
Le CECRL et l'évaluation : de la « bonne théorie » aux effets pervers

Assessment through the Common European Framework of Reference: from “good theory” to “adverse effects”

Leistungsbeurteilung mit dem Europäischen Referenzrahmen: Ideal und unerwünschte Nebenwirkungen

Odile Schneider-Mizony

 <https://www.ouvroir.fr/cpe/index.php?id=685>

DOI : 10.57086/cpe.685

@reference_electronique

Odile Schneider-Mizony, « Le CECRL et l'évaluation : de la « bonne théorie » aux effets pervers », *Cahiers du plurilinguisme européen* [en ligne], 6 | 2014,

@mis_en_ligne_le 01 janvier 2014, @consulte_le 25 juin 2024. URL : <https://www.ouvroir.fr/cpe/index.php?id=685>

@licence

Licence Creative Commons – Attribution – Partage dans les mêmes conditions 4.0 International (CC BY-SA 4.0)

Le CECRL et l'évaluation : de la « bonne théorie » aux effets pervers

Assessment through the Common European Framework of Reference: from “good theory” to “adverse effects”

Leistungsbeurteilung mit dem Europäischen Referenzrahmen: Ideal und unerwünschte Nebenwirkungen

Odile Schneider-Mizony

@TEXTE_INTEGRAL

- 1 Le titre « CECRL et évaluation », ainsi que les adjectifs axiologiques qui l'explicitent (« bonne », « pervers ») pourraient être lus comme une évaluation de l'évaluation du CECRL, alors que l'intention de cette contribution est d'en comprendre la logique évaluatrice (partie 1), de la prolonger sur certaines lignes (partie 2 sur la certification), et d'en exposer ouvertement certaines difficultés, dont certaines ne sont pas facilement solubles (partie 3). Cette réflexion d'ordre plus théorique que pratique¹ s'appuie sur les réflexions déjà amorcées par d'autres auteur/e/s, qui ont par exemple fait remarquer à quel point l'évaluation occupe une place centrale dans le CECRL, se retrouvant de façon transversale à de nombreux endroits autres que le chapitre 9 qui lui est consacré (Huver, 2009 : 72-73).
- 2 Pour ce faire, nous mettrons en regard ce qu'écrit le *Cadre* lui-même, avec les propos tenus par les acteurs institutionnels qui le mettent en œuvre en France ainsi qu'avec les rapports d'expérience qu'en font les enseignant/e/s.
- 3 Différents principes directeurs pour évaluer à l'aide du CECRL sont de bons principes pour les langues vivantes, comme le poids renforcé de l'évaluation de l'oral. Le *Rapport* de 2007 de l'Inspection sur l'évaluation en langues vivantes écrivait que la généralisation de l'évaluation orale obligerait les enseignants qui n'évaluaient qu'à l'écrit à changer, au moins partiellement, leurs pratiques. Ce souhait d'une évolution vers plus d'oral n'est pas sans objet : parents et élèves rapportent encore de nos jours tel propos d'un/e enseignant/e de langue vivante indiquant à l'élève déçu par sa moyenne ou par son appréciation sur le bulletin que les résultats écrits sont plus importants.

Le CECRL relaye en cela le BO du 8 juin 2006, qui indique que l'enseignement doit « privilégier l'apprentissage de l'oral » : les exemples d'évaluation proposés dans le chapitre 9 concernent ainsi quasi exclusivement des activités orales : p. 136, 137 (liste de contrôle décomposée en descripteurs d'activités orales), 141 (exemples oraux pour l'évaluation directe), 142 (exemples oraux pour l'évaluation de la performance), 142 (pour l'évaluation analytique), etc. L'objectif de changer les pratiques scolaires par le formatage de l'examen (principe du *back-wash*) se justifie par ses bons motifs.

