



HAL
open science

La démarche qualité chez les certificateurs européens en langues étrangères. Lieu de rencontres, curricula cachées et représentations stéréotypées

Patrick Riba

► To cite this version:

Patrick Riba. La démarche qualité chez les certificateurs européens en langues étrangères. Lieu de rencontres, curricula cachées et représentations stéréotypées. Raisons, comparaisons, éducations : la revue française d'éducation comparée, 2013, L'enseignement des langues étrangères face aux évolutions des systèmes éducatifs, 1 (9), pp.75-86. halshs-01395787

HAL Id: halshs-01395787

<https://shs.hal.science/halshs-01395787>

Submitted on 14 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*La démarche qualité chez les certificateurs
européens en langues étrangères. Lieu de
rencontres, curricula cachés et
représentations stéréotypées*

P. Riba

Université des Antilles et de la Guyane

Les procédures que s'imposent les membres de l'Association des certificateurs européens en langue (Alte), dans le cadre de leur démarche qualité, ont mis en évidence la dichotomie de cultures évaluatives distinctes. Cet article interroge les tensions méthodologiques inscrites dans une démarche qui semble ne pas être capable aujourd'hui de conjuguer inférence de compétences actionnelles en contexte et mesure de la validité et de la fiabilité, tout en soulignant qu'aucun des organismes en présence n'ignore cependant la préoccupation des autres, rendant sans doute possible le développement de recherches communes en ce sens.

Les certificateurs européens en langues, regroupés au sein de l'Association of Language Testers in Europe (Alte), ont mis en place il y a quelques années une démarche qualité visant à garantir la validité des certifications proposées par chacun de ses membres.

Cette démarche s'inscrivait dans une période de croissance du marché des certifications en langues, les tests et diplômes proposés par les membres de cette association étant généralement retenus comme prérequis pour de nombreuses formalités administratives, comme c'est le cas en France pour l'aide à l'insertion avec le diplôme initial de langue française (Dilif) pour la régulation de la mobilité internationale des étudiants ressortissants hors Union européenne, avec le test de connaissance du français pour la demande d'admission préalable, TCF DAP, ou encore l'accès à la nationalité avec le TCF éponyme, TCF-AN.

Face à cette tendance, ces organismes, le Centre international d'études pédagogiques (Ciep) et la chambre de commerce et d'industrie de Paris (CCIP), pour la partie française, l'Université de Cambridge et Trinity College pour l'anglais, l'Instituto Cervantes pour l'espagnol ou encore le Goethe Institute pour l'allemand, pour n'en citer que quelques-uns (la liste complète des 27 membres du groupe est consultable sur le site www.alte.org), ont donc décidé en 2005 d'harmoniser leurs procédures et leurs standards afin de se doter d'une instance régulatrice interne d'évaluation et de labellisation par les pairs, qui verra le jour en 2007, le Alte Q-mark Label. Depuis cette date, de nombreux audits ont été effectués par des comités d'experts de l'association, et plusieurs certifications ont reçu ce label, d'autres étant en cours d'évaluation.

(www.alte.org/setting_standards/the_alte_q_mark)

Cette initiative de normalisation de la certification en langues étrangères en Europe est à bien des égards un lieu de rencontre intéressant qui cristallise dans sa finalité même nombre de principes éducatifs et didactiques de pays différents. Sans prétendre pénétrer le champ désormais constitué de la comparaison en éducation (Groux D., Chane-Davin F., 2009, pp. 11-15), (Schubauer-Leoni, 2009, pp. 15-30), nous tenterons d'illustrer comment, au travers des choix méthodologiques retenus, les auditeurs ont plébiscité de manière implicite certaines approches didactiques aux dépens d'autres, au risque d'impacter, pour le meilleur et pour le pire, l'évaluation et donc l'enseignement des langues étrangères. Sans remettre nullement en cause la nature de la procédure elle-même, nous émettons le vœu que cette contribution à la *Revue Française d'Éducation Comparée*, consacrée à l'enseignement des langues étrangères face aux évolutions des systèmes éducatif et médiatique, attirera l'attention de spécialistes en éducation comparée, qui sauront, par leur approche, guider et orienter les spécialistes en évaluation certificative.

