité;
- la volonté, dans la continuité de la pratique initiée en Licence, de rendre les étudiants acteurs de leur apprentissage en langues, et non de leur fournir des cours « clés en main »;
- l'influence de la certification CLES, et donc le désir d'organiser le travail de l'étudiant autour d'une « tâche » à mener à bien pendant le semestre.

Une collaboration très fructueuse a été mise sur pied avec des enseignants de Sciences de la Terre et d'Ingénierie Environnementale, aboutissant au schéma de fonctionnement suivant:

- En début de semestre, la tâche est expliquée aux étudiants. Il s'agit de réaliser une présentation orale sur une thématique de recherche directement liée à la spécialité étudiée, en exploitant au moins deux publications (de vulgarisation scientifique en première année de Master – M1, scientifiques en seconde année – M2) en langue étrangère (généralement en anglais). Les étudiants de M1 peuvent réaliser cette tâche en binôme; en M2, c'est un travail individuel.
- Les étudiants choisissent eux-mêmes la thématique et les sources, mais celles-ci doivent être validées par l'enseignant de spécialité.
- Les étudiants rendent compte régulièrement à l'enseignant de langue et au groupe de l'avancement de leur travail, ce qui permet d'exercer les compétences orales (parler en continu et interagir), de travailler les aspects phonologiques et lexicaux, de remédier le cas échéant aux difficultés rencontrées.

- Une fois les documents suffisamment exploités, les étudiants rédigent un abstract de leur future présentation qui est retravaillé avec l'enseignant (travail sur la compétence Ecrire), mais non évalué.
- Ils conçoivent leur présentation PowerPoint qui est également revue avec l'enseignant de langue (travail sur la méthodologie et la langue).
- En fin de semestre, ils réalisent leur présentation devant le groupe et les deux enseignants qui évaluent conjointement la prestation. La présentation est suivie de questions des enseignants et / ou du groupe.

La complexité de la tâche à réaliser et le type de sources à utiliser correspondent sans aucun doute au niveau C1, voire au-delà, du *Cadre européen commun de référence pour les langues* (CECRL). L'évaluation tient compte de l'aptitude à communiquer (compétences opérationnelles et pragmatiques), de la qualité de la langue, mais aussi de la qualité de la compréhension des documents et donc de la justesse du contenu scientifique restitué. L'étudiant agit dans sa vraie vie d'étudiant, dans la réalité de son domaine de spécialité; la tâche, inspirée des pratiques de la communauté scientifique, est non seulement réaliste, mais bien réelle pour lui.

Document 5: Grilles d'évaluation

A. Grilles existantes utilisées dans les contextes CLES et UNiCert

Grilles d'évaluation du niveau B2 du CECRL

Production orale

CERTIFICATION NATIONALE CLES

Grille d'évaluation de l'interaction orale CLES 2

N° d'anonymat:

Critères	Niveau B2	Validé	Non validé
LINGUISTIQUES PRAGMATIQUES	1. Situation Joue de manière adéquate le rôle qui lui a été attribué dans le cadre de la situation donnée.	<input type="checkbox"/>	<input type="checkbox"/>
	2. Contenu Utilise des arguments variés et pertinents issus des documents en ajoutant éventuellement des idées personnelles.	<input type="checkbox"/>	<input type="checkbox"/>
	3. Interaction Sait interagir: prend son tour et l'initiative de la parole quand il convient, sait relancer l'échange si nécessaire.	<input type="checkbox"/>	<input type="checkbox"/>
	4. Aisance Exprime ses idées avec fluidité sans faire de longues pauses (hésitations tolérées).	<input type="checkbox"/>	<input type="checkbox"/>
	5. Phonologie A une prononciation et une intonation suffisamment claires pour être compris(e).	<input type="checkbox"/>	<input type="checkbox"/>

	6. Correction grammaticale Maîtrise suffisamment la grammaire pour ne pas commettre d'erreurs conduisant à des malentendus (erreurs non systématiques tolérées).	<input type="checkbox"/>	<input type="checkbox"/>
	7. Cohérence Dispose d'outils linguistiques pour lier, nuancer et adapter son discours.	<input type="checkbox"/>	<input type="checkbox"/>
	8. Lexique Utilise un lexique juste et varié (quelques lacunes sont tolérées).	<input type="checkbox"/>	<input type="checkbox"/>
RÉSULTAT: B2 est validé <u>si et seulement si</u> chacun des critères est validé			
OUI <input type="checkbox"/>		NON <input type="checkbox"/>	

Grille 1: Grille d'évaluation de l'oral CLES niveau B2 du CECRL

Grille d'évaluation (B2.1, B2.2)

Langue / Cours: _____	Nom: _____
Semestre: _____	Prénom: _____
Enseignant/e: _____	N° d'étudiant: _____
Sujet: _____	

Tâche orale:

Date: _____

Il / Elle est en mesure de s'exprimer à l'oral, de manière appropriée, sur de nombreux sujets d'ordre culturel ou relevant d'un domaine de spécialité, et de participer activement à des discussions en utilisant des structures de phrases complexes et du vocabulaire spécialisé.

Par rapport au niveau visé:	très bien maîtrisé	bien maîtrisé	satisfaisant	encore suffisant	insuffisant
Elaboration de la présentation / Compétence pragmatique:	1	2	3	4	5
Gestion de la tâche (richesse du texte): <ul style="list-style-type: none"> - Compréhension et gestion correcte du problème / de la tâche. - Capacité à traiter la problématique. - Présentation claire et convaincante des contenus. - Présentation nuancée et convaincante du point de vue / de la perspective de l'étudiant/e - Capacité à mobiliser des connaissances générales sur le thème choisi. 					
Justesse des contenus: <ul style="list-style-type: none"> - Le contenu correspond au sujet. - Les contenus sont pertinents et correctement présentés. - Certains aspects sont présentés ou analysés de manière critique si nécessaire. - La solution proposée est plausible et convaincante. - Les exemples sont bien choisis. 					
Logique et cohérence de la présentation: <ul style="list-style-type: none"> - Présentation clairement structurée. - Bonne introduction. - Bonne structuration et articulation des arguments. - Les points essentiels et les transitions sont clairement indiqués. - La conclusion ou transition finale est bien formulée et mise en valeur. - L'étudiant/e a pu introduire une discussion.. 					
Evaluation de l'élaboration de la présentation / Compétence pragmatique:					

Discussion et prise de position argumentative / opinions exprimées:	1	2	3	4	5
Capacité à mettre en question les propos d'autrui: <ul style="list-style-type: none"> - L'étudiant/e met en question les propos tenus lors de la présentation. - Ses questions sont claires et précises. 					
Capacité à défendre sa position: <ul style="list-style-type: none"> - L'étudiant/e réagit de manière adéquate aux propos des interlocuteurs et fournit des réponses claires et convaincantes. - Il / Elle peut défendre ou préciser sa position et lever les malentendus. - Il / Elle peut relancer un débat et l'élargir. 					
Evaluation de la capacité à débattre:					

Par rapport au niveau visé:	très bien maîtrisé	bien maîtrisé	satisfaisant	encore suffisant	insuffisant
Compétence linguistique	1	2	3	4	5
Phonétique / Prosodie: <ul style="list-style-type: none"> - Présentation fluide. - Prononciation correcte. - Intonation et accentuation correctes. - Bonne intensité vocale. 					
Grammaire / Morphosyntaxe: <ul style="list-style-type: none"> - Morphosyntaxe et syntaxe correctes. - Utilisation correcte des constructions en fonction du contexte. - Souplesse grammaticale. - Maintien de la correction grammaticale lors de réponses spontanées aux questions et commentaires. 					
Vocabulaire: <ul style="list-style-type: none"> - Usage convaincant d'un vocabulaire étendu. - Usage convaincant d'expressions variées du domaine de spécialité. - Usage correct de la terminologie du domaine de spécialité. 					
Aisance: <ul style="list-style-type: none"> - Expression spontanée complexe dans la présentation et les interventions (moyens d'expression variés et nuancés). - Utilisation maîtrisée de modalisateurs et connecteurs. - Reformulations appropriées en cas de lacunes lexicales et de faux départs. - Utilisation consciente et maîtrisée des pauses pour réfléchir, mettre en relief, 					
Registre / Style: <ul style="list-style-type: none"> - Présentation et interventions lors de la discussion correspondent au registre linguistique des situations professionnelles ou universitaires correspondantes. - Style adapté à la situation professionnelle ou universitaire correspondante. 					
Evaluation des aspects linguistiques:					

Aspects paralinguistiques et techniques de présentation:	1	2	3	4	5
Gestuelle / Mimique: <ul style="list-style-type: none"> - Gestuelle et mimique adaptées à la langue cible. - Gestuelle et mimique naturelles et convaincantes. - Gestuelle et mimique soulignent les propos. 					
Maintien: <ul style="list-style-type: none"> - Maintien adapté à la situation. - Contenu mis en valeur le cas échéant par le maintien. 					
Techniques de présentation: <ul style="list-style-type: none"> - Parle sans notes. - Contact visuel avec le public qui favorise l'interaction. 					
Utilisation d'outils: <ul style="list-style-type: none"> - Usage convaincant de PowerPoint. - Graphiques et tableaux illustrent les contenus. - Bon usage d'outils complémentaires. - Suscite l'attention des auditeurs par l'utilisation des outils. 					
Evaluation des aspects paralinguistiques et techniques de présentation:					

	Calcul	Note
Elaboration de la présentation / Compétence pragmatique (30%)	_____ x 30%	
Discussion et prise de position argumentative / opinions exprimées (20%)	_____ x 20%	
Compétence linguistique (40%)	_____ x 40%	
Aspects paralinguistiques et techniques de présentation (10%)	_____ x 10%	
Note finale production orale (100%)		

Signatures: _____

Grille 2: Université de Göttingen (UNICert) – Grille d'évaluation de l'oral niveau B2 du CECRL

Production écrite

CERTIFICATION NATIONALE CLES
Grille d'évaluation de la production écrite CLES 2

N° d'anonymat:

Critères	Niveau B2	Validé	Non validé
PRAGMATIQUES	1. Accomplit la mission définie par la mise en situation. Respecte: • la forme (compte rendu), • le contenu (neutralité), • la longueur recommandée.	<input type="checkbox"/>	<input type="checkbox"/>
	2. Utilise dans l'ensemble un registre de langue approprié à la situation (en fonction des destinataires).	<input type="checkbox"/>	<input type="checkbox"/>
	3. Réinvestit une variété d'informations pertinentes issues des documents.	<input type="checkbox"/>	<input type="checkbox"/>
	4. Structure et articule son discours de façon cohérente.	<input type="checkbox"/>	<input type="checkbox"/>
LINGUISTIQUES	5. Maîtrise la syntaxe de la phrase simple et sait aussi utiliser des phrases complexes adaptées à la mission.	<input type="checkbox"/>	<input type="checkbox"/>
	6. Maîtrise suffisamment la grammaire pour permettre une lecture fluide du texte (erreurs non systématiques tolérées).	<input type="checkbox"/>	<input type="checkbox"/>
	7. Utilise un lexique pertinent et varié.	<input type="checkbox"/>	<input type="checkbox"/>
RÉSULTAT: B2 est validé si et seulement si chacun des critères est validé			
OUI <input type="checkbox"/>		NON <input type="checkbox"/>	

Grille 3: Grille d'évaluation de l'écrit CLES niveau B2 du CECRL

Grille d'évaluation (B2.1, B2.2)

Langue / Cours: _____	Nom: _____
Semestre: _____	Prénom: _____
Enseignant/e: _____	N° d'étudiant: _____
Sujet: _____	

Tâche écrite:

Date: _____

Il / Elle peut s'exprimer à l'écrit de manière appropriée sur un grand nombre de sujets d'ordre culturel ou technique en utilisant aussi jusqu'à un certain point des structures complexes et un vocabulaire spécialisé.