- 4 Ce souhait de renforcer en premier lieu l'évaluation de l'interaction orale, activité contrastée avec l'oral du monologue, s'impose également comme une évidence à tout enseignant/e qui aura un jour constaté combien tel élève ou étudiant/e, participant activement en classe ou capable d'un exposé très correct en langue, donnait brusquement dans le bredouillement peu intelligible lorsqu'il était mis en face d'un vrai locuteur étranger qui lui posait une vraie question au cours d'un voyage dans le pays par exemple².
- 5 Évaluer positivement la capacité à faire quelque chose, même de petites choses, dans la langue étrangère, au lieu de corriger constamment les fautes, est d'un intérêt pédagogique si évident que la profession n'avait pas attendu le *Cadre* pour s'en pénétrer. Présentant à Grenoble les avantages de l'évaluation du CECRL, l'Inspecteur Général Francis Goullier formule qu'« on ne peut pas motiver pour quelque chose pour lequel on échoue » (2008), le premier « on » représentant les enseignant/e/s, le second les élèves. Cette évaluation positive et son orientation à la tâche est la forme d'évaluation qui intéresse les non-linguistes, les apprenants non-spécialistes et le grand public, comme le remarque Pierre Frath (2006 : 2), c'est-à-dire ceux que l'on appelle maintenant les clients de l'institution scolaire. Évaluer en partant de A1 en direction de compétences plus développées, et non en prenant la compétence du natif (C2+) comme étalon par rapport auquel est mesuré un (gros) écart, permet le relevé de quelque chose : toute obtention d'un niveau de certification, si bas soit-il, valorise l'apprenant, au lieu qu'une note basse dans des travaux complexes le démotive.
- 6 Le développement réfléchi de l'auto-évaluation est également un progrès, mais non pas en raison de l'argument développé par Goullier

« Comme on ne sait pas de quelle langue ils auront besoin plus tard, il est important qu'ils apprennent à s'évaluer aussi, afin de pouvoir juger de leurs compétences dans de nouvelles langues » (2008) ; car penser tous les apprenant/e/s comme de futurs polylingues *in progress* n'est pas vraisemblable. Mais l'individu peut ainsi dépasser sa dernière hétéro-évaluation par une institution s'il a fait de vrais progrès en raison de nouvelles circonstances de vie. La potentialité d'autonomie donnée par l'auto-évaluation n'est pas simplement une caresse psychologique, elle peut être un gain réel sur un CV ou dans un cursus professionnel.

- 7 Enfin, dernier argument mis en avant par l'Inspection (2007 : 11), le remplacement de la note moyenne de langue par ce qui est joliment appelé un « profil linguistique », distinguant la compréhension de l'expression, l'écrit de l'oral, et l'oral continu de l'oral interactif, donne une image plus véridique des différentes compétences de l'apprenant³. Dans l'infinie variété des individus, de leurs biographies, caractères, intérêts, on peut (en théorie) différencier la bonne compréhension de l'écrit de la mauvaise orthographe, la facilité à interagir de l'exposé discipliné, etc. Dans une conférence, ce profil différencié était vanté, à l'exemple des Pays Bas, comme permettant de distinguer entre les compétences de consommateur (voir des films en VO, lire des notices de produits dans d'autres langues) et les compétences productives, potentiellement utiles dans les métiers de services dont ce pays serait riche (Goullier, 2007).
- 8 On peut être plus réservé devant la louange des autres avantages de cette nouvelle façon d'évaluer, comme le *distinguo* entre les types d'évaluation, qui intéresse moins les élèves ou leurs parents que les institutions ou gouvernements. Les parents s'intéressent à l'évaluation diagnostique de type : « Jusqu'où pourra-t-il aller ? », qui contient le secret espoir d'études brillantes. Ce grand public, rassemblant parents, amis et observateurs non experts, peut aussi souhaiter une évaluation sommative : « Qu'a-t-il appris ? » qui lui permettra de conclure : « Il en sait moins ou plus que de mon temps ! » L'évaluation dite formative par le *Cadre*, qui demande la valeur ajoutée par la formation, intéresse relativement peu la société en dehors des experts budgétaires. La distinction vantée entre les différentes formes d'évaluation ne sortira donc guère des cercles d'initiés pour acquérir une utilité sociale.

