

Évaluation et autoévaluation en français langue étrangère : approche empirique du *Cadre européen commun de référence pour les langues* (CECR) du Conseil de l'Europe à travers le TEF et l'échelle d'autoévaluation DIALANG

DEMEUSE Marc, *Professeur*, Institut d'Administration scolaire, Université de Mons-Hainaut, Mons, Belgique, marc.demeuse@umh.ac.be

Institut d'Administration scolaire, Faculté de Psychologie et des Sciences de l'Education, Université de Mons-Hainaut, Académie universitaire Wallonie-Bruxelles, Place du Parc, 18, B-7000 Mons, Belgium.

ARTUS Frédérique, *Chercheuse*, Institut d'Administration scolaire, Université de Mons-Hainaut, Mons, Belgique, frederique.artus@umh.ac.be

CASANOVA Dominique, *Responsable pédagogique*, Chambre de Commerce et d'Industrie de Paris, Paris, France, dcasanova@ccip.fr

CRENDAL Alexandra, *Responsable pédagogique*, Chambre de Commerce et d'Industrie de Paris, Paris, France, acrendal@ccip.fr

DESROCHES Franck, *Responsable du TEF*, Chambre de Commerce et d'Industrie de Paris, Paris, France, fdesroches@ccip.fr

RENAUD François, *Responsable pédagogique*, Chambre de Commerce et d'Industrie de Paris, Paris, France, frenaud@ccip.fr

Introduction

La mise en équivalence des niveaux linguistiques en langue seconde dans différentes langues est aujourd'hui une question politique et économique importante qui touche aussi bien le problème du recrutement de fonctionnaires fédéraux bilingues en Belgique que l'accès à l'immigration économique dans un pays bilingue comme le Canada (Citoyenneté et Immigration Canada, 2002). A côté du développement d'outils d'évaluation critériés à finalité certificative, il importe, principalement lorsque les résultats comportent de forts enjeux pour les candidats (Engle & Engle, 2004), de leur permettre d'avoir à la fois une bonne compréhension de la manière dont leurs compétences sont appréciées et une bonne capacité à s'auto-évaluer en regard du référentiel, en particulier le *Cadre européen commun de référence pour les langues* (CECR) du Conseil de l'Europe (2000).

Le CECR constitue en effet un outil de référence de plus en plus important et qui devient même incontournable dans le secteur des langues étrangères, du moins en Europe. Il constitue la référence aussi bien en matière d'élaboration de tests de langue critériés (Conseil de l'Europe, Division des politiques linguistiques, 2003) que dans le domaine de la production de manuels d'apprentissage de langues étrangères. Les tests de français langue étrangère s'y réfèrent et le *Test d'Evaluation de Français* (TEF)¹ y est donc lié. Une grille de correspondance permet de faire concorder les scores du TEF répartis en sept niveaux de compétences avec les six niveaux de l'échelle du CECR (Noël-Jothy & Sampsonis, 2006).

¹ Le *Test d'Evaluation de Français* (TEF) a été mis au point par la *Chambre de Commerce et d'Industrie de Paris* dès 1998 - <http://www.fda.ccip.fr>.

Le CECR sert également de référence aux différents modèles nationaux de portfolios européens des langues. Ces derniers présentent aux apprenants d'âges et de milieux variés, des propositions leur permettant de faire une autoévaluation de leurs connaissances et savoir-faire dans quatre compétences communicatives (compréhension orale, compréhension écrite, expression orale, expression écrite)².

DIALANG³, premier système majeur d'évaluation mis gratuitement à disposition sur internet depuis mars 2004, est par ailleurs un outil entièrement conçu à partir du CECR et développé par plus de 20 universités européennes avec le soutien financier de la Commission européenne, Direction-Générale de l'Éducation et de la Culture (Programme SOCRATES, LINGUA Action D). Cet outil permet d'établir un diagnostic équivalent dans 14 langues de l'UE : l'allemand, l'anglais, le danois, l'espagnol, le finnois, le français, le grec, l'irlandais, l'islandais, l'italien, le néerlandais, le norvégien, le portugais, le suédois. DIALANG couvre cinq compétences : l'écoute, la lecture, la rédaction, la grammaire et le vocabulaire.

Ce test n'a pas de vocation certificative (Danish Technological Institute, Nov. 2004) mais il fournit des résultats et commentaires dès la fin du passage du test. Ce test est précédé par deux étapes (pré-tests) qui sont proposées aux candidats afin de procéder à une première estimation de leurs compétences et de choisir le niveau de difficulté adéquat pour le test qui en présente trois. Ce pré-positionnement des candidats comporte deux parties distinctes :

1. Un test de reconnaissance des mots (Test de Niveau sur l'étendue du Vocabulaire) : une liste de mots (verbes) qui doit être validée dans la langue cible (existe ou n'existe pas). Ce pré-test permet de poser un premier diagnostic, néanmoins assez sommaire, sur le niveau.
2. Une deuxième étape est un questionnaire d'autoévaluation qui permet d'apprendre à évaluer soi-même ses compétences acquises au cours de l'apprentissage d'une langue étrangère. Cette activité est disponible pour les compétences de lecture, d'écoute, ou de rédaction.