Par rapport au niveau visé:	très bien maîtrisé	bien maîtrisé	satisfaisant	encore suffisant	insuffisant
(Elaboration du) contenu des textes	1	2	3	4	5
Réalisation de la tâche (Richesse du texte): <ul style="list-style-type: none"> - Compréhension et réalisation adaptée de la tâche. - Traitement pertinent de la problématique. - Présentation claire et convaincante des contenus. - L'étudiant/e peut rédiger des textes clairs et détaillés sur un grand nombre de thèmes familiers. 					
Justesse des contenus: <ul style="list-style-type: none"> - Le contenu correspond au sujet. - Les informations sont restituées avec justesse. - Certains aspects sont présentés de manière critique si nécessaire. - L'étudiant/e peut analyser les causes, conséquences et hypothèses. 					
Pertinence des contenus: <ul style="list-style-type: none"> - Contenus pertinents et adaptés, illustrés par des exemples appropriés. 					
Originalité: <ul style="list-style-type: none"> - Mise en forme originale du contenu qui suscite l'intérêt du lecteur. 					
Evaluation (de l'élaboration) du contenu:					

Compétence pragmatique	1	2	3	4	5
Fonctionnalité: Effet sur le destinataire (incl. prise en compte du type de texte): <ul style="list-style-type: none"> - L'étudiant/e peut écrire différents types de textes. - L'étudiant/e peut écrire des courriers officiels, des courriels et des comptes rendus. - L'étudiant/e peut écrire des lettres, dans lesquelles il / elle décrit ses sentiments de façon nuancée, souligne l'importance d'événements et expériences vécus et commente les nouvelles données par son / sa partenaire et ses opinions. - L'étudiant/e peut présenter des nouvelles, des opinions et des sentiments et prendre position sur ceux / celles d'autres personnes. - L'étudiant/e peut rédiger un essai ou un compte rendu sur un sujet donné, en donnant des arguments pour ou contre un certain point de vue et en examinant les avantages et inconvénients de différentes solutions. - L'étudiant/e peut écrire une brève critique d'un livre ou d'un film. 					

<p>Organisation / Cohérence / Cohésion / Présentation:</p> <ul style="list-style-type: none"> - Texte clairement structuré. - Bonne introduction. - Maîtrise de l'articulation du texte, de l'enchaînement des arguments et des transitions. - Les points principaux sont clairement formulés. - L'étudiant/e peut examiner et comparer différentes idées et solutions à un problème. - L'étudiant/e peut résumer des informations et arguments tirés de différentes sources. - L'étudiant/e peut rédiger dans un style accessible des descriptions claires, détaillées et cohérentes d'événements ou expériences réels ou fictifs en explicitant les liens entre différentes idées. - Bonne conclusion. - L'étudiant/e peut restituer des informations dans un essai ou un compte rendu ou présenter les arguments pour ou contre un certain point de vue. - Présentation et paragraphes sont corrects et participent à l'organisation du texte. 					
. Evaluation de la compétence pragmatique:					

Par rapport au niveau visé:	très bien maîtrisé	bien maîtrisé	satisfaisant	encore suffisant	insuffisant
Compétence linguistique	1	2	3	4	5
Correction linguistique:					
<p>Grammaire:</p> <ul style="list-style-type: none"> - Syntaxe correcte. - Utilisation correcte de constructions adaptées au contexte. 					
<p>Vocabulaire:</p> <ul style="list-style-type: none"> - Utilisation convaincante d'un vocabulaire étendu. - Utilisation pertinente de connecteurs pour clarifier les liens logiques. 					
<p>Orthographe / Ponctuation:</p> <ul style="list-style-type: none"> - Orthographe et ponctuation correctes. 					
Souplesse des moyens d'expression:					
<p>Grammaire:</p> <ul style="list-style-type: none"> - Grande souplesse grammaticale. 					

Par rapport au niveau visé:	très bien maîtrisé	bien maîtrisé	satisfaisant	encore suffisant	insuffisant
Compétence linguistique	1	2	3	4	5
Vocabulaire: <ul style="list-style-type: none"> - L'étudiant peut nuancer l'expression de points de vue, sentiments et expériences. - L'étudiant peut est capable de transmettre des nouvelles et des points de vue. - L'étudiant peut rédiger différents textes suivis sur des sujets en lien avec ses centres d'intérêt personnels, d'études ou professionnels. 					
Registre / Style: <ul style="list-style-type: none"> - L'étudiant peut résumer des textes sur des sujets concernant sa personne, ses études ou sa profession, et des informations factuelles tirées de différentes sources et médias. - L'étudiant peut rédiger des courriers officiels standard en utilisant le registre approprié et en respectant les usages. 					
Evaluation de la compétence linguistique:					

	Calcul	Note
Elaboration du contenu (25%)	_____ x 25%	
Compétence pragmatique (25%)	_____ x 25%	
Compétence linguistique (50%)	_____ x 50%	
Note finale production écrite (100%)		

Signatures: _____

Grille 4: Université de Göttingen (UNICert) – Grille d'évaluation de l'écrit niveau B2 du CECRL

Grilles d'évaluation niveau C 1 du CECRL

Production orale

CERTIFICATION NATIONALE CLES

Grille d'évaluation de l'oral CLES 3

N° d'anonymat:

Niveau C1	Production		Interaction	
	validé	non validé	validé	non validé
1. Réalise la tâche demandée en utilisant un ou des supports adéquats	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Prend en compte les questions du jury et réagit de manière pertinente dans l'échange	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Gère correctement le temps imparti	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Restitue les informations et arguments issus des documents avec justesse et précision	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Structure son propos de manière claire et cohérente	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. S'exprime en se détachant de ses notes, regarde ses interlocuteurs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. S'exprime avec fluidité, aisance et spontanéité	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8. Utilise un registre de langue adapté à la tâche	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Maîtrise une syntaxe adaptée à l'argumentation orale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Maintient un niveau élevé de correction grammaticale (erreurs occasionnelles)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Utilise un lexique étendu, précis et approprié à la tâche	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. Fait preuve d'une maîtrise satisfaisante du système phonologique de la langue (prononciation, intonation, accentuation...)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
RÉSULTAT: C1 est validé <u>si et seulement si</u> chacun des critères est validé				
OUI <input type="checkbox"/>		NON <input type="checkbox"/>		

Grille 5: Grille d'évaluation CLES niveau C1 du CECRL

Grille d'évaluation (C1.1, C1.2)

Langue / Cours: _____	Nom: _____
Semestre: _____	Prénom: _____
Enseignant/e: _____	N° d'étudiant: _____
Sujet: _____	

Examen oral:

Date: _____

Il / Elle possède des connaissances et des compétences linguistiques de haut niveau relatives à des domaines scientifiques généraux et inhérents à la spécialité qui lui permettent de communiquer sur des sujets pointus en utilisant des moyens linguistiques variés. Il / Elle est en mesure, à l'oral, de s'exprimer aisément et couramment sur des sujets pointus et complexes de son domaine de spécialité, qui sont pertinents pour la vie universitaire ou professionnelle dans le cadre d'un séjour à l'étranger. Il / Elle est à même d'exposer son point de vue de manière synthétique et logique dans un style adéquat.

Par rapport au niveau visé:	très bien maîtrisé	bien maîtrisé	satisfaisant	encore suffisant	insuffisant
Elaboration de la présentation / Compétence pragmatique:	1	2	3	4	5
Gestion de la tâche (richesse du texte): <ul style="list-style-type: none"> - Compréhension et gestion correcte du problème / de la tâche. - Capacité à traiter la problématique. - Présentation claire et convaincante des contenus. - Présentation nuancée et convaincante du point de vue / de la perspective de l'étudiant/e - Capacité à mobiliser des connaissances générales sur le thème choisi. 					
Justesse des contenus: <ul style="list-style-type: none"> - Le contenu correspond au sujet. - Les contenus sont pertinents et correctement présentés. - Certains aspects sont présentés ou analysés de manière critique si nécessaire. - La solution proposée est plausible et convaincante. - Les exemples sont bien choisis. 					
Logique et cohérence de la présentation: <ul style="list-style-type: none"> - Présentation clairement structurée. - Bonne introduction. - Bonne structuration et articulation des arguments. - Les points essentiels et les transitions sont clairement indiqués. - La conclusion ou transition finale est bien formulée et mise en valeur. - L'étudiant/e a pu introduire une discussion.. 					
Evaluation de l'élaboration de la présentation / Compétence pragmatique:					

Discussion et prise de position argumentative / opinions exprimées:	1	2	3	4	5
Capacité à mettre en question les propos d'autrui: <ul style="list-style-type: none"> - L'étudiant/e met en question les propos tenus lors de la présentation. - Ses questions sont claires et précises. 					
Capacité à défendre sa position: <ul style="list-style-type: none"> - L'étudiant/e réagit de manière adéquate aux propos des interlocuteurs et fournit des réponses claires et convaincantes. - Il / Elle peut défendre ou préciser sa position et lever les malentendus. - Il / Elle peut relancer un débat et l'élargir. 					
Evaluation de la capacité à débattre:					

Par rapport au niveau visé:	très bien maîtrisé	bien maîtrisé	satisfaisant	encore suffisant	insuffisant
Compétence linguistique	1	2	3	4	5
Phonétique / Prosodie: <ul style="list-style-type: none"> - Présentation fluide. - Prononciation correcte. - Intonation et accentuation correctes. - Bonne intensité vocale. 					
Grammaire / Morphosyntaxe: <ul style="list-style-type: none"> - Morphosyntaxe et syntaxe correctes. - Utilisation correcte des constructions en fonction du contexte. - Souplesse grammaticale. - Maintien de la correction grammaticale lors de réponses spontanées aux questions et commentaires. 					
Vocabulaire: <ul style="list-style-type: none"> - Usage convaincant d'un vocabulaire étendu. - Usage convaincant d'expressions variées du domaine de spécialité. - Usage correct de la terminologie du domaine de spécialité. 					
Aisance: <ul style="list-style-type: none"> - Expression spontanée complexe dans la présentation et les interventions (moyens d'expression variés et nuancés). - Utilisation maîtrisée de modalisateurs et connecteurs. - Reformulations appropriées en cas de lacunes lexicales et de faux départs. - Utilisation consciente et maîtrisée des pauses pour réfléchir, mettre en relief, corriger. 					
Registre / Style: <ul style="list-style-type: none"> - Présentation et interventions lors de la discussion correspondent au registre linguistique des situations professionnelles ou universitaires correspondantes. - Style adapté à la situation professionnelle ou universitaire correspondante. 					