Quelques rappels

L'association Alte a été créée en 1989 à l'initiative des universités de Cambridge, Grande-Bretagne, et Salamanque, Espagne, afin d'établir des standards communs pour la certification en langues étrangères en Europe, l'association parlant d'ailleurs dans sa constitution, uniquement rédigée en anglais, de « testing »¹⁷. Elle est aujourd'hui constituée de trente-trois organismes européens et d'une quarantaine d'organismes observateurs, représentant plusieurs dizaines de langues, du basque au gallois (par ordre alphabétique), et de pays européens.

Rappelons que l'Université de Cambridge a toujours été pionnière dans ce domaine, puisque c'est probablement elle qui a instauré la première certification en langues, le Certificate of Proficiency in English qui, dès 1913, visait à certifier le seuil de maîtrise en anglais langue étrangère des enseignants non Britanniques. Sa revue, *Research Notes*, reste l'une des meilleures références internationales dans le domaine de l'évaluation, et son département Esol Examination est l'un des plus importants certificateurs d'anglais de la planète, avec plusieurs dizaines de certifications proposées dans plus de 140 pays. C'est le directeur de ce département, M. Milanovic, qui a créé Alte, qu'il préside depuis sa création. Comme le précise l'annexe B du Cadre européen commun de référence pour les langues du Conseil de l'Europe, désormais le Cadre, les premières échelles de compétence d'Alte, Can Do Statements, ont par ailleurs fortement inspiré les concepteurs des échelles du Cadre, qui connaissent aujourd'hui un succès majeur.

Concernant la France, le Ciep adhéra à Alte en 1992, et la CCIP, en 2008.

L'histoire et le fonctionnement d'Alte sont donc indissociables de celui d'Esol Examination, et par son action à la fois sur les certifications elles-mêmes, mais aussi

¹⁷ « *Recognising the need for a coherent approach to language testing, ALTE was established in 1989 by the Universities of Cambridge and Salamanca* », www.alte.org/about_alte

sur son influence auprès des instances dirigeantes européennes, cette association a eu et continue d'avoir une forte influence dans la certification en langues étrangères en Europe. C'est, par exemple, sous la direction de Cambridge Esol Examination qu'un certain nombre de membres d'Alte, réunis en consortium (dont le Ciep) qu'a été réalisée pour le compte de l'Union européenne une enquête sur le niveau en langue étrangère des étudiants âgés de 15 ans, SurveyLang, dont les résultats viennent d'être publiés (cf. sitographie).

Quelques (im)précisions lexicales

La plupart des membres d'Alte proposent des tests et des diplômes ; c'est ainsi par exemple que les deux organismes français présentent leur offre de certifications. La différence entre ces deux types d'examens à visée évaluative, utilisés, dans le cadre d'Alte, à des fins sommatives et certificatives, semble aller de soi. Aucun des deux organismes ne l'explique cependant au grand public, et elle n'est d'ailleurs pas si simple.

Dans le cas qui nous intéresse, on parle de test lorsque le candidat est soumis à une ou plusieurs séries d'exercices basés sur un continuum de difficulté croissante ou décroissante permettant de positionner le candidat sur une échelle donnée (de A1 à C2, par exemple, comme c'est le cas pour le test de connaissance du français, TCF, ou pour le test d'évaluation de français, TEF). Ce diagnostic peut donner lieu à la délivrance d'une attestation de niveau, qui devient souvent caduque au bout d'un certain temps. Remarquons que les épreuves et les attestations qui en découlent peuvent interpréter de façon distincte le niveau des candidats pour chacune des activités langagières (compréhensions et productions orales ou écrites), rendant compte de manière explicite de l'hétérogénéité des profils des locuteurs, puisque l'on peut obtenir un niveau A1 en production écrite et B1 en compréhension orale, par exemple.