- 9 On sera également dubitatif sur l'accroissement en lisibilité d'évaluations codées en B1, B2, voire en B1+, B2-, par rapport à un ancien 14/20, et pensera que toutes ces notions n'étaient pas aussi ignorées des enseignant/e/s que l'institution fait semblant de le croire maintenant : un professeur de langues qui mettait un 17/20 auparavant avait bien le sentiment de réaliser une évaluation positive et en connaissait la vertu, même si ce n'était pas habillé dans les mêmes formulations. De même, tout enseignant débutant expérimentait, puis perfectionnait le fait que l'enseignement et l'évaluation sont des choses différentes, distinction dont il est dit que le CECRL en rend conscient. Semblant dénigrer les pratiques antérieures, l'institution parle d'une faiblesse particulière de l'évaluation des langues vivantes en France, mais les arguments utilisés paraissent spécieux :
- l'évaluation dite «sauvage» des compétences langagières de l'élève par rapport au natif de la langue n'est pas une spécificité des parents et citoyens français ; l'institution nomme ainsi les situations où le parent demande à son enfant de lui traduire telle lettre arrivée en langue étrangère ou de lui expliquer tel propos tenu en langue étrangère en situation réelle, situation dont le jeune se tire souvent fort mal, en France comme dans d'autres pays européens...
 - les dispositifs à notes sont en usage dans de nombreux pays, qu'ils embrassent de 0 à 20 (orientation positive vers le chiffre élevé) comme en France, ou de 1 à 6 comme en Allemagne (orientation négative vers le chiffre élevé) ; et dans tous ces pays, on est conscient aussi bien des dégâts psychologiques causés par des notes abusives et mal vécues que des vertus d'orientation et de socialisation scolaire qu'elles représentent également (Winter, 2008 : 47).
 - la remise en question du «sens» d'un 8/ 20 ou d'un 12/20 globalisant diverses compétences et connaissances se poserait tout aussi bien dans d'autres disciplines scolaires : mathématique (algèbre, géométrie, statistiques), musique (solfège, dictée musicale, histoire de la musique ou pratique d'un instrument), enseignement de langue maternelle, d'activités sportives, etc.
- 10 Ces reproches faits par le *Rapport* 2007 de l'Inspection à l'enseignement français des langues vivantes (évaluation sauvage manifestant un échec, notes démotivantes et d'une globalité inadéquate) ne constituent pas une tare méthodologique qui serait spécifique à l'enseignement des langues en France. Mais constituer cette représentation rend plus attractive pour la communauté nationale l'harmonisa-

tion par le CECRL : un changement s'avérera d'autant plus nécessaire que le stade actuel aura été décrit comme imparfait.

- 11 Le CECRL n'est pas né de réflexions françaises sur l'enseignement des LV⁴, il est une déclinaison nationale de mesures éducatives prises par les institutions communautaires et intergouvernementales où s'élaborent les principes et outils façonnant les politiques éducatives : sur le manuel, le logo du Conseil de l'Europe et la signature de la Division des politiques linguistiques à Strasbourg rappellent cette paternité au lecteur. Le souci d'aligner les pratiques éducatives européennes est très marqué dans le domaine de l'évaluation : l'application des mêmes principes d'évaluation dans différents pays permettrait idéalement de comparer les divers diplômes nationaux. Un exemple classique des difficultés antérieures à cet exercice est celui de la comparaison des notes entre la France et l'Allemagne, riche de multiples échelles de correspondance entre les notes allemandes (de 1 à 6) et les notes françaises (de 20 à 0), avec une moyenne française posée tantôt à 3 (6 divisé par 2), 3,5 (milieu de l'intervalle) ou à 4 (qualifié en Allemagne de niveau suffisant « ausreichend »). Le zéro français y constitue un problème insoluble, parce qu'il ne correspond pas à une pratique allemande, qui ne donne jamais la note la plus basse de 6, et l'équivalence du 1 allemand n'est guère plus convaincante, parce que la note maximale de 20/20 est rare en France. L'évaluation par l'intermédiaire de cotes en A1 ou B2 devient par comparaison un modèle de simplicité, reproductible avec d'autant plus d'efficacité que les pays européens adhéreront en plus grand nombre. Cette comparabilité est vue par les collègues français comme un bienfait. Citons le jugement à ce titre dans le dossier sur le CECRL publié par les *Langues Modernes* (2008 : 20) :

Leur avantage majeur⁵ est de rendre possible la comparaison entre les différents diplômes nationaux, les notes traditionnelles ne suffisant pas à remplir cette tâche. On gagne ainsi en cohérence et en transparence à l'échelle internationale.