Par rapport au niveau visé:	très bien maîtrisé	bien maîtrisé	satisfaisant	encore suffisant	insuffisant
Compétence linguistique	1	2	3	4	5
Evaluation des aspects linguistiques:					

Aspects paralinguistiques et techniques de présentation:	1	2	3	4	5
Gestuelle / Mimique: <ul style="list-style-type: none"> - Gestuelle et mimique adaptées à la langue cible. - Gestuelle et mimique naturelles et convaincantes. - Gestuelle et mimique soulignent les propos. 					
Maintien: <ul style="list-style-type: none"> - Maintien adapté à la situation. - Contenu mis en valeur le cas échéant par le maintien. 					
Techniques de présentation: <ul style="list-style-type: none"> - Parle sans notes. - Contact visuel avec le public qui favorise l'interaction. 					
Utilisation d'outils: <ul style="list-style-type: none"> - Usage convaincant de PowerPoint. - Graphiques et tableaux illustrent les contenus. - Bon usage d'outils complémentaires. - Suscite l'attention des auditeurs par l'utilisation des outils. 					
Evaluation des aspects paralinguistiques et techniques de présentation:					

	Calcul	Note
Elaboration de la présentation / Compétence pragmatique (30%)	_____ x 30%	
Discussion et prise de position argumentative / opinions exprimées (20%)	_____ x 20%	
Compétence linguistique (40%)	_____ x 40%	
Aspects paralinguistiques et techniques de présentation (10%)	_____ x 10%	
Note finale production orale (100%)		

Signatures: _____

Grille 6: Université de Göttingen (UNiCert) – Grille d'évaluation de l'oral niveau C1 du CECRL

CERTIFICATION NATIONALE CLES

Grille d'évaluation de l'oral CLES 3

N° d'anonymat:

Niveau C1	validé	non validé
1. Effectue la tâche demandée dans le cadre du scénario	<input type="checkbox"/>	<input type="checkbox"/>
2. Respecte la longueur demandée (600 mots +/-10%)	<input type="checkbox"/>	<input type="checkbox"/>
3. Structure le discours de façon cohérente (usage pertinent de connecteurs et autres marqueurs d'articulation)	<input type="checkbox"/>	<input type="checkbox"/>
4. Réinvestit les informations pertinentes issues des documents	<input type="checkbox"/>	<input type="checkbox"/>
5. Intègre à son écrit les remarques faites lors de l'entretien avec le jury	<input type="checkbox"/>	<input type="checkbox"/>
6. Tient compte des points de vue et argumente afin de parvenir à une conclusion appropriée à la tâche	<input type="checkbox"/>	<input type="checkbox"/>

7. Utilise un registre de langue approprié à la tâche	<input type="checkbox"/>	<input type="checkbox"/>
8. Maîtrise les phrases complexes	<input type="checkbox"/>	<input type="checkbox"/>
9. Maintient un niveau élevé de correction (grammaticale, orthographique ...)	<input type="checkbox"/>	<input type="checkbox"/>
10. Utilise un lexique étendu et précis	<input type="checkbox"/>	<input type="checkbox"/>
RESULTAT: C1 est validé <u>si et seulement si</u> chacun des critères est validé		
OUI <input type="checkbox"/>	NON <input type="checkbox"/>	

Grille 7: Grille d'évaluation de la production écrite CLES niveau C1 du CECRL

Grille d'évaluation (C1.1, C1.2)

Langue / Cours: _____	Nom: _____
Semestre: _____	Prénom: _____
Enseignant/e: _____	N° d'étudiant: _____
Sujet: _____	

Tâche écrite:

Date: _____

Il / Elle possède des connaissances et des compétences linguistiques de haut niveau relatives à des domaines scientifiques généraux et inhérents à la spécialité qui lui permettent de communiquer sur des sujets pointus grâce à l'utilisation de moyens linguistiques variés. Il / Elle est en mesure, à l'écrit, de s'exprimer aisément et couramment sur des sujets pointus et complexes de son domaine de spécialité, qui sont pertinents pour la vie universitaire ou professionnelle dans le cadre d'un séjour à l'étranger. Il / Elle est à même d'exposer son point de vue de manière synthétique et logique dans un style adéquat.

Par rapport au niveau visé:	très bien maîtrisé	bien maîtrisé	satisfaisant	encore suffisant	insuffisant
(Elaboration du) contenu des textes	1	2	3	4	5
Réalisation de la tâche (Richesse du texte): <ul style="list-style-type: none"> - Compréhension et réalisation adaptée de la tâche. - Traitement pertinent de la problématique. - Présentation claire et convaincante des contenus. - Présentation nuancée et convaincante du point de vue / de la perspective de l'étudiant/e - L'étudiant/e a su mobiliser des connaissances générales sur le thème choisi. - L'étudiant/e peut présenter des contenus complexes de manière claire et détaillée à l'écrit. 					
Justesse des contenus: <ul style="list-style-type: none"> - Le contenu correspond au sujet. - Les informations sont restituées avec justesse. - Certains aspects sont présentés ou analysés de manière critique si nécessaire. - La solution proposée est plausible et convaincante. - Les exemples sont bien choisis. - L'étudiant/e est capable de présenter ses arguments avec clarté et efficacité dans des lettres formelles complexes (par exemple formuler une réclamation, prendre position dans une controverse). 					
Pertinence des contenus: <ul style="list-style-type: none"> - Contenus pertinents et adaptés. - L'étudiant/e est capable de traiter de contenus complexes dans des lettres, essais ou comptes rendus et d'en souligner les aspects importants. - Il / Elle est capable de présenter et d'étayer des points de vue par des arguments supplémentaires et des exemples adéquats. 					
Originalité / Richesse: <ul style="list-style-type: none"> - Mise en forme originale du contenu qui suscite l'intérêt du lecteur. - Réinvestissement de nombreuses informations pertinentes dans la présentation. 					
Evaluation (de l'élaboration) du contenu:					

Compétence pragmatique	1	2	3	4	5
Fonctionnalité: Effet sur le destinataire (incl. prise en compte du type de texte): <ul style="list-style-type: none"> - L'étudiant/e peut écrire différents types de textes. - L'étudiant/e peut écrire des courriers officiels, des courriels et des comptes rendus. - L'étudiant/e peut rédiger des rapports. - L'étudiant/e peut écrire des lettres standard, pour transmettre ou obtenir des informations détaillées (par exemple répondre à une annonce, postuler à un emploi). - L'étudiant/e peut rédiger des descriptions ou des textes de fiction clairs, détaillés et bien structurés, dans un style adapté au lecteur, convaincant et naturel. 					
Organisation / Cohérence / Cohésion / Présentation: <ul style="list-style-type: none"> - Texte clairement structuré. - Bonne introduction. - Maîtrise de l'articulation du texte, de l'enchaînement des arguments, et des transitions. - Les points principaux sont clairement formulés et complétés par des détails appropriés. - L'étudiant/e écrit de manière claire et fluide. - L'étudiant/e peut présenter son opinion en détail. - Bonne conclusion. - Présentation et paragraphes sont corrects et participent à l'organisation du texte. 					
Evaluation de la compétence pragmatique:					

Compétence linguistique	1	2	3	4	5
Correction linguistique:					
Grammaire: <ul style="list-style-type: none"> - Syntaxe correcte. - Utilisation juste de constructions adaptées au contexte. 					
Vocabulaire: <ul style="list-style-type: none"> - Utilisation convaincante d'un vocabulaire étendu. - Utilisation correcte du lexique de spécialité. - Utilisation pertinente de nombreux connecteurs, pour clarifier les liens logiques. 					
Orthographe / Ponctuation: <ul style="list-style-type: none"> - Orthographe correcte. - Ponctuation correcte. 					
Etendue des moyens d'expression:					
Grammaire: <ul style="list-style-type: none"> - Grande souplesse grammaticale. - Utilisation de structures complexes. 					

Compétence linguistique	1	2	3	4	5
Vocabulaire: <ul style="list-style-type: none"> - Utilisation convaincante d'un vocabulaire général et spécialisé étendu. - Expression précise et nuancée. - Utilisation adéquate de synonymes. - L'étudiant/e peut nuancer l'expression d'opinions et de propos, par exemple en fonction de la certitude / de l'incertitude, de la conviction / du doute, de la probabilité. - L'étudiant/e peut exprimer les points de vue, les sentiments et les expériences de manière nuancée. 					
Registre / Style: <ul style="list-style-type: none"> - L'étudiant/e peut choisir le style adapté aux différents lecteurs en utilisant le registre approprié. - L'étudiant/e peut adapter son vocabulaire et son style au contexte. - L'étudiant/e utilise de manière adéquate le registre exigé dans le contexte professionnel ou universitaire. 					
Evaluation de la compétence linguistique:					

	Calcul	Note
Elaboration du contenu (25%)	_____ x 25%	
Compétence pragmatique (25%)	_____ x 25%	
Compétence linguistique (50%)	_____ x 50%	
Note finale production écrite (100%)		

Signatures des examinateurs: _____

Grille 8: Université de Göttingen (UNICert) – Grille d'évaluation de l'écrit niveau C1 du CECRL

B. Grille d'évaluation élaborée par les participants de la réunion de réseau GULT à Graz, les 13 et 14 Janvier 2011

Grille d'évaluation adaptée à l'approche actionnelle pour la production écrite ET orale

Contexte de réalisation du produit

Atelier lors de la réunion de réseau GULT au CELV: comparer les grilles en usage pour le CLES et pour UNiCert, proposer des lignes directrices pour une grille GULT.

Les composants à évaluer

Après examen des divers modèles en présence, le groupe s'est orienté vers une composition tripartite des aspects à prendre en compte dans l'évaluation:

1. la réalisation effective de la tâche dans la perspective actionnelle: 30% du score;
2. les compétences discursives mises en évidence lors de la performance: 30% du score;
3. la maîtrise du code linguistique: 40% du score.

L'attribution d'une note

Le groupe a majoritairement souligné la nécessité d'attribuer une note à la prestation, pour compléter la mention du niveau globalement atteint, afin de faciliter l'intégration de la certification dans la reconnaissance des unités et crédits qui composent les cursus universitaires européens (système ECTL).

Les catégories de la grille

Il ne s'agit ici que de catégories générales à décliner selon les caractéristiques des tâches effectivement demandées en production.

1. REALISATION DE LA TACHE

(La communication du candidat est-elle efficace pour l'action envisagée ?)

- 1.1. Respect des consignes d'examen
- 1.2. Gestion du rôle social attribué

1.3. Pertinence thématique et socio-culturelle des informations / arguments avancés

2. COMPETENCES DISCURSIVES

(Le candidat construit-il un discours selon les règles discursives socialement requises ?)

(Détailler les descripteurs selon le type de discours requis par la tâche)

2.1. Capacité à gérer le discours présentatif ou narratif

2.2. Capacité à gérer le discours argumentatif

2.3. Capacité à gérer l'interaction langagière

3. MAITRISE DU CODE LINGUISTIQUE

(Le candidat maîtrise-t-il le code linguistique au niveau requis ?)

3.1. Syntaxe

3.2. Morphologie

3.3. Lexique

3.4. Phonétique, prosodie (production orale) / orthographe (production écrite)

Document 6: Définitions de la « tâche »

Le *Multilingual Glossary of Language Testing Terms*, conçu par les membres d'ALTE, définit une tâche de la manière suivante: « Combinaison de consignes, d'entrées et de réponses. Exemple: texte à lire, accompagné d'items à choix multiple, auxquels il est possible de répondre en se référant à la consigne » (ALTE / UCLES, 1998: 239).

Les membres de l'équipe GULT ont une conception différente de la « tâche ». Les définitions réunies ci-dessous (en anglais, français et allemand) reflètent au mieux la conception qu'a GULT de la ou des « tâche / tâches »:

Les tâches sont des activités du monde réel « que les gens réalisent dans la vie quotidienne et qui nécessitent la langue pour être menées à bien » (original en anglais: « that people do in everyday life and which require language for their accomplishment »)

(Norris, 1998: 33).

Une tâche est « une activité qui amène les individus à utiliser la langue pour atteindre un but ou un objectif particulier dans une situation particulière » (original en anglais: « an activity that involves individuals in using language for the purpose of achieving a particular goal or objective in a particular situation »)

(Bachman et Palmer, 1996: 44).

A task is an activity promoting language learning in which meaning is primary, and through which communication must take place that has a link with real-life activities and its completion has priority, because the success of its completion is assessed by outcomes.

(Wesche et Skehan, 2002)

Throughout the paper, the notion of “task” is used in a broader sense to refer to any structural language learning endeavour which has a particular objective, appropriate content, a specified working procedure, and a range of outcomes for those who undertake the task. “Task” is therefore assumed to refer to a range of workplans which have the overall purpose of facilitating language learning – from the simple and brief exercise type to more complex and lengthy activities such as group problem-solving simulations and decision-making.

(Breen, 1987: 23)

[O]ne of a set of differentiated, sequenceable problem-posing activities involving learners and teachers in some joint selection from a range of varied cognitive and communicative procedures applied to existing and new knowledge in the collective exploration and pursuance of foreseen or emergent goals within a social milieu.