Le mot diplôme est quant à lui plus polysémique, puisqu'il désigne, en français tout au moins, à la fois le titre délivré par une autorité qui atteste d'un degré d'aptitude ou de connaissance, mais aussi l'examen ou le concours permettant d'obtenir ce titre. En général, l'épreuve est proposée sur un seul niveau (par exemple B2) qu'elle sanctionne de façon binaire, réussite/échec, et elle est composée d'une ou de plusieurs séries d'exercices ouverts et/ou fermés.

Tests et diplômes se distinguent donc par rapport au document final remis ; les deux attestent et certifient, seul le deuxième pouvant titrer. Ils se distinguent aussi par leur structure, basée pour le test sur un continuum, et pour le diplôme sur un seul niveau donné. Notons cependant que l'un et l'autre peuvent recourir à des exercices ouverts et/ou fermés, voire à des tâches. Pourtant, on considère généralement que le test fait appel à des exercices moins ouverts que l'examen du diplôme, ceci étant peut-être dû à un trait sémiologique plus scientifique du test, le diplôme relevant davantage du domaine académique. En soi, cette conception est cependant non avérée, puisque l'on trouve des exercices fermés dans les épreuves de nombreux diplômes (épreuves de compréhension du Delf, par exemple), et que les tests ne rechignent pas à donner une place à des épreuves ouvertes (production orale du TCF, par exemple). Ceci dit, il est vrai que les tests ont été conçus dès l'origine avec un souci explicite et technique de fiabilité de la notation, ce qui n'était pas forcément le cas des diplômes.

Il convient enfin de rappeler que, jusqu'à l'apparition du TCF et du TEF, la France n'avait pas de test de français langue étrangère. L'Alliance française de Paris offrait depuis de nombreuses années ses diplômes, le Ciep ayant créé en 1985 pour sa part trois diplômes, les Delf 1^{er} et 2^e degrés, et le Dalf. La chambre de commerce de Paris et le Centre international d'études pédagogiques sont donc allés chercher l'expertise pour la conception des tests chez des partenaires européens, Belgique et Grande-Bretagne, des

pays dans lesquels la culture du test est traditionnellement plus répandue. Le succès de ces certifications et les demandes de la société à leur égard prouvent que l'initiative répondait bel et bien à de véritables besoins, complémentaires des besoins en diplômes puisque le Delf et le Dalf n'ont par ailleurs jamais cessé de se développer également.

La démarche qualité d'Alte

En 2007 donc, Alte s'est dotée d'un code de bonnes pratiques, afin de « rendre explicites les normes qu'ils entendent respecter et reconnaître les obligations qui prévalent dans leur activité. [...] Les certificateurs d'Alte s'engageant à protéger les droits des candidats en s'efforçant de répondre aux normes de ce code dans quatre domaines :

- la conception des examens de langue ;
- l'interprétation des résultats des examens ;
- la recherche de l'équité ;
- l'information des candidats ».

Pour ce faire, Alte a établi 17 « normes minimales relatives à l'établissement de profils qualité pour les examens d'Alte » dans chacun de ces domaines, allant par exemple de l'explicitation des concepts hypothétiques des examens au respect de normes dans la communication des résultats aux candidats. C'est sur ces bases, et selon une procédure d'audit très codifiée, que des experts membres d'Alte, membres d'organismes certificateurs, analysent une ample documentation fournie par l'organisme audité, se rendent sur place et vérifient, sur la base de check-lists, les assertions de l'organisme audité pour lui délivrer, ou non, le label Q-Mark. Cette procédure s'applique donc à toutes les certifications proposées par les pays membres, diplômes et tests, et ses vertus contraignantes offrent d'indéniables garanties de sérieux et de professionnalisme à des candidats dont nous avons vu qu'ils choisissaient souvent de se soumettre à ces certifications par nécessité.