La partie autoévaluation de DIALANG est une composante centrale de l'outil, non seulement dans la phase de pré-positionnement mais également dans le système de rapport des résultats et du test en lui-même (Alderson, 2005a). Les descripteurs proviennent en grande majorité du CECR, certains ayant été adaptés, d'autres rajoutés, selon les besoins spécifiques de l'épreuve. Tous ont été calibrés par le premier Centre de Coordination du projet DIALANG (96-99) en Finlande dont les résultats ont démontré des indices de fidélités très élevés (North, 2000 et Kaftandjieva, 2002).

La philosophie qui sous-tend l'intérêt de développer chez l'apprenant une auto-évaluation de ses capacités linguistiques et communicatives est que l'autoévaluation non seulement permet une meilleure compréhension de son niveau de compétences, mais contribue également à évaluer ses propres besoins de formation ultérieurs et donc au développement de son autonomie. Dans la démarche du test diagnostique de DIALANG qui propose de faire le lien entre les résultats de l'autoévaluation et ceux du test proprement dit, le but poursuivi n'est pas de démontrer les possibles disparités, voire de suggérer l'imprécision de l'autoévaluation, mais bien d'aider l'apprenant à prendre conscience des éventuelles divergences, d'en comprendre les raisons et de l'aider à poursuivre sa réflexion sur les implications de celles-ci. (Alderson, 2005a).

² Ces modèles de portfolio, tout en étant calibrés sur le CECR, présentent des descripteurs ciblés en fonction du public auquel ils sont adressés et doivent être validés par le Conseil de l'Europe pour être homologués.

³ <http://www.dialang.org/project/francais/index.html>

La recherche en cours

Dans le domaine du français langue étrangère, les auteurs qui contribuent à l'élaboration du *Test d'Évaluation de Français* (TEF) ont souhaité s'assurer de la bonne compréhension des informations fournies aux candidats à l'issue d'un test à forts enjeux⁴ (Demeuse *et al*, 2004 et 2005). Pour ce faire, l'idée qui a été poursuivie est de mettre en relation les résultats obtenus au test critérié (TEF) et la manière dont s'apprécient les candidats, partant du principe que les résultats au test seront d'autant mieux compris que la manière dont les candidats s'auto-évaluent correspond à ces résultats. L'échelle d'autoévaluation mise au point dans le cadre du projet DIALANG a été adoptée dans cette expérimentation. Cette échelle est présentée dans la langue maternelle des candidats de manière à éviter les problèmes de compréhension et de vocabulaire auprès des candidats les plus faibles.

Les résultats

Une première expérimentation a été menée auprès de 36 candidats hispanophones, inscrits dans un même centre de langue agréé en France, avant une passation de la version électronique du TEF, baptisée e-TEF (Desroches *et al*, 2005). L'autoévaluation a été proposée dans sa version courte (utilisée dans les tests pilotes en compréhensions écrite (CE) et orale (CO)) de DIALANG (Alderson, 2005b). Le niveau de chaque candidat, sur l'échelle en six niveaux du CECR, a alors été calculé à partir de l'échelle d'autoévaluation d'une part et du TEF d'autre part, puis comparé. Cette approche n'a pas conduit à des résultats satisfaisants, au sens où la concordance entre évaluation et autoévaluation est très limitée (Desroches *et al*, 2005). Le manque de correspondance constaté entre ces deux instruments, dans une perspective d'autodiagnostic, met en évidence les difficultés qui peuvent subsister face à un référentiel lorsqu'il doit être maîtrisé par des usagers qui ne sont pas des experts du domaine linguistique, mais qui possèdent néanmoins une certaine compétence dans au moins une autre langue que leur langue maternelle.

De manière à mieux comprendre les problèmes posés par la mise en correspondance des compétences « objectivées » par un test et des compétences telles que les sujets les évaluent eux-mêmes, l'expérience est poursuivie. Celle-ci porte sur une extension de l'échantillon à d'autres locuteurs qu'hispanophones et sur la prise en compte de l'échelle DIALANG dans sa version longue et non plus dans sa version réduite, soit 36 items en compréhension écrite (CE) et 44 en compréhension orale (CO). L'échantillon hétérogène se compose de 6 autoévaluations en anglais dont 4 non-anglophones, 5 allemands, 6 espagnols et 1 portugais. Les candidats ont passé pour la plupart l'autoévaluation dans leur propre langue, excepté lorsque la langue maternelle n'était pas disponible parmi les 14 langues proposées par DIALANG, en l'occurrence pour quatre candidats. Le protocole est le même que pour la première expérimentation (Desroches *et al*, 2005).