(Candlin, 1987: 10)

A task is a “workplan”; that is, it takes the form of materials for researching or teaching language. A workplan typically involves the following: (1) some input (i.e., information that learners are required to process and use); and (2) some instructions relating to what outcome the learners are supposed to achieve. As Breen (1989) has pointed out, the task-as-workplan is to be distinguished from the task-as-process (i.e., the activity that transpires when particular learners in a particular setting perform the task). As we will see, the activity predicted by the task-as-workplan may or may not accord with the activity that arises from the task-as-process. Definitions of “task” typically relate to task-as-workplan.

(Ellis, 2000: 195)

From a psycholinguistic perspective a task is a device that guides learners to engage in certain types of information-processing that are believed to be important for effective language use and / or for language acquisition from some theoretical standpoint. This perspective is predictive, and, in some cases, deterministic. That is, it assumes that there are properties in a task that will predispose, even induce, learners to engage in certain types of language use and mental processing that are beneficial to acquisition. As Skehan, Foster and Mehnert (1998) put it “task properties have a significant impact on the nature of performance” (p. 245). The claim is, therefore, that there is a close correlation between the task-as-workplan and the task-as-process because the activity that results from the task-as-workplan is predictable from the design features of the task.

(Ellis, 2000: 197-198)

Il y a « tâche » dans la mesure où l'action est le fait d'un (ou de plusieurs) sujet(s) qui y mobilise(nt) stratégiquement les compétences dont il(s) dispose(nt) en vue de parvenir à un résultat déterminé.

(Conseil de l'Europe, 2001: 15)

Als Merkmale einer *task* werden deshalb unter anderem die folgenden beschrieben: Eine *task* nennt den Zweck und das erwartete Ergebnis einer Aktivität, sie legt den Schwerpunkt auf die Bedeutung dessen, was gesagt wird, und nicht auf die Verwendung einer bestimmten Form (z. B. die Anwendung einer grammatischen Struktur), und sie versucht, die Sprache so zu verwenden, wie sie im Alltag vorkommen könnte (*real or authentic language use*).

(Müller-Hartmann et Schocker-v. Ditfurth, 2005: 2)

A task is an activity which requires learners to use language, with emphasis on meaning, to attain an objective.

(Skehan, 2003: 3)

A task is intended to result in language use that bears a resemblance, direct or indirect, to the way language is used in the real world.

(Ellis, 2003: 16)

Im TBL [*Task-Based Learning*] sieht man die Aufgabe demgegenüber weniger als Zielpunkt einer Unterrichtssequenz, sondern als deren Ausgangs- und Angelpunkt. Das befrachtet die Aufgabe mit dem Anspruch, genuine Lernsituation zu sein, in der neues Wissen und Können erworben werden können. Zuweilen erhofft man sich in den konkreten Anforderungen der Aufgabenstellung auch ein erhebliches Motivationspotential, das bei den Schülern selbstgesteuertes Lernen auslöst, damit diese die Aufgabe zufriedenstellend bewältigen können.

(Klippel, 2006: 111)

Börner sieht im Aufgabenbegriff die Möglichkeit, die Implikationen eines starren Übungsbegriffs zu überwinden, der „zum Kampfbegriff für unkommunikative Sprachlerntätigkeiten mutierte“ (1999, 212) und an dessen Stelle nunmehr eine differenzierter Aufgabenbegriff treten zu lassen, der sowohl die lehrerseitige Initiierung sprachlicher Aktivitäten des Lerners umfasse, die als problemhaft antizipiert werden, sowie die Betrachtung lernerseitiger sprachlicher und kognitiver Handlungen (Lösen einer Aufgabe) als auch das Produkt eben dieser Handlungen (Lösung einer Aufgabe).

(Königs, 2006: 115)

Est définie comme tâche toute visée actionnelle que l'acteur se représente comme devant parvenir à un résultat donné en fonction d'un problème à résoudre, d'une obligation à remplir, d'un but qu'on s'est fixé. Il peut s'agir tout aussi bien, suivant cette définition, de déplacer une armoire, d'écrire un livre, d'emporter la décision dans la négociation d'un contrat, de faire une partie de cartes, de commander un repas dans un restaurant, de traduire un texte en langue étrangère ou de préparer en groupe un journal de classe.

(Conseil de l'Europe, 2001: 16)

Aufgaben bedeuten für die Lehrenden mehr und anderes als für die Lernenden, *Aufgaben* sind mehr und etwas anderes als *Aufgaben-Formulierungen*. Aufgaben-Formulierungen sind für die Lernenden da. Sie verlangen obligatorisch eine Fragestellung bzw. Aufforderung (die eigentliche Aufgabe), Angaben zu Themen und Materialien, die benutzt werden sollen, Angaben zu erwarteten Aktivitäten der Lernenden und zu den Resultaten, die aus diesen erwachsen sollen, und optimalerweise enthalten sie auch Angaben, die die Aufmerksamkeit auf Sprache sowie die Fähigkeit zur Selbststeuerung und Selbstbeurteilung der Lernenden fördern. Ihr Zweck ist erfüllt, wenn sie ihre Adressaten richtig informieren und orientieren.

(Portmann-Tselikas, 2006: 185)

A *task* is defined as any purposeful action considered by an individual as necessary in order to achieve a given result in the context of a problem to be solved, an obligation to fulfil or an objective to be achieved. This definition would cover a wide range of actions such as moving a wardrobe, writing a book, obtaining certain conditions in the negotiation of a contract, playing a game of cards, ordering a meal in a restaurant, translating a foreign language text or preparing a class newspaper through group work.

(Conseil de l'Europe, 2001: 10)

Références

ALTE (1998). *Multilingual glossary of language testing terms: Studies in language testing 6*. Cambridge: Cambridge University Press.

Bachmann, Lyle F. et Palmer, Adrian S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Breen, M. (1987). « Learner contributions to task design ». In: Candlin, C. and Murphy D. (éd.). *Language learning tasks*. Englewood Cliffs NJ: Prentice-Hall, 23-46.

Candlin, C.N. (1987). « Towards task-based learning ». In: Candlin, C. N. and Murphy, D. (éd.). *Lancaster practical papers in English language education. Vol. 7. Language learning tasks*. Englewood Cliffs, NJ: Prentice Hall, 5-22.

Conseil de l'Europe (2001). *Cadre européen commun de référence pour les langues: apprendre, enseigner, évaluer*. Strasbourg: Conseil de l'Europe / Paris: Didier.

Ellis, R. (2000). « Task-based research and language pedagogy ». In: *Language Teaching Research* 4(3), 193-220.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2009). « Task-based language teaching: sorting out the misunderstandings ». In: *International Journal of Applied Linguistics* 19, 221–246.

Klippel, F. (2006). « Aufgabenorientierung im Fremdsprachenunterricht. Wirklich ein neues Paradigma? ». In: Bausch, K.-R., Burwitz-Melzer, E., Königs, F. G. et Krumm, H.-J. (éd.). *Aufgabenorientierung als Aufgabe. Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Gunter Narr Verlag, 109-115.

Königs, F. G. (2006). « Aufgabenorientierung als Aufgabe. Überlegungen zur aktuellen Diskussion um ein ‘neues’ Konzept für den Fremdsprachenunterricht ». In: Bausch, K.-R., Burwitz-Melzer, E., Königs, F. G. et Krumm, H.-J. (éd.). *Aufgabenorientierung als Aufgabe. Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Gunter Narr Verlag, 115-122.

Müller-Hartmann, A. et Schocker-v. Ditfurth, M. (éd.) (2005). *Aufgabenorientierung im Fremdsprachenunterricht. Task-Based Language Learning and Teaching*. Tübingen: Narr.

Norris, J. M. et al. (1998). *Designing second language performance assessments*. Honolulu: University of Hawaii Press.

Portmann-Tselikas, P. R. (2006). « Was müssen Lehrende über Aufgaben wissen? » In: Bausch, K.-R., Burwitz-Melzer, E., Königs, F. G. et Krumm, H.-J. (éd.) *Aufgabenorientierung als Aufgabe. Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Gunter Narr Verlag, 182-191.

Skehan, P. (2003). « Task-based instruction ». In: *Language Teaching* 36, 1-14.

Wesche, M. B. et Skehan, P. (2002). « Communicative, task-based, and content-based language instruction ». In: Kaplan, R. B. (éd.). *The Oxford handbook of applied linguistics*. Oxford: Oxford University Press, 207-228.

Document 7: Définitions de l'« authenticité »

« Authenticité »:

Authenticity: the degree to which test materials and test conditions succeed in replicating those in the target use situation.

(McNamara, 2000: 131)

On a pu affirmer que l'authenticité est la résultante de l'interaction entre le lecteur et le texte et pas seulement une caractéristique du texte [...]. L'identité du lecteur, son projet de lecture, l'intention du scripteur et le degré de proximité sociale et culturelle entre le lecteur et le texte ont une incidence sur la nature de l'interaction entre un lecteur et un texte donné.

(*Evaluation de compétences en langues et conception de tests*, 2002: 24)

L'authenticité interactionnelle peut se définir comme l'interaction entre l'activité d'évaluation (la tâche) et le candidat; elle suppose que les rédacteurs et les concepteurs de tests devraient:

- proposer des textes, des situations et des tâches qui simulent la « vraie vie » sans essayer de la reproduire à l'identique;
- essayer de proposer des situations et des tâches qui ont des chances d'être pertinentes pour le candidat potentiel à un niveau donné;
- clarifier la *finalité* de chaque tâche ainsi que le *public cible* en mettant en contexte adéquat;
- expliciter les *critères de réussite* de la tâche.

(*Evaluation de compétences en langues et conception de tests*, 2002: 24)

« Ressources authentiques »:

There are a number of definitions related to authentic materials and three are presented here:

An authentic text is a stretch of real language, produced by a real speaker or writer for a real audience and designed to convey a real message of some sort.

(Morrow 1977: 13)

Authentic texts (either written or spoken) are those which are designed for native speakers: they are real texts designed not for language students, but for the speakers of the language in question.

(Harmer 1983: 146)

A rule of thumb for authentic here is any material which has not been specifically produced for the purposes of language teaching.

(Nunan 1989: 54)

(*Teaching Functional English Through Authentic Materials*, 2011)

The definitions of authentic materials are slightly different in literature. What is common in these definitions is “exposure to real language and its use in its own community”. Rogers (1988) defines it as “appropriate” and “quality” in terms of goals, objectives, learner needs and interest and “natural” in terms of real life and meaningful communication (p. 467). Harmer (1991), cited in Matsuta (n.d., para. 1) defines authentic texts as materials which are designed for native speakers; they are real text; designed not for language students, but for the speakers of the language. Jordan (1997, p. 113) refers to authentic texts as texts that are not written for language teaching purposes. Authentic material is significant since it increases students' motivation for learning, makes the learner be exposed to the ‘real’ language as discussed by Guariento & Morley (2001, p. 347). The main advantages of using authentic materials are (Philips and Shettlesworth 1978; Clarke 1989; Peacock 1997, cited in Richards 2001):

- 1) They have a positive effect on learner motivation.
- 2) They provide authentic cultural information.
- 3) They provide exposure to real language.
- 4) They relate more closely to learners' needs.
- 5) They support a more creative approach to teaching.

We can claim that learners are being exposed to real language and they feel that they are learning the ‘real’ language. These are what make us excited and willing to use authentic materials in our classrooms, but while using them, it is inevitable that we face some problems.

(Kilickaya, 2004)

Peacock (1977) dit que les ressources authentiques sont des « [...] ressources qui ont été produites pour remplir un certain objectif social dans la communauté linguistique ». (Version originale en anglais: « materials that have been produced to fulfil some social purpose in the language community »).

Il faut mentionner ici la distinction de Widdowson (1990) entre les termes « *authentic material* » (en français: « ressources authentiques ») et « *genuine material* » (en français: « ressources réalistes »):

Authentic would be material designed for native speakers of English used in the classroom in a way similar to the one it was designed for. For example, a radio news report brought into the class so students discuss the report on pollution in the city where learners live. Most of the time, though, this material is used in a *genuine* way, in other words, not in the way it was intended, but in a somewhat artificial way. For example, news articles where the paragraphs are cut up and jumbled so students have to put them back together in the correct order.