- 12 De la même façon, l'étiquette « européenne » assure une légitimité supplémentaire à l'évaluation : ce n'est pas moi qui vous note de cette façon, c'est l'Europe, comme le fait remarquer Pierre Frath (2006 : 12).

- 13 Certains collègues allemands, de culture nationale plus réticente à ce qu'ils appellent « der gläserne Mensch », l'homme transparent, y voient au contraire un danger politique, celui d'une sorte de traçabilité mondialisée ne permettant plus aux individus de nouveaux départs dans d'autres cultures organisées différemment (Tönshoff, 2003 : 187-188). Cependant cette harmonisation européenne ne résulte pas d'une contrainte explicite, ce qui amène l'institution à parler de « liberté » française à appliquer telle ou telle option du CECRL. Même si le B.O. a déjà publié les fiches d'évaluation correspondant à ces nouvelles cotes, ce qui éminemment directif, le discours institutionnel fait comme si les acteurs étaient volontaires, et prône l'émulation entre langues comme dans une compétition ludique. La souplesse vantée du CECRL, avec la métaphore de la caisse à outils de laquelle on sort ce dont on a besoin pour sa situation professionnelle précise, peut donner un sentiment de liberté. Cette liberté à élaborer ses outils, à traduire l'esprit du *Cadre* dans ses classes et ses activités et donc à fabriquer ses propres évaluations, représente une tâche d'une telle ampleur que de nombreux collègues en langues ont entrepris de collaborer et mutualiser leurs travaux dans les établissements : le travail en équipe résultant est alors aussi bien un espace d'harmonisation de l'établissement qu'un lieu de « bon esprit ».
- 14 Enfin, un espace de liberté est dessiné par le Portfolio et son auto-évaluation, sous-type d'évaluation présenté par le *Cadre* (p. 144-145). Bien que reconnue comme un important facteur de motivation, l'auto-évaluation est cependant « recadrée » par le manuel au sens de ramenée à la seule interprétation acceptée par l'institution : elle se voit acceptée si l'enjeu n'est pas important, dans un contexte où les apprenants ont été formés à l'évaluation et sont surveillés par l'institution – toutes choses qui sont aux antipodes de l'auto-évaluation. Cette forme inhabituelle à l'institution est manifestement accueillie de façon réservée : sans doute l'auto-évaluation par l'élève remplit-elle difficilement la fonction de certification de compétences linguistiques dévolue à présent à l'évaluation en langues.
- 15 L'évaluation du CECRL est déclarée bonne en fonction de caractéristiques (évolutivité, professionnalisation) qui la rendraient à présent incontournable. Mais le fait que le CECRL soit omniprésent dans la politique éducative française et qu'il devienne même une « référence mondiale » pour le site www.Franc-parler.org⁶ n'en fait pas de facto

un progrès sauf à considérer que la fréquence génère obligatoirement la qualité.