Un point aveugle, l'analyse du test

Il est cependant intéressant d'analyser certaines difficultés de cette entreprise, qui s'expliquent peut-être par ces références à des cultures évaluatives étrangères que nous évoquions précédemment.

Les normes 13 et 14 portent ainsi sur l'analyse des épreuves (mais la traduction française du document parle, de manière significative, d'analyse du test).

« 13. Vous collectez et analysez des données sur un échantillon de candidats adéquat et représentatif et vous êtes certain que leur note reflète fidèlement les compétences mesurées à l'examen et n'est pas influencée par des facteurs tels que la L1, le pays d'origine, le genre, l'âge et l'origine ethnique.

14. Les données concernant les niveaux des items (ex. pour calculer la difficulté, la discrimination, la fiabilité et les erreurs standard d'évaluation de l'examen) sont collectées à partir d'un échantillon adéquat de candidats et analysées. »

Ces deux standards reposent sur des données bien connues en évaluation, la validité (de construct) et la fiabilité, qui ont toujours été une préoccupation majeure chez les concepteurs d'épreuves, y compris en France. La nouveauté imposée ici réside dans l'obligation qui est faite de devoir le prouver. Car si rien n'est dit sur la façon dont on parviendra à s'assurer « que la note reflète fidèlement la compétence mesurée », les auditeurs en interne se sont naturellement mis d'accord sur les procédures à suivre, en s'appuyant d'ailleurs sur une littérature qui s'imposait parallèlement autour du Cadre européen – et à laquelle certains d'entre eux avaient d'ailleurs largement contribué.

De nombreuses techniques éducatives issues du « testing » anglo-saxon avaient en effet été utilisées pour étalonner les descripteurs du Cadre (Conseil de l'Europe, 2001, annexe B). Pressentant la nécessité d'asseoir une méthodologie qui guiderait l'utilisation du Cadre par les certificateurs, sans avoir nécessairement à recourir à des

techniques que peu d'entre eux possèdent, la division des Politiques linguistiques du Conseil de l'Europe avait publié (Conseil de l'Europe, 2003) un « *guide pour relier les examens au Cadre européen* », qui permettait d'illustrer une corrélation de fait entre les niveaux du Cadre et les examens au moyen de listes de contrôle et d'analyse ; mais très vite cependant, ce guide fut complété par un supplément théorique qui redonna à l'approche psychométrique une place prépondérante (Conseil de l'Europe, 2008).

Il s'agit principalement d'une approche centrée sur la notion de mesure, qui tente de justifier l'attribution d'une valeur chiffrée à une performance donnée, afin d'inférer une compétence ou un savoir. En langue, elle consiste à soumettre chaque composante de l'épreuve à une série de prétests afin d'en valider les traits internes à partir des résultats obtenus sur une population supposée représentative. Ces tests doivent permettre de vérifier par exemple si l'on peut identifier de manière incontestable le construit mesuré dans l'épreuve proposée (validité de contenu ou validité apparente), si celle-ci possède des qualités discriminantes suffisantes pour distinguer les construits s'il en révèle plusieurs (validités discriminante et convergente), ou encore si différents jurys attribuent tous une même note à une même performance (fiabilité).

Elle nécessite la mise en place d'une logistique complexe pour organiser les prétests, d'un savoir-faire précis en traitement des données (outils informatiques et compétences théoriques), et d'une capacité à la mise en banque des exercices validés qui permet de rentabiliser le processus de conception. Le Ciep s'est doté des moyens nécessaires pour réaliser ces études concernant le TCF, et il travaille à leur mise en place pour le Delf, mais de nombreux organismes ont fait le choix de sous-traiter ces études à d'autres organismes partenaires comme Cambridge Esol ou Cito. Dans la maîtrise de ces nouvelles procédures, l'outil psychométrique constitue donc une indéniable valeur ajoutée qui modifie de facto le rapport de forces entre partenaires.