« Authenticité du Texte et de la Tâche »:

Authentic texts or materials have been defined by Villegas Rogers and Medley (1988) as “[...] language samples – both oral and written – that reflect a naturalness of form, and an appropriateness of cultural and situational context that would be found in the language as used by native speakers” (p. 468). Texts that are prepared *for* native speakers *by* native speakers reflect the culture and societal values of everyday life. “No textbook culture note on the Hispanic family, for example, can replace the study of authentic birth or christening, wedding and death announcements, where, under the observable linguistic conventions, lie the rituals of events, the connotations of rites of passage, the meaning of family, and the dynamic nature of culture”.

(Galloway et Labarca, 1990: 139)

For our purposes, any text that is *purposeful, meaningful*, and has a real *communicative intent* for a real audience can be considered to be authentic. In other words, it is authentic in the sense that it was not originally produced for language-teaching purposes but rather for the purpose of communicating meaning (Brinton *et al.*, 1989, p. 17). This means that an e-mail message sent via the Internet by a student of German to another student of German is “authentic” as long as the message is meaningful (even though the message was not written by a native speaker for another native speaker). Furthermore, authenticity in a deeper sense does not reside in the text itself but rather is determined by how that text is used (Hutchinson & Waters, 1987), i.e., the authenticity of the task. For example, if a teacher uses an article from a target culture magazine for the sole purpose of having the students underline all of the instances in which the subjunctive appears, the authenticity of the task disappears.

Let’s examine a task and consider ways in which it can be slightly altered to become more authentic. Imagine that students are engaged in a unit on Costa Rica (or any other target country). As a culminating activity at the end of the unit, the teacher decides to have students create travel brochures in the target language to demonstrate their knowledge of what they have learned. Such a task asks that the students pretend to act as

native speakers, which they clearly are not. Kramersch (1993) would argue that authenticity involves having students be who they are – learners of the target language. To revise the task somewhat with an eye toward greater authenticity, the teacher can have students create travel itineraries for a group of students who will be traveling to Costa Rica, the intent being to demonstrate their knowledge of what they have learned by communicating it to other students.

Another example would involve having students at the beginning of the unit write letters in the target language to various travel agencies, tourist bureaus, and “Chamber of Commerce” equivalents to indicate that they (1) are students of Spanish, (2) are studying about Costa Rica, and (3) are interested in receiving travel information in Spanish. Such a task has a real purpose and a real audience. The added benefit is that it will also lead to additional authentic materials for classroom use!

A final example of an authentic task for this instructional setting is to have students write to Costa Rican students about Minnesota (i.e., their home state), given what they have learned about Costa Rica. A letter written for this task might include, for example, a comparison between Minnesota's Boundary Waters and Costa Rica's Tortuguero National Park in terms of their environmental restrictions.

These suggestions highlight the importance of creating tasks that involve students in using language for real communicative purposes and for real audiences. For example, a teacher might have high school students write children's stories that are then shared with an elementary language program in the same district. It is important to note, however, that it is not possible to make every task or text authentic in the language classroom. Sometimes students need to pretend to be native speakers for a role play; sometimes they need to write for a hypothetical audience; sometimes they need to read a text that has been adapted for nonnative speakers of the language. Such activities are valuable and certainly have a place in the language curriculum. What is important (and possible!), however, is for teachers to find a good balance in their curriculum between tasks and texts that are less authentic and those that represent the principles of authenticity as described above. Teachers should also make sure that some of the texts they use in the curriculum contain language as used by native speakers so as to incorporate cultural and linguistic authenticity. A number of authentic texts (i.e., written by native speakers for native speakers of the target language) are used in CoBaLTT lessons/units found at the Web Resource Center.

(Tedick, 2003)

« Evaluation authentique »

Authentic assessment: a form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills.

(Mueller, 2011)

Authentic assessment aims to evaluate students' abilities in "real-world" contexts. In other words, students learn how to apply their skills to authentic tasks and projects. Authentic assessment does not encourage rote learning and passive test-taking. Instead, it focuses on students' analytical skills; ability to integrate what they learn; creativity; ability to work collaboratively; and written and oral expression skills. It values the learning process as much as the finished product.

In authentic assessment, students:

- do science experiments
- conduct social-science research
- write stories and reports
- read and interpret literature
- solve math problems that have real-world applications.

(TeacherVision, 2011)

Références

Evaluation de compétences en langues et conception de tests (2002). Edité par les membres d'ALTE. Disponible sur le site internet:

<http://www.coe.int/t/DG4/Portfolio/documents/Guide%20pour%20l'Evaluation%20-%20rev%20octobre%202002.doc> [consulté le 9 septembre 2011].

Kilickaya, F. (2004). *Authentic materials and cultural content in EFL classrooms*. Disponible sur le site internet: <http://iteslj.org/Techniques/Kilickaya-AuthenticMaterial.html> [consulté le 9 septembre 2011].

McNamara, T. (2000). *Language testing*. Oxford *et al.*: Oxford University Press.

Mueller, J. (2011). *What is authentic assessment?* Disponible sur le site internet: <http://jonathan.mueller.faculty.noctrl.edu/toolbox/whatisit.htm> [consulté le 9 septembre 2011].

Peacock, M. (1997). « The effect of authentic materials on the motivation of EFL learners ». In: *English Language Teaching Journal* 51: 2. Cité sur le site internet: <http://www3.telus.net/linguisticsissues/authenticmaterials.html> [consulté le 9 septembre 2011].

Teaching functional English through authentic materials (2011). Disponible sur le site internet:

http://www.teachingstylesonline.com/teaching_functional_english_through_authentic_materials.html [consulté le 9 septembre 2011].

TeacherVision® (2011). *Authentic assessment overview*. Disponible sur le site internet: <http://www.teachervision.fen.com/teaching-methods-and-management/educational-testing/4911.html#ixzz1ERN2vV00> [consulté le 9 septembre 2011].

Tedick, D. J. (2003). *CAPRII: Key concepts to support standards-based and content-based second language instruction*. Minneapolis, MN: University of Minnesota, Center for Advanced Research on Language Acquisition (CARLA). La publication originale se trouve sur le site CARLA:

<http://www.carla.umn.edu/cobaltt/modules/strategies/CAPRII/READING1/caprii.htm> [consulté le 9 septembre 2011]. Utilisation autorisée.

Widdowson, H. (1990). *Aspects of language teaching*. Oxford: Oxford University Press. Cité sur le site internet:

<http://www3.telus.net/linguisticsissues/authenticmaterials.html> [consulté le 9 septembre 2011].

Lecture complémentaire

Dans son article « Authentic assessment: Implications for EFL performance testing in Korea », publié en 2002 dans *Secondary Education Research* 49, 89-122, Andrew Finch fournit une description complète de l'évaluation authentique et la compare aux tests standardisés. Pour plus d'informations:

http://www.finchpark.com/arts/Authentic_Assessment_Implications.pdf.

Voir aussi:

Wiggins, G. (1990). « The case for authentic assessment ». In: *Practical Assessment, Research & Evaluation*, 2(2). Disponible sur le site internet: <http://pareonline.net/getvn.asp?v=2&n=2>.

Document 8: Etapes de la conception d'un test par tâches

La procédure suivante peut vous aider à préparer votre propre test selon une approche actionnelle:

Conception du scénario:

1. Pensez à vos candidats: Quel est leur niveau de compétence et quel est leur domaine d'études ? Ont-ils des centres d'intérêt spécifiques ?
2. Pensez à une situation authentique dans laquelle ils devraient utiliser leurs compétences en langue étrangère dans leur domaine (dans le cadre de leurs études ou de leur vie professionnelle future). Quelle situation pourrait servir de cadre à votre épreuve ?
3. Pensez à un problème ou un projet qu'ils devraient être capables de gérer dans la langue cible pendant leurs études ou dans leur future profession. Quelle macro-tâche devraient-ils avoir à réaliser pour fournir un produit final spécifique adapté aux besoins particuliers dans la situation donnée ?
4. Pensez aux productions écrites et orales individuelles qu'ils ont à fournir pour réaliser la macro-tâche: qu'ont-ils à préparer à l'écrit et à présenter oralement ?

Collecte de documents:

5. Maintenant rassemblez les documents nécessaires pour l'écoute et la lecture. Commencez par rechercher des documents d'écoute, car cela nécessitera sans doute plus de temps et sera plus compliqué.

Rédaction du test: Maintenant préparez votre sujet !

6. Décrivez la situation.
7. Concevez une macro-tâche (réaliste), c'est-à-dire décrivez l'objectif général.
8. Décidez comment vous voulez tester les compétences réceptives, choisissez les textes (oraux et écrits) et concevez les fiches de travail.

9. Préparez ensuite un corrigé pour la partie réceptive du test.
10. Pensez à un lien logique entre la partie réceptive et la partie productive du test.
11. Pour la partie de production, décidez si les candidats devront travailler individuellement ou (partiellement ou entièrement) en binômes ou en groupes.
12. Pour la partie de production, concevez des tâches intermédiaires pour les différentes activités langagières, et définissez clairement les productions attendues (à l'écrit et à l'oral).
13. Définissez les rôles des différents candidats / des membres d'un groupe, leurs tâches respectives et les productions attendues.
14. Puis définissez quelles productions vous attendez et adaptez vos grilles d'évaluation. A ce stade, fixez le minimum requis pour réussir au test.
15. Pensez à l'ordre logique des différentes tâches. Associez-les conformément à une situation authentique dans la vraie vie (par exemple: préparez d'abord un rapport écrit et ensuite présentez-le oralement à un auditoire, par exemple l'équipe).
16. Lorsque vous adaptez le test aux besoins de votre institution, soyez attentifs à l'importance respective des différentes tâches intermédiaires (activités productives et réceptives) dans l'évaluation de la macro-tâche.

Test GULT: Fiche récapitulative

Titre:	
Niveau du CECRL:	
Scénario (problématique):	
Problème / Simulation / Projet:	
Contexte:	
Situation:	
Rôle(s):	
Tâches:	
Macro-tâche:	
Tâches intermédiaires:	
Tâche d'écoute / production:	
Tâche de lecture / production:	
Tâche écrite / production:	
Tâche orale / production:	
Document(s) d'écoute:	
Document(s) de lecture:	
Ressources documentaires:	
Corrigé pour les évaluateurs:	
Grilles pour les évaluateurs:	
Commentaires pour les évaluateurs:	

Liste de contrôle pour la conception d'un test GULT – version 1

1. Y a-t-il une macro-tâche ? <input type="checkbox"/> oui <input type="checkbox"/> non
2. Le test est-il construit à partir d'un problème à résoudre, ou s'agit-il d'une simulation / d'un projet ? Le problème: La simulation / le projet:
3. La tâche est-elle intégrée dans un contexte spécifique ? <input type="checkbox"/> oui <input type="checkbox"/> non La situation:
4. Les rôles des candidats sont-ils clairement définis ? <input type="checkbox"/> oui <input type="checkbox"/> non Les rôles:
5. Y a-t-il des tâches intermédiaires ? <input type="checkbox"/> oui <input type="checkbox"/> non Détaillez-les: a) b) c) d) e) f)
6. Dans quelle mesure la macro-tâche est-elle réaliste ? Et les tâches intermédiaires ? Et le problème ou la simulation / le projet ? Et la situation ? Et le dossier documentaire ?
7. Le problème / la simulation / le projet, la situation, les tâches et les rôles sont-ils en rapport avec le candidat et ses centres d'intérêt ? <input type="checkbox"/> oui <input type="checkbox"/> non