- 16 Sans doute l'évaluation par des descripteurs formulés en capacités pratiques représente-t-elle une meilleure adéquation entre les connaissances des formés et les compétences recherchées par les employeurs : le monde de l'éducation et celui du travail sont alors rapprochés par le *Cadre*, et c'est en ce sens que l'on peut comprendre le « pas de géant vers la professionnalisation de l'évaluation en langues » (Goullier, 2007). Évaluer par compétences revient à évaluer un savoir par l'intermédiaire d'une tâche : c'est un progrès pratique, mais ce n'est pas un progrès docimologique. De Ketele & Gerard, qui constatent que l'évaluation des compétences n'est pas à la hauteur des exigences édumétriques⁷ en la matière, font penser que l'évaluation prônée par le CECRL pourrait être fiable et valide, mais qu'elle ne remplit pas le critère de pertinence (2005 : 4). Par ailleurs, les exigences méthodologiques à respecter pour qu'une évaluation soit fiable et valide sont si pointues qu'on peut douter que des enseignant/e/s non formé/e/s à cette matière complexe fabriquent fortuitement, en recherchant « librement » un type d'évaluation qui conviendrait bien à la séquence ou tâche exercée avec leurs élèves, un test qui n'aurait aucun biais méthodologique.
- 17 L'évaluation favorisée par le CECRL est cependant infalsifiable en ceci qu'elle n'est guère soumise à l'épreuve de la réalité : les tâches pouvant se décliner en évaluation sont rarement précisées dans le *Cadre* lui-même, le seul et même exemple servant dans les deux conférences de l'inspecteur recueillies sur la Toile ; la présentation badine sur le mode du « Si vous voulez rester jusqu'à demain matin, je pourrai(s) vous donner d'autres exemples » décourage d'ailleurs, dans la conférence visionnée, le public d'en demander plus. Que le CECRL s'engage peu dans la déclinaison concrète est compréhensible face à la diversité des situations pédagogiques ; mais cette réserve a l'effet avantageux de rejeter les possibles écueils de la pratique hors de son domaine de responsabilité. L'ouvrage lui-même se pose comme un discours scientifique étant allé jusqu'au bout de la question, doté d'une bibliographie d'apparence scientifique et utilisant des termes à forte aura comme les « items d'ancrage du modèle de Rasch » (CECRL, 175). Lorsqu'on prend cette scientificité au mot, le chapitre 9 (*Evaluation*) révèle une bibliographie dont le cœur temporel date des

années 1990, dans laquelle on cherchera vainement aussi bien des titres de George Rasch que des théoriciens récents, comme les auteurs de la revue *Mesure et évaluation en éducation*. Une recherche extérieure au *Cadre* nous apprend que Rasch, statisticien pionnier dans la psychométrie des tests fondant par exemple les enquêtes Pisa, est si mal utilisé en statistique éducative et que ses critères sont si fortement critiqués⁸ qu'il semble audacieux de fonder un ensemble de politiques scolaires européennes sur un modèle docimologique si peu consensuel.

- 18 Mais la pensée circulaire empêche le doute de s'insinuer : l'avertissement du CECRL précise page 5 que les six niveaux d'évaluation se fondent sur la pratique d'organismes publics de certification, qui réutilisent ensuite les critères du *Cadre* pour leur certification, où ils se révèlent, sans surprise, opératoires. Les défauts de cohérence entre le CECRL et l'évaluation sont déclarés malentendus ou effets d'optique (Goullier, 2008). Par ailleurs, l'habitus courtois (on n'impose rien à personne) et évolutif, en chronologie (on ne décline pas le cadre exactement de la même façon en 2010 qu'en 2001) comme en lieux (adaptez ça à votre propre situation) ne permet pas le contre-exemple. Les discours autour de l'évaluation du CECRL prennent rapidement un air de publi-information, comme dans le dossier « Évaluer avec le CECRL » où l'on cherche vainement, à une rubrique intitulée « oui, mais... », des considérations qui relèveraient du « mais ». L'examen critique s'épuise à courir derrière une matière aussi protéiforme.
- 19 La souplesse empirique du CECRL est iconique de la souplesse éducative jugée idéale à notre société : elle permettra une mise à jour plus régulière des compétences souhaitées pour les individus formés par le système éducatif. Car l'évaluation perd peu à peu de son ancrage scolaire pour devenir une accréditation à valeur sociale.
- 20 La langue française distingue l'évaluation « action d'évaluer, mesure à l'aide de critères déterminés, des acquis d'un élève ou de la valeur d'un enseignement » de la certification « attestation de conformité d'une denrée ou d'un produit à des caractéristiques ou à des normes pré-établies »⁹. L'évaluation met l'accent sur le rapport élève / mesure de valeur (les critères sont l'outil), alors que la certification insiste sur le rapport entre les critères mesurés et les critères visés