Pour rappel, la consistance interne de chaque outil est assez élevée, avec un α de Cronbach de 0,80 (CE) et de 0,82 (CO) pour DIALANG, et de 0,80 (CE) et 0,67 (CO) pour le TEF. Par contre, les corrélations entre les résultats obtenus aux épreuves du TEF et le nombre de réponses positives à l'autoévaluation sont toujours positives, mais faibles ($r_{36}=0,14$ en CE et $r_{44}=0,30$ en CO) et donc non significatives ($p<0,51$ en CE et $p<0,35$ en CO).

⁴ Le TEF est en effet, notamment, reconnu par Citoyenneté et Immigration Canada et est utilisé pour l'admission officielle des étudiants non francophones dans les premiers cycles des universités françaises.

Si l'on compare nos résultats à ceux recueillis par Alderson (2005b) dans une étude confrontant l'autoévaluation aux épreuves du test DIALANG en langue anglaise sur un échantillon plus grand (n=718 en CE et n=606 en CO), les corrélations ne sont pas plus satisfaisantes. Elle s'élève à 0,54 en CE entre l'autoévaluation globale⁵ et l'épreuve elle-même et est encore plus modeste entre l'autoévaluation par habilité et l'épreuve équivalente en CE (r=0,49). Concernant la CO, la corrélation est de 0,47 entre l'autoévaluation globale et l'épreuve correspondante, elle est du même ordre de grandeur (r=0,49) entre l'autoévaluation par habilité en CO et l'épreuve (Alderson, 2005a).

La deuxième expérience ayant été basée sur la version longue d'autoévaluation proposée par DIALANG alors que la première expérience reprenait la version courte, nous avons repris les items d'autoévaluation communs aux deux versions afin de les comparer sur un ensemble identique d'items d'autoévaluation. Dans la version longue de DIALANG, les corrélations r_{36} et r_{44} correspondent aux corrélations entre les 36 items pour la CE et 44 items pour la CO de la version longue et les épreuves du TEF correspondantes. Ces corrélations sont comparées aux corrélations respectives (r_{18}) des 18 items de la version courte de DIALANG et les résultats aux épreuves correspondantes du TEF. Les résultats restent globalement stables :

- pour la compréhension écrite (CE) : $r_{36} = 0,14 \rhd r_{18} = 0,16$
- pour la compréhension orale (CO) : $r_{44} = 0,30 \rhd r_{18} = 0,24$.

L'échantillon sélectionné pour la deuxième expérience présente des candidats d'origines plus variées que lors de la première, réalisée sur un échantillon homogène (composé exclusivement de candidats hispanophones). Malgré cette plus grande variété de langues, si on prend en compte l'accroissement du nombre d'items dans la seconde version des autoévaluations, la dispersion des résultats obtenus n'est pas beaucoup plus large que dans la première expérimentation en CE (écart type de 1,73 pour la première contre 4,11 pour la deuxième). La différence se creuse en CO (écart type de 2,38 pour la première contre 7,24 pour la deuxième).

Ces résultats ne sont donc pas plus satisfaisants et les conclusions se maintiennent à propos de la faible liaison statistique observée entre les deux outils (TEF et DIALANG), pourtant liés au même référentiel. Les candidats apparaissent, comme relevé par Alderson, ne pas avoir une bonne perception de leurs propres compétences, en regard des évaluations « objectives » qui leur sont administrées. Cette conclusion, qui se maintient lorsque l'échantillon est plus large, pose bien les limites du recours à des descripteurs tels que ceux du CECR lorsque les usagers ne maîtrisent pas bien le référentiel dans son ensemble.

⁵ L'étude pilote de DIALANG inclut une autoévaluation globale et une autoévaluation par habilité ; seule la seconde sera retenue dans la version mise en ligne (Alderson, 2005a).

Les perspectives

Les résultats des deux expérimentations nous conduisent à réfléchir à la manière de mieux communiquer les résultats aux candidats et à les informer, lors de la formation linguistique, du référentiel qui est utilisé pour les positionner sur l'échelle de niveaux pour chacune des compétences linguistiques et communicatives. Si dans le cadre de DIALANG, test sans enjeu et à vocation purement diagnostique, l'autoévaluation est utilisée pour alimenter les commentaires et les conseils dans une perspective de formation, les divergences de jugement et de niveau font partie intégrante de la démarche diagnostique et se justifient dans le choix d'une orientation formative. Par contre, le recours à l'autoévaluation dans le cadre d'un test à visée certificative tel que le TEF dont la fonction principale n'est pas, *a priori*, formative, doit être interrogé : l'outil d'autoévaluation est-il le plus adéquat dans une démarche de compréhension et d'interprétation des résultats du TEF ? Ne faudrait-il pas plutôt étayer les commentaires fournis sur l'attestation afin de préciser les liens entre le score obtenu, le niveau sur l'échelle du CECR auquel il correspond et ce que cela traduit en termes de compétences et d'habilités ? Le recours à l'autoévaluation ne se justifierait alors sans doute que lorsque le TEF est utilisé comme test de positionnement dans le cadre d'un plan de formation.