<p>8. Le test est-il intéressant et stimulant ?</p> <p><input type="checkbox"/> oui <input type="checkbox"/> non</p>
<p>9. Dans quelle mesure les candidats sont-ils déjà familiarisés avec la thématique et le contenu des documents ?</p>
<p>10. Dans quelle mesure la thématique permet-elle une discussion contradictoire?</p>
<p>11. Quel est le niveau de complexité de la thématique ?</p>
<p>12. Les rôles des candidats sont-ils clairs ?</p> <p><input type="checkbox"/> oui <input type="checkbox"/> non</p>
<p>13. Les étapes successives que les candidats doivent réaliser pendant l'épreuve sont-elles présentées de manière claire et structurée ?</p> <p><input type="checkbox"/> oui <input type="checkbox"/> non</p>
<p>14. Y a-t-il assez de ressources disponibles ?</p> <p><input type="checkbox"/> oui <input type="checkbox"/> non</p> <p>Pour chaque aspect de la thématique ?</p> <p><input type="checkbox"/> oui <input type="checkbox"/> non</p>
<p>15. Quelles productions doivent-elles être réalisées et présentées ?</p> <p>Productions écrites:</p> <p>Productions orales:</p>
<p>16. Est-il possible de réaliser la tâche sans utiliser les documents fournis (c'est-à-dire les candidats utilisant uniquement leurs propres connaissances sur le sujet) ?</p> <p><input type="checkbox"/> oui <input type="checkbox"/> non</p>

Liste de contrôle pour la conception d'un test GULT – version 2

	Oui	Non
1. Les apprenants sont-ils déjà familiers de l'approche par tâches ?		
2. Y a-t-il une macro-tâche ?		
3. Cette tâche est-elle intégrée dans un contexte spécifique ?		
4. Les rôles des candidats sont-ils clairement définis ?		
5. Y a-t-il des tâches intermédiaires appropriées ?		
6. La macro-tâche est-elle suffisamment réaliste ?		
7. Les tâches intermédiaires sont-elles suffisamment réalistes ?		
8. Le problème / la simulation / le projet sont-ils suffisamment réalistes ?		
9. Les ressources documentaires (supports d'évaluation) sont-elles authentiques ?		
10. Le problème / la simulation / le projet sont-ils en rapport avec le candidat et ses centres d'intérêt ?		
11. La situation et les tâches sont-elles en rapport avec le candidat et ses centres d'intérêt ?		
12. La thématique permet-elle de développer une argumentation ?		
13. Le niveau de complexité de la thématique est-il adapté aux apprenants ?		
14. Les rôles des candidats sont-ils à la fois clairs et adaptés à leurs besoins et à leurs centres d'intérêt ?		
15. Les consignes sont-elles claires pour chaque étape de l'épreuve ?		
16. Les candidats disposent-ils d'assez de ressources documentaires pour réaliser la tâche ?		
17. Les candidats doivent-ils utiliser les documents fournis pour réaliser la tâche ?		

Document 9: Glossaire³⁷

Analyse d'items: a lieu une fois le test administré et noté, lorsqu'il est temps de le réexaminer en détail. Tout enseignant-testeur devrait procéder à une simple analyse d'items, qui consiste à calculer les indices de difficulté et de discrimination.

Approche actionnelle dans l'apprentissage (et l'évaluation) en langues: suit une approche communicative, mais va plus loin. L'objectif n'est plus de réaliser un jeu de rôles dans lequel les apprenants peuvent montrer ce qu'ils ont acquis, mais de l'insérer dans un contexte et une situation spécifiques, avec un objectif bien défini. On ne jugera pas si les candidats se sont exprimés de manière (linguistiquement) correcte (par exemple s'ils étaient capables de demander leur chemin au niveau A du CECR ou de faire une présentation sur un sujet spécialisé au niveau C), mais s'ils ont réussi à réaliser l'objectif de la tâche, c'est-à-dire dans les exemples cités plus haut, s'ils ont trouvé l'endroit où ils voulaient aller (niveau A) ou si les auditeurs ont compris le contenu de la présentation et si le contenu a eu un impact sur les activités du groupe (niveau C). Dans l'approche GULT des tests de langues, l'évaluation de la performance de l'étudiant ne repose pas seulement sur le produit de l'activité, mais aussi sur des descripteurs prenant en compte le contenu autant que les compétences linguistiques et pragmatiques, en suivant une approche interactionniste.

Approches par les tâches / actionnelles de l'évaluation en langues: elles sont mises en œuvre lorsque la tâche que l'examineur demande au candidat de réaliser incite celui-ci à agir. Ceci signifie que dans une approche par tâches, une macro-tâche décrit l'activité générale qui doit être menée à bien, un résultat bien défini étant attendu à la fin. Afin d'atteindre ce résultat, le candidat doit élaborer des productions prédéfinies, à savoir un rapport écrit expliquant le problème à résoudre et ses solutions possibles, et une présentation orale de ces solutions à des interlocuteurs donnés dans un contexte donné. Toutes ces activités et productions sont liées entre elles. Pour parvenir à ces résultats, le candidat doit souvent réaliser des tâches intermédiaires, ou micro-tâches, précises.

Authenticité: le point jusqu'auquel les supports et les conditions de passation du test parviennent à reproduire ceux de la situation d'utilisation cible. (Traduction de la définition du terme « *authenticity* » tirée de McNamara 2000: 131)

37 En l'absence de source, les entrées ont été rédigées par l'équipe de projet.

Capacité langagière: ensemble de compétences qui définissent la capacité d'un individu à utiliser la langue dans des buts communicatifs variés. (ALTE, 1998: 213)

Choix multiple: c'est un format de question dans un test où les candidats doivent choisir parmi plusieurs alternatives proposées la seule réponse correcte. (Traduction de la définition du terme « *multiple-choice format* » tirée de McNamara, 2000: 135)

Compétence communicative: capacité à utiliser un langage approprié dans des situations et des dispositifs de communication variés. (ALTE, 1998: 214)

Compétences productives: parler et écrire.

Compétences réceptives: écouter et lire.

Construct / construit: attribut hypothétique des individus ou opération mentale qui ne peut être directement ni mesurée ni observée (par exemple, en évaluation des langues, la capacité de compréhension orale). Les tests de langue essaient de mesurer les différents constructs qui sous-tendent les capacités langagières. En dehors de la capacité langagière, la motivation, l'attitude et l'acculturation constituent des constructs pertinents. (ALTE, 1998: 216)

Critère: 1) l'aspect de comportement pertinent pour la conception du test; ou 2) un aspect de la performance évalué dans la notation du test, par exemple la fluidité, la justesse, etc.

Descripteurs de niveau: dans les échelles de notation, critères décrivant les caractéristiques d'une performance minimale acceptable à un niveau donné. (Traduction de la définition du terme « *level descriptors* » tirée de McNamara, 2000: 134)

Difficulté de l'item: la proportion de réponses correctes par rapport au total des réponses à un item, par exemple si 20 étudiants sur 30 donnent la bonne réponse à un item, la difficulté de l'item est de 66% (20/30) et l'index de difficulté est $p=0,66$.

Document authentique: texte ou document utilisé dans un test et qui n'a pas été conçu à des fins pédagogiques ou pour des situations d'évaluation. (ALTE, 1998: 219)

Domaine: ensemble défini de contenus et / ou de niveaux de capacité que l'on va tester à l'aide d'une tâche spécifique ou de l'une des composantes d'un examen. (ALTE, 1998: 219)

Etablir le barème: fixer le nombre de points par item, par tâche et par partie et décider ce qui est exigé pour obtenir la note maximale. Le point le plus important dans un barème est le point de césure, en d'autres termes, le point où passe la ligne

entre ceux qui réussissent le test et ceux qui échouent. Le point de césure optimal ne repose pas seulement sur l'intuition du testeur mais aussi sur la recherche.

Etalonnage: syn.: standardisation. Processus par lequel on s'assure que les examinateurs sont d'accord avec une procédure donnée et appliquent correctement les échelles de notation. (ALTE, 1998: 223)

Evaluation analytique: évaluation séparée de chaque aspect d'une performance, par exemple pour la production écrite la grammaire, la structure du texte, le contenu, etc. (Traduction de la définition du terme « *analytic rating* » tirée de McNamara, 2000: 131)

Evaluation (à référence) critériée: signifie que le candidat doit satisfaire à certains critères afin de réussir le test ou d'obtenir une certaine note. La plupart des tests de langue à l'université sont critériels, ce qui signifie que l'on sait au préalable ce qui est nécessaire pour obtenir par exemple le score maximum, mais on ne sait pas combien d'étudiants obtiendront telle ou telle note ou échoueront au test.

Evaluation formative: évaluation qui prend place pendant l'apprentissage plutôt qu'à la fin. Les résultats peuvent permettre à l'enseignant de pratiquer dès le début une remédiation ou de changer si besoin est le déroulement de son enseignement. Les résultats peuvent également aider un étudiant à identifier et cibler ses faiblesses. (ALTE, 1998: 223)

Evaluation indirecte: *stricto sensu*, couvre toutes les formes d'évaluation des compétences réceptives, car nous ne pouvons pas tester la compréhension de l'écrit ou de l'oral en tant que telles – dans un test de lecture ou d'écoute, nous devons recourir à la forme orale ou écrite pour établir ce qui a été compris. Un exemple plus concret d'évaluation indirecte: on donne la moitié d'un dialogue sous forme écrite au candidat et on lui demande de rédiger l'autre moitié. En d'autres termes, la compétence de production orale est testée ici à travers la production écrite.

Evaluation normative: les notes sont déterminées par les résultats antérieurs de la population testée et sont attribuées aux candidats suivant le percentile dans lequel ils se situent. Par exemple, nous savons déjà avant même de noter le premier test quel pourcentage des candidats obtiendra la meilleure note ou échouera. Cela signifie que le score brut utilisé comme point de césure peut différer d'un test à l'autre, mais pas au point de rendre le test non fiable. Les tests normatifs ne sont pas appropriés pour évaluer des petits effectifs.

Evaluation sommative: évaluation qui a lieu à la fin d'un cursus ou d'une séquence d'enseignement (traduction de la définition du terme « *summative evaluation* » tirée du *Testing Glossary*, 2011).

Fidélité: uniformité, constance ou stabilité des mesures dans un test. Plus un test est fidèle, moins il contient d'erreurs accidentelles. Un test présentant une erreur systématique, par exemple un biais vis-à-vis de certains groupes, peut être fidèle, mais pas valide. (ALTE, 1998: 224-225)

Fidélité inter-individuelle: on la calcule quand plusieurs examinateurs notent le même test, et que les candidats ont besoin de savoir que tous les examinateurs évaluent les tests de la même façon.

Fidélité intra-individuelle: estimation de la fidélité de l'évaluation, basée sur le degré de similitude de notation d'une même performance effectuée à différentes occasions. (ALTE, 1998: 225)

Indice de difficulté: dans la théorie classique du test, la difficulté d'un item est fonction de la proportion (p) de candidats qui y a répondu correctement. L'indice de difficulté est lié à la population considérée et varie selon le niveau d'habileté des candidats. (ALTE, 1998: 227). L'indice de difficulté (symbolisé par la lettre « p » dans les tableaux) se calcule en divisant le nombre de réponses correctes par le nombre de candidats. L'index de difficulté varie de 0 à 1. Plus la valeur est élevée, plus l'item est facile. Il est néanmoins usuel d'approuver des indices de difficulté plus élevés pour les premiers et les derniers items, afin que les candidats ne soient pas effrayés par le test au début et qu'ils le terminent en ayant le sentiment qu'ils étaient à la hauteur. Cependant, les concepteurs de tests doivent au préalable connaître assez bien la population cible pour avoir idée de ce que sera la difficulté des items du test.