Mais au-delà de ces considérations, cette approche implique de ne retenir que des typologies d'épreuves adaptées aux techniques d'interprétation utilisées. C'est la raison pour laquelle, pendant de nombreuses années, les tests étaient constitués en majorité d'exercices fermés (questions à choix multiples, appariements, etc.), ne laissant qu'une faible part à des questions ouvertes (argumentation, etc.). Il n'était d'ailleurs pas rare de voir ces tests se limiter à la mesure de compétences en compréhension et en maîtrise des formes, toute compétence en production semblant, par nature, plus rebelle à l'observation de traits isolés. Même si les techniques évoluent aujourd'hui, cette approche conditionne une certaine défiance vis-à-vis de la production orale en particulier. À l'heure où l'approche actionnelle plébiscite la mise en place de tâches dans l'enseignement-apprentissage, peu de certifications ont osé franchir le pas, et celles qui le font semblent s'inscrire en faux par rapport à l'approche qualité prônée par Alte.

De fait, tous les tests soumis à la procédure qualité ont obtenu le label Q-Mark, ce qui n'est pas le cas de tous les diplômes.

En France, la large place laissée dans le Delf aux épreuves de production (50 % du total de l'examen), n'a pas permis pour l'instant au Ciep de parvenir au seuil fixé par Alte pour rendre compte des points 13 et 14. Si cet organisme veut y parvenir, il devra certainement revoir le format de ces épreuves. Il a, en revanche, facilement atteint les standards d'exigence de l'association pour le TCF.

Des préoccupations légitimes mais différentes

En simplifiant, nous pourrions dire que deux cultures didactiques se rencontrent, chacune tentant d'ailleurs d'aller vers l'autre sans abandonner ce qui constitue sa propre valeur ajoutée. D'un côté, la prééminence d'un louable souci de justice vis-à-vis des candidats pousse des certificateurs (plutôt anglo-saxons, même si cette schématisation accepte de nombreux contre-exemples) à ne

proposer que des épreuves pour lesquelles des études statistiques permettent à la fois de rendre observables les traits saillants évalués et de renforcer ainsi la corrélation inter jurys. De l'autre, des tenants d'épreuves plus complexes à observer et interpréter, mais sans doute plus représentatives d'une approche actionnelle sollicitant à la fois des compétences de production et de compréhension, où la maîtrise des formes ne constitue plus que l'un des traits observés.

Dans la culture éducative française, les écarts de notes semblent assumés comme une fatalité. Dans le cas du sacrosaint baccalauréat, par exemple, véritable rite initiatique français au poids social si fort, l'Institut de recherche sur l'éducation, Iredu, relève, depuis de nombreuses années, l'arbitraire des notations (Suchaut, 2008) sans que le format des épreuves ou la formation des évaluateurs ne soient véritablement remis en question ; pourtant, une même copie peut obtenir de 3/20 à 18/20, selon les jurys. Tant d'articles ont été publiés sur ce sujet qu'il semble presque indécent de le rappeler, et la création de comités d'harmonisation n'a pas réglé un problème... qui ne semble pas en être un d'ailleurs pour les enseignants. Car pour eux, remplacer une épreuve de type argumentatif par un exercice moins ouvert constituerait une perte de validité impardonnable

Il est donc indispensable de sanctionner des niveaux de compétence en langue au travers d'épreuves qui reflètent une (ou plusieurs) authentique(s) compétence(s) en situation de la langue. Tant que la démarche qualité, ou tout au moins les techniques retenues par les auditeurs pour se mettre en place, imposera des analyses sur des dimensions uniques, il y a fort à parier que des examens qui recourent à la simulation globale ou à la mise en jeu de compétences multiples pour résoudre une tâche, seul ou en groupe, n'atteignent pas les seuils requis d'exigences. C'est une limite dommageable à la qualité même de ces certifications. Mais tant que les épreuves de certaines évaluations certificatives se satisferont de seuils de fiabilité faibles, voire non calculés, avec des écarts de notes très grands pour