Par ailleurs, une réflexion approfondie à propos des items d'autoévaluation, de l'utilisation de la langue maternelle ou non, ou encore de la réceptivité et compréhension de l'outil par ses utilisateurs, mérite d'être poursuivie. Les travaux rapportés par Alderson (2005a) appellent au recueil de données empiriques supplémentaires d'une part, et d'autre part à une réflexion sur l'amélioration de l'outil d'autoévaluation de DIALANG sans en altérer le lien au référentiel européen. Nos conclusions vont dans la même direction.

Bibliographie

ALDERSON, J.C. (2005a) : *Diagnosing Foreign Language Proficiency – The Interface between Learning and Assessment*, London & New York, Continuum.

ALDERSON J.C., HUHTA A. (2005b) : *The Development of a suite of computer-based diagnostic tests based on the Common European Framework*, Language Testing 2005-22 (3)

CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS (2005) : *Le Test d'Évaluation de Français de la Chambre de Commerce et d'Industrie de Paris*. Paris : Chambre de Commerce et d'Industrie de Paris, Centre de Langue de la Direction des relations internationales de l'Enseignement.

CITOYENNETE ET IMMIGRATION CANADA (2002) : *Standards Linguistiques Canadiens 2002*.

CONSEIL DE L'EUROPE (2000) : *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*, Paris, Didier.

CONSEIL DE L'EUROPE, DIVISION DES POLITIQUES LINGUISTIQUES (2003) : *Relier les examens de langue au Cadre européen de référence pour les langues : apprendre, enseigner, évaluer (CECR) – Manuel avant-projet*, Strasbourg : DGIV/EDU/LANG.

DEMEUSE, M., CRENDAL, A., DESROSCHE, F., OSTER, P., RENAUD, F., LEROUX, X. (2004) : *L'évaluation des compétences linguistiques des adultes en français langue étrangère dans une perspective de multi-référentialisation. L'exemple du Test d'Évaluation de Français (TEF) de la Chambre de Commerce et d'Industrie de Paris (CCIP)*, In Actes du 17^e colloque international de l'Association pour le Développement des Méthodologies d'Évaluation en Éducation (ADMEE-Europe), Lisbonne, 18-20 novembre 2004

DEMEUSE, M., DESROSCHE, F., CRENDAL, A., RENAUD, F., CASANOVA, D., ARTUS, F. (2005) : *La fiabilité de l'évaluation des compétences linguistiques pour des adultes non francophones : présentation d'un protocole d'évaluation*, In Actes du 18^e colloque international de l'Association pour le Développement des Méthodologies d'Évaluation en Éducation (ADMEE-Europe), Reims, 24-26 octobre 2005.

DESROSCHE, F., CASANOVA, D., CRENDAL, A., RENAUD, F., MAHY, C., DEMEUSE, M. (2005) : *Mise en équivalence des versions papier-crayon et électronique d'un test de français langue étrangère et mise en perspective des résultats par une procédure d'autoévaluation*, In Actes du 18^e colloque international de l'Association pour le Développement des Méthodologies d'Évaluation en Éducation (ADMEE-Europe), Reims, 24-26 octobre 2005.

ENGLE, L., ENGLE, J. (2004) : *Assessing Language Acquisition and Intercultural Sensitivity Development in Relation to Study Abroad Program Design*, *Frontiers*, in The Interdisciplinary Journal of Study Abroad, Volume X, Carlisle, PA.

KAFTANDJIEVA F., TAKALA S. (2002) : *Council of Europe scales of language proficiency: a validation study*. In Alderson, 2002, *Common European Framework of References for languages (2002): Learning, Teaching, Assessment - Case studies* : pp.106-29, Strasbourg, Council of Europe.

NOËL-JOTHY F., SAMPSONIS B. (2006) : *Certifications et Outils d'évaluation en FLE*, Paris, Hachette.

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

Sites Internet

Citoyenneté et Immigration Canada: <http://www.cic.gc.ca/french/skilled/qual-3-1.html>

Chambre de Commerce et d'industrie de Paris : <http://www.fda.ccip.fr>

Conseil d'Europe: <http://www.coe.int>

"Defining a Strategy for the Direct Assessment of Skills", Danish Technological Institute (Nov.2004): <http://europa.eu.int/comm/education/doc/reports/doc/skills.pdf>

DIALANG: <http://www.dialang.org/project/francais/index.html>