Indice de discrimination: un outil qui aide les concepteurs du test à établir si les items discriminent vraiment les bons des mauvais apprenants, même si les items du test donnent lieu au nombre attendu de réponses correctes. En d'autres termes, si l'indice de difficulté prévu est 0,5 (= la moitié de la population testée a fourni la bonne réponse à l'item), l'item est de mauvaise qualité si ceux qui ne sont pas bons dans la langue ont la réponse juste tandis que les bons apprenants n'ont pas fourni la réponse correcte à cet item. On peut calculer l'indice de discrimination (« D » dans les tableaux) en utilisant une formule simple. La population testée est répartie en une bonne moitié et une moitié faible en fonction des scores bruts totaux ou d'autres informations concernant leurs compétences en langue. La formule est la suivante: (le nombre de réponses correctes dans la bonne moitié moins le nombre de réponses correctes dans la moitié faible) divisé par le nombre total de bonnes réponses. L'indice D varie entre +1 (toutes les réponses

correctes sont dans la bonne moitié) et -1 (toutes les réponses correctes sont dans la moitié faible).

Item: une entrée ou question dans un test ou un questionnaire. Exemple d'un item: « Rédigez une phrase en utilisant le terme 'salubre' ». (Traduction de la définition du terme « *test item* » tirée du *Testing glossary*, 2011)

Item à choix multiple: type d'item qui consiste en une question ou une phrase incomplète (stem), accompagnée d'un choix de réponses ou de propositions pour compléter la phrase (options). Le candidat devra choisir l'option correcte (clé) parmi trois, quatre ou cinq possibilités. Aucune production langagière ne lui est demandée. C'est pour cette raison qu'on utilise habituellement les items à choix multiple dans les tests de compréhension écrite et orale. Ils peuvent être discrets ou basés sur du texte. (ALTE, 1998: 228)

Méthode communicative / tests communicatifs: des tests lors desquels les candidats communiquent véritablement dans la langue cible, en transmettant et en échangeant des informations pertinentes.

Norme: 1) une distribution des notes pour un test provenant d'une analyse empirique qui fournit des données de référence pour des groupes de candidats correspondants; les résultats pour le TOEFL (*Test of English as a Foreign Language*) par exemple sont communiqués avec référence aux normes pour que les étudiants puissent se situer par rapport à l'ensemble des étudiants étrangers (traduction de la définition du terme « *norm* » tirée du *Testing Glossary*, 2011); 2) standard de performance. Dans un test standardisé, on détermine la norme en enregistrant les notes d'un grand groupe. La norme ou les standards basés sur la performance de ce groupe sont utilisés pour évaluer la performance de groupes suivants composés de candidats du même type. (ALTE, 1998: 232)

Notation analytique: méthode de notation qui peut être utilisée dans les tests de production langagière comme ceux d'expression écrite et orale. L'examineur évalue à partir d'une liste de points spécifiques. Dans un texte de production écrite par exemple, l'échelle analytique peut inclure des critères de grammaire, de vocabulaire, d'utilisation des articulateurs, etc. (ALTE, 1998: 232)

Point de coupure / point de césure / note de césure: une note qui marque la limite entre les réussites et les échecs.

Pondération: action d'assigner un nombre différent maximal de points à un item, une tâche ou une composante, afin de changer sa contribution relative à l'ensemble des points, en fonction des autres parties du test. Si par exemple, on attribue une note double à tous les items de la tâche n° 1 d'un test, la tâche n° 1 sera proportionnellement plus importante que les autres tâches dans le total des points obtenus. (ALTE, 1998: 235)

Questions / tâches intégrées: se réfère[nt] à des questions ou tâches à réaliser qui mettent en jeu plus d'une habileté ou sous-habileté. Exemples: compléter un test de closure, participer à un entretien oral, lire une lettre et y répondre. (ALTE, 1998: 236)

Scores bruts: le total des points avant toute manipulation ou ajustement.

Standardisation: voir « étalonnage ».

Tâche / tâches: 1) [...] une activité qui exige des apprenants qu'ils utilisent la langue, en privilégiant le sens, pour atteindre un objectif (traduction de la définition du terme « *task* » tirée de Skehan 2003: 3); 2) [...] une activité favorisant l'apprentissage de la langue dans laquelle le sens est essentiel, et qui doit provoquer la communication [...], qui est en rapport avec des activités de la vie réelle, et dont la réalisation est prioritaire, parce que sa réussite est évaluée en fonction du résultat obtenu (traduction de la définition du terme « *task* » tirée de Weideman 2006: 87); 3) [...] sont des activités qui ont le sens pour objet premier. La réussite est évaluée en termes de réalisation d'une production, et les tâches sont en général proches de l'usage de la langue dans la vraie vie. Ainsi l'enseignement par les tâches s'appuie assez fortement sur l'approche communicative dans l'enseignement des langues. (Traduction de la définition du terme « *tasks* » tirée de Skehan 1996: 20)

Test chronométré: test à durée limitée. Les candidats les plus lents obtiendront une note inférieure car ils ne pourront pas achever le test. Dans un test chronométré, le niveau de difficulté des questions doit être tel qu'il permette aux candidats d'y répondre correctement, s'il n'y avait pas de contrainte de temps. Syn.: test de vitesse. (ALTE, 1998: 240)

Test de puissance intellectuelle: test dont la durée est suffisante pour que tous les candidats puissent le terminer, mais qui contient des tâches ou des items d'un degré de difficulté qui rend peu probable que la majorité des candidats répondent à tous correctement. (ALTE, 1998: 241)

(Note: Evidemment, l'enseignement des langues étrangères à l'université ne recourt ni aux tests chronométrés, ni aux tests de puissance intellectuelle à l'état pur. Mais lorsqu'ils projettent des tests, les concepteurs doivent avoir une idée très claire du temps qui sera accordé pour chaque item et pour l'ensemble du test. Il faudra se baser sur une durée estimée moyenne: jamais tous les étudiants n'auront besoin de la même durée pour effectuer un test.)

Test diagnostique: utilisé pour découvrir les points forts et les lacunes d'un apprenant. Les résultats peuvent servir à des prises de décisions concernant la formation future, l'apprentissage ou l'enseignement. (ALTE, 1998: 242)

Tests à enjeux élevés: ce sont des épreuves importantes dont le résultat entraîne des décisions majeures telles que l'admission dans une université. En général, les tests à enjeux élevés ne peuvent également être passés qu'un nombre limité de fois et peuvent être payants.

Tests à enjeux limités: épreuves mineures qui peuvent être repassées, ou s'il y a une série de tests, un résultat faible dans un test peut être compensé par de bons résultats dans un autre test. Les quizzes ou les petits tests de langue sont des exemples typiques de tests à enjeux limités.

Tests à items discrets / tests à éléments isolés: sont composés de parties qui ciblent des phénomènes linguistiques clairement identifiables et qui sont totalement séparées les unes des autres, avec leurs propres totaux de points; et le score est la somme des points donnés pour les différents items.

Tests authentiques: représentent la forme extrême de l'évaluation communicative et devraient consister à évaluer en tant que tests de performance des situations authentiques d'utilisation de la langue étrangère. Pour un tel test, on pourrait envoyer des étudiants dans une bibliothèque où ils doivent se procurer un certain livre en utilisant la langue cible. L'évaluateur filmerait les échanges entre les étudiants et le bibliothécaire, et évaluerait ensuite la production orale des étudiants.

Tests critiques: test[s] dans lequel [lesquels] la performance du candidat est interprétée en fonction de critères prédéterminés. L'accent est mis davantage sur la maîtrise d'objectifs que sur le classement que la note du candidat lui impose dans un groupe. (ALTE, 1998: 240)

Tests d'acquisition: voir « tests de connaissances ».

Tests d'aptitude: sont conçus pour mesurer la capacité ou le potentiel permettant à un candidat de suivre avec succès un programme universitaire, d'apprendre une langue étrangère, de se former à une profession donnée ou d'acquérir toute autre compétence. (Traduction de la définition du terme « *aptitude tests* » tirée du *Testing Glossary*, 2011)

Tests de closure: type de tâche lacunaire dans laquelle des mots entiers sont supprimés d'un texte. Dans un test de closure traditionnel, on supprime un mot tous les x mots. On appelle également test de closure l'exercice dans lequel des phrases courtes sont supprimées d'un texte, ou lorsque l'élaborateur choisit les mots qui seront supprimés, comme c'est le cas dans le test de closure sélectif. Les candidats devront fournir les mots manquants (test ouvert) ou les choisir dans une liste (choix multiple ou test de closure à lacunes sélectives). Le corrigé d'un test ouvert peut comporter soit le mot exact (le mot supprimé du texte original

étant seul accepté comme réponse correcte), soit les mots acceptables (dans ce cas une liste de mots acceptables est donnée au correcteur). (ALTE, 1998: 241)

Tests de connaissances: montrent ce que les étudiants ont acquis des contenus linguistiques qui leur ont été enseignés. Pour ce faire, les tests de connaissances doivent porter presque exclusivement sur les contenus de cours et ce sont typiquement les tests terminaux utilisés dans les cours de langues. On les appelle également « tests d'acquisition ».

Tests de niveau de capacité: test[s] qui mesure[nt] la capacité ou l'aptitude générale, sans référence à quelque formation spécifique ou ensemble de matériel que ce soit. (ALTE, 1998: 241)

Tests de performance: ne portent pas sur un corpus linguistique prédéfini mais montrent la capacité du candidat à agir efficacement dans certaines situations. Les étudiants savent quoi étudier pour un test de connaissances alors que les tests de performance peuvent couvrir toutes sortes de compétences, sans tenir compte des études de langues antérieures des candidats. De nombreux tests de langues commerciaux à diffusion internationale sont des tests de performance.

Tests directs: test[s] qui mesure[nt] la capacité de production orale ou écrite, et où la performance est directement mesurée. Par exemple lors de l'évaluation de la production écrite lorsque la tâche du candidat est d'écrire une lettre. (ALTE, 1998: 242)

Tests indirects (tâches): test[s] ou tâche[s] à réaliser qui tente[nt] de mesurer les capacités sous-jacentes à une aptitude langagière, plutôt que de tester la performance au test lui-même. Un exemple est de tester la capacité de production écrite en demandant au candidat de noter dans un texte les structures incorrectes. (ALTE, 1998: 242)

Tests intégrés: peuvent tester plusieurs compétences à la fois ou la langue en général sans mettre l'accent sur, par exemple, la grammaire ou le vocabulaire. Un test de closure est un test intégré car afin de compléter un blanc, le candidat doit comprendre les indices contextuels aussi bien que syntaxiques.

Tests normatifs: test[s] dans lequel [lesquels] les notes sont interprétées par rapport à la performance d'un groupe donné composé de personnes comparables aux individus auxquels le test est administré. On tend à utiliser ce terme pour des tests dont l'objectif est de classer les individus en fonction du groupe de référence ou les uns par rapport aux autres. (ALTE, 1998: 242)

Tests objectifs / notés objectivement: ce sont typiquement les tests à choix multiples ou les tests vrai-faux avec une grille de réponses justes (normalement une par item) que chaque évaluateur utilise de la même manière.

(*Note: Il n'existe pas de tests de langues vraiment objectifs, car chaque épreuve repose sur la conception que l'auteur du test a de ce qui est important dans une langue et des compétences requises pour réussir une épreuve de langue.*)

Tests pronostiques: des épreuves de langues qui peuvent être utilisées pour prédire la réussite future d'un étudiant dans ses études de langues sont des tests pronostiques.

Tests réussite / échec: on les rencontre fréquemment à l'université, où de nombreux tests sont non pas notés sur une échelle mais réussis ou non. Dans une épreuve de langue, le point de césure et les limites entre les notes peuvent être décidées sur la base d'une évaluation normative ou d'une évaluation critériée.