une même performance, ou des procédures d'harmonisation relativement flexibles, cela constituera également une limite inacceptable de qualité. C'est toute la gageure qui attend par exemple le diplôme de compétences en langue, DCL, dont nous avons déjà souligné la richesse en termes de validité, (Riba, 2008) mais qui, par sa nature, reste une certification qui devra faire ses preuves dans le domaine de la fiabilité, alors même que les formations et les habilitations reçues par les correcteurs et examinateurs du Delf et du Dalf ne semblaient pas suffisantes, il y a encore peu de temps, pour garantir un seuil de qualité acceptable au regard des normes d'Alte.

Il est bien entendu réducteur de considérer qu'il y a une seule et unique culture anglo-saxonne, rebelle à l'évaluation de la production. Rappelons que l'Oral Proficiency Interview, OPI, d'American Council on the Teaching of Foreign Languages, ACFTL, a probablement été à l'initiative des grilles d'évaluation guidée utilisées dans de nombreux cas aujourd'hui pour les épreuves de production ; et notons aussi que de nombreuses équipes travaillent depuis longtemps en Europe sur la recherche de modèles de mesures multidimensionnelles (Pienemann, Johnston, Brindley, 1988). De son côté, le Ciep a fourni de notables efforts pour former ses évaluateurs et obtenir un ratio de fiabilité satisfaisant (sans toutefois publier ses résultats qui pourraient apporter du crédit aux typologies d'épreuves ouvertes de production). Il n'en demeure pas moins qu'un vrai clivage semble perdurer que le peu d'intérêt de notre université pour ces questions ne pourra que conforter.

Références bibliographiques

Chhane-Davin F., Groux D., (2009). « Méthodologie de la comparaison en éducation, introduction au dossier », in *Raisons, comparaisons, éducations, la Revue française d'éducation comparée*, septembre 2009, n° 5, pp. 11-14.

- Irtenemann M., Johnston M., Brindley G. (1988). « Constructing an Acquisition-Based Procedure for Second Language Assessment », in *Studies in Second Language Acquisition*, n° 10, pp. 217-243.
- Riba P. (2008). « Conception et validité de tâches évaluatives dans une perspective actionnelle », *Recherche et application*, n° 45, pp 124-133.
- Schubauer-Leoni M.-I. (2009). « Les Outils de la comparaison en éducation », in *Raisons, comparaisons, éducations, la revue française d'éducation comparée*, septembre, n° 5, pp. 15-30
- Suchaut B. (2008). « La Loterie des notes au bac, un réexamen de l'arbitraire de la notation des élèves », *Les Documents de travail de l'Iredu*, mars.
- Sitographie (dernière consultation, 25 novembre 2012)
ACFTL www.language-testing.com/lacad_opi.htm
Alle www.alle.org
Surveylang www.surveylang.org/fr

RAISONS
COMPARAISONS
ÉDUICATIONS

La Revue française d'éducation comparée

**L'enseignement des langues
étrangères face aux évolutions
des systèmes éducatifs**

Rédacteur en chef
Dominique Groux

Coordination
Fabrice Barthélémy

L'HARMATTAN

© L'HARMATTAN, 2013

5-7, rue de l'École-Polytechnique, 75005 Paris

<http://www.librarieharmattan.com>

diffusion.harmattan@wanadoo.fr

harmattan1@wanadoo.fr

ISBN : 978-2-343-00557-7

EAN : 9782343005577

ÉDUCATION COMPARÉE
Raisons - Comparaisons - Éducations

L'ENSEIGNEMENT DES LANGUES ÉTRANGÈRES
FACE AUX ÉVOLUTIONS DES SYSTÈMES ÉDUCATIFS

Rédacteur en chef : DOMINIQUE GROUX
Coordination : FABRICE BARTHÉLÉMY

L'Harmattan