Tests standardisés: un instrument de mesure qui a été piloté (généralement sur un échantillon large, représentant différents types de personnes interrogées) et pour lequel des données d'interprétation telles que des normes, des coefficients de fidélité et de validité ont été fournis; a été administré à un large groupe de candidats dans une population cible, souvent plus de 1 000 personnes, et a été analysé et normé pour être utilisé avec d'autres échantillons de la même population. (Traduction de la définition du terme « *standardised tests* » tirée du *Testing Glossary*, 2011)

Tests subjectifs / tests à barème subjectif: ce sont des tests où la personnalité de l'évaluateur ainsi que ses préférences personnelles jouent un rôle plus important parce qu'il n'y a pas de corrigé, et que les consignes écrites peuvent souvent être interprétées de plusieurs manières. C'est pourquoi, dans des épreuves à enjeux décisifs, les examinateurs suivent des sessions de formation avant de commencer à noter des tests sommatifs.

Tests: évaluation officielle d'une compétence, annoncée à l'avance et nécessitant une durée de passation substantielle (à savoir une heure ou plus). Par exemple, un test pourra inclure un extrait de texte à lire avec des questions, une partie grammaticale, et une série de mots à utiliser dans des phrases. (Traduction de la définition du terme « *tests* » tirée du *Testing Glossary*, 2011). Il s'agit aussi d'une procédure servant à évaluer le niveau de compétence ou de connaissance d'un individu.

- 1) Une série d'éléments dont l'ensemble constitue une procédure d'évaluation souvent équivalente à un examen.
- 2) Une tâche ou un élément unique servant à évaluer un domaine spécifique de compétence ou de connaissance: productions orale et écrite. Dans ce cas, le test ne sera qu'une des composantes (exemple: test de production orale) ou une des tâches (exemple: test de closure) d'un examen complet.

- 3) Une procédure d'évaluation simple et de passation facile, souvent conçue et administrée au sein d'un établissement (exemple: test de progrès). Elle peut également être utilisée dans le cadre d'un programme de recherche ou à des fins d'évaluation (exemple: test d'ancrage). (ALTE, 1998: 240)

Transparence: signifie que les tests sont décrits suffisamment en détail pour que tous ceux qui en traitent les résultats sachent ce qui a été testé et comment, et ce que signifie le résultat, c'est-à-dire ce que le candidat ayant obtenu une certaine note ou un certain score est capable de faire dans la langue cible.

Validation a posteriori: a lieu après que le test a été administré, est établie en utilisant toutes sortes de mesures et d'analyses.

Validation a priori: a lieu après que le test a été conçu mais avant qu'il soit utilisé. Elle repose sur l'étude et la pondération de théories antérieures sur l'évaluation, et sur l'utilisation de recherches antérieures portant sur des tests et des situations de tests similaires.

Validité: a) On peut démontrer qu'un test est valide par une validation a priori ou une validation a posteriori. Un test est valide s'il mesure ce qu'il est supposé mesurer et qu'il en convainc les utilisateurs. Il y a plusieurs sortes de validité.
b) Degré auquel les notes d'un test permettent de tirer des conclusions appropriées, significatives et utiles, en relation avec l'objectif du test. On distingue différents aspects de la validité, tels que la validité de contenu, la validité critérielle et la validité de construit; elles donnent différentes sortes de preuves permettant de juger la validité globale d'un test en fonction de ses objectifs. (ALTE, 1998: 244)

Validité apparente: qualité d'un test ou de toute autre mesure qui semble correcte et adéquate à l'objet mesuré. Il s'agit là d'un jugement subjectif plus que d'un jugement basé sur une analyse objective du test. La validité apparente est souvent considérée comme une fausse forme de validité. On l'appelle également « attrait d'un test » (*test appeal*). (ALTE, 1998: 244)

Validité concourante: on dit d'un test qu'il a une validité concourante si les notes obtenues sont en corrélation élevée avec un critère externe reconnu qui mesure le même domaine de connaissance ou de capacité. (ALTE, 1998: 244)

Validité consécutive: la manière dont la mise en place d'un test peut affecter l'interprétabilité des notes du test; les conséquences pratiques de l'introduction d'un test. (Traduction de la définition du terme « *consequential validity* » tirée de McNamara 2000: 132)

Validité de construit / de construct (validité conceptuelle): 1) On dit d'un test qu'il a une validité de construit si les notes obtenues peuvent être interprétées comme une théorie sur la nature d'un construit ou sur le rapport de ce construit avec

d'autres. On pourrait prédire, par exemple, que deux tests valides de compréhension orale classent les apprenants de la même façon, mais chacun d'entre eux aurait un rapport plus éloigné avec les notes obtenues à un test de compétence grammaticale. (ALTE, 1998: 245)

- 2) Elle est liée au cadre théorique qui sous-tend le test, c'est-à-dire si le test teste vraiment un construct tel que la compréhension écrite en LSP (langues à fins spécifiques).

Validité de contenu: montre si le test a mesuré ce qu'il était supposé mesurer, par exemple que le contenu d'un test de connaissances ne va pas au-delà de ce qui a été enseigné avant le test.

Validité externe: dépend de la représentativité du test et de la possibilité d'en généraliser les résultats.

Validité interne: porte sur le contenu du test, et en particulier sur le fait que le test lui-même ne contient pas d'éléments non pertinents qui dérangerait le candidat.

Validité prédictive: signifie que le test peut être utilisé pour prédire la réussite future du candidat dans l'usage et / ou l'apprentissage de cette langue ou d'une autre.

Références

ALTE (1998). *Multilingual glossary of language testing terms: Studies in language testing 6*. Cambridge: CUP.

McNamara, T. (2000). *Language testing*. Oxford et al.: Oxford University Press.

Testing Glossary (2011). Consultable sur le site internet:

<https://facultystaff.richmond.edu/~rterry/Middlebury/testgloss.htm>

[consulté le 9 septembre 2011].

Skehan, P. (1996). « Second language acquisition research and task-based instruction ». In: Willis, J. et Willis, D. (éd.). *Challenge and change in language teaching*, Oxford: Heinemann, 17-30.

Skehan, P. (2003). « Task-based instruction ». In: *Language Teaching* 36, 1-14.

Weideman, A. (2006). « Assessing academic literacy in a task-based approach ». In: *Language Matters* 37 (1), 81-101.

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

BELGIUM/BELGIQUE

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

Jean De Lannoy/DL Services
Avenue du Roi 302 Koningslaan
BE-1190 BRUXELLES
Tel.: +32 (0)2 538 43 08
Fax: +32 (0)2 538 08 41
E-mail: jean.de.lannoy@dl-servi.com
<http://www.jean-de-lannoy.be>

BOSNIA AND HERZEGOVINA/ BOSNIE-HERZÉGOVINE

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

CANADA

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

CROATIA/CROATIE

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

CZECH REPUBLIC/ RÉPUBLIQUE TCHÈQUE

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

DENMARK/DANEMARK

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

FINLAND/FINLANDE

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

FRANCE

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

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

GERMANY/ALLEMAGNE

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

GREECE/GRÈCE

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

HUNGARY/HONGRIE

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

ITALY/ITALIE

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

NORWAY/NORVÈGE

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

POLAND/POLOGNE

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

PORTUGAL

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

RUSSIAN FEDERATION/ FÉDÉRATION DE RUSSIE

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

SPAIN/ESPAGNE

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

Diaz de Santos Madrid

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

SWITZERLAND/SUISSE

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

UNITED KINGDOM/ROYAUME-UNI

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

UNITED STATES and CANADA/ ÉTATS-UNIS et CANADA

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

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

FR-67075 STRASBOURG Cedex

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

Evaluer par les tâches les langues à fins spécifiques à l'université: un guide

Johann Fischer, Catherine Chouissa, Stefania Dugovičová et Anu Virkkunen-Fullenwider

Cette publication s'adresse:

- aux responsables de l'enseignement et de l'évaluation en langues à l'université (directeurs des centres de langues universitaires ou des départements de langues, etc.),
- aux enseignants/rédacteurs de tests de langues à des fins spécifiques dans l'enseignement supérieur,
- aux formateurs d'enseignants,
- à toute personne impliquée dans la formation et l'évaluation en langues au niveau universitaire.

Ce manuel pratique est conçu pour les enseignants et rédacteurs de tests qui recherchent un outil valable pour mesurer, de manière pertinente, les capacités langagières de leurs étudiants. Il montre comment lier les savoir-faire enseignés et ceux qui seront nécessaires dans les études et la vie professionnelle. Il aide également les formateurs en langues, qui enseignent déjà dans une perspective actionnelle, à concevoir les tests correspondants et à évaluer les performances langagières de leurs étudiants. Par ailleurs, il souligne les avantages de l'évaluation par les tâches pour toutes les parties prenantes.

Pour obtenir de plus amples informations et accéder aux matériels accompagnant ce manuel, consultez le site internet <http://gult.ecml.at>.

Le Conseil de l'Europe regroupe aujourd'hui 47 Etats membres, soit la quasi-totalité des pays du continent européen. Son objectif est de créer un espace démocratique et juridique commun, organisé autour de la Convention européenne des droits de l'homme et d'autres textes de référence sur la protection de l'individu. Créé en 1949, au lendemain de la seconde guerre mondiale, le Conseil de l'Europe est le symbole historique de la réconciliation.



<http://www.coe.int>

<http://book.coe.int>
Editions du Conseil de l'Europe

12 € / 24 US\$



ISBN 978-92-871-7164-1

POUR L'EXCELLENCE DANS L'EDUCATION AUX LANGUES

Le Centre européen pour les langues vivantes (CELV) est un « Accord partiel élargi » du Conseil de l'Europe regroupant trente-quatre Etats membres¹. L'institution vise à rapprocher les théories issues des politiques linguistiques et les pratiques d'apprentissage en classe. A une époque de changement et de mobilité sans pareils, le Centre propose des approches concrètes permettant de répondre aux questions et aux défis posés aux sociétés culturellement diverses en Europe.

Le CELV cherche à opérer un changement positif pour les professionnels de l'éducation aux langues en :

- promouvant des approches innovantes ;
- développant la qualité de l'enseignement et de l'apprentissage des langues ;
- soutenant la mise en œuvre de politiques linguistiques éducatives ;
- encourageant le dialogue entre les praticiens et les décideurs dans le domaine de l'éducation aux langues.

Au sein du Conseil de l'Europe, les activités du CELV sont complémentaires à celles de la Division des politiques linguistiques, chargée de développer des politiques et des outils de planification en matière d'éducation aux langues, et à celles du Secrétariat de la Charte européenne des langues régionales ou minoritaires.

La présente série de publications est issue du programme d'activités 2008-2011 du CELV intitulé *Valoriser les professionnels en langues : Compétences – Réseaux – Impact – Qualité*. Le programme a été conduit dans un contexte marqué, dans le domaine de l'éducation, par des développements politiques majeurs sur la scène internationale, qui requièrent de la part des enseignants des qualifications professionnelles de plus en plus poussées. Le corps enseignant est censé contribuer aux processus de réforme de l'éducation nationale et répondre à un grand nombre de défis liés, entre autres, à l'application de normes dans l'enseignement, à l'évaluation axée sur les résultats, à une plus grande autonomie des institutions éducatives et à une hétérogénéité ethnique et culturelle croissante parmi les apprenants.

Les publications reflètent l'implication et l'engagement actif de tous ceux qui ont participé à une série de 24 projets internationaux, en particulier des équipes de coordination.

L'ensemble des publications du CELV et des matériels d'accompagnement est disponible sur le site internet <http://www.ecml.at/publications>.

¹ Les 34 Etats suivants sont membres de l'Accord partiel élargi du CELV : Albanie, Andorre, Arménie, Autriche, Bosnie-Herzégovine, Bulgarie, Croatie, Chypre, République tchèque, Estonie, Finlande, France, Allemagne, Grèce, Hongrie, Islande, Irlande, Lettonie, Liechtenstein, Lituanie, Luxembourg, Malte, Monténégro, Pays-Bas, Norvège, Pologne, Roumanie, République slovaque, Slovénie, Espagne, Suède, Suisse, « L'ex-République yougoslave de Macédoine », Royaume-Uni (situation au 30 juin 2011).