(l'élève est l'objet). À la lumière de cette définition, l'évaluation est individuelle alors que la certification est sociale. L'usage moderne des deux termes en discours tend à les rapprocher, voire à les confondre. Une petite recherche sur Internet des couples « évaluer et certifier » et « évaluer ou certifier » fait remonter un usage quasi-synonymique, que ce soit l'usage de l'ambassade de France au Nigéria sous le titre « Pourquoi évaluer et certifier son niveau en langue française ? », ou celui de l'AFNOR, qui indique que sa certification propose des prestations d'évaluation. Ce brouillage sémantique fait comprendre l'hésitation du *Cadre* entre l'évaluation d'un savoir individuel, de culture et de formation cognitive personnelle, et l'évaluation d'actions en langues dans la société et la profession. Pierre Frath fait remarquer (p. 8) que proposer la certification par les niveaux du CECRL comme évaluation scolaire le retransforme de facto en examen individuel. Ce mélange de sens se retrouve dans les informations à destination du grand public, comme dans l'article de l'encyclopédie grand public wikipedia « Le CECRL et les **examens** de langues (sic)¹⁰ », où un tableau propose une lecture en A1-C2 des **certifications** en langues les plus utilisées en Europe. L'utilisation réelle de l'évaluation façon CECRL en fait plus un test de niveau que l'évaluation formative (progression) ou sommative (somme des acquisitions) qu'elle prétend être. De là provient la difficulté ultérieure à faire correspondre les critères nécessaires aux tests (pertinence¹¹, fiabilité¹², validité¹³) avec ce qui se passe en classe, qui correspond à d'autres logiques : centration sur l'apprenant individuel, évolution en fonction des lieux et circonstances, difficulté à cibler une compétence isolée (Walter, 1999).

- 21 L'IGN présente la certification par des partenaires extérieurs comme « levier pour ôter la suspicion » (Goullier, 2007). Le BO du 8 juin 2006 la prévoit « en liaison étroite avec au moins l'un des pays européens dont la langue concernée est la langue officielle ». Ces formules recouvrent un changement de paradigme qui est la remise en question du principe de l'évaluation interne. Alors que depuis des siècles, l'école connaissait un mode de régulation interne par lequel les enseignants jugent de la valeur des travaux des élèves en fonction de critères assis sur une connaissance disciplinaire approfondie, seule vaudrait à présent l'évaluation externe, l'audit (on trouve le terme dans le CECRL p. 174 par ex.), le cabinet de consultant extérieur prétendument impartial. En déplaçant l'évaluation hors du système, on la

déclare indépendante, ce qui signifie indépendante des enseignant/e/s, mais non des autres acteurs. Si l'évaluation extérieure à l'école devient la seule légitime, elle se basera sur les représentations sociales répandues dans le grand public. Même si une évaluation interne à l'école n'est pas obligatoirement scientifique, on ne voit pas pourquoi des extérieurs sans connaissance des conditions de l'apprentissage (savoir antérieur, contexte social et pédagogique, progression didactique visée) ni de l'épistémologie de la discipline auraient des attitudes par principe plus scientifiques. L'évaluation extérieure peut certes être vue comme un contrôle d'efficacité, mais selon les critères émis par l'organisme qui évalue. Si une certification semble plus sérieuse en dehors de l'école qu'en son sein, c'est en raison de son effet de réalisme (ça vient de la vraie vie...), qui efface malheureusement les compétences séculaires accumulées dans la culture scolaire. Et reléguer le rôle d'évaluateur, tout incommode qu'il soit parfois, hors de l'école, fera perdre à cette dernière la légitimité pour mener avec l'élève le dialogue d'évaluation susceptible de le faire progresser

- 22 L'évaluation externe par certification a une autre conséquence, sa déconnection de l'école confirme et soutient le développement grandissant de l'éducation hors système scolaire : facteur de progrès s'il s'agit de donner de nouvelles chances de formation aux individus et de les leur attester, façon « *zweiter Bildungsweg* » allemand (= deuxième chemin de formation), il s'agit d'un outil à double-tranchant si l'éducation est remise entre les mains de chaque individu, charge à lui de saisir cette chance, de se former comme et au prix où il l'entend, quitte à faire certifier la nouvelle compétence acquise par un des multiples organismes qui s'en occupent. La situation actuelle est une situation de l'entre-deux : on ne demande pas encore aux élèves de payer les certifications en langues réalisées par des organismes extérieurs, et c'est alors le contribuable qui enrichit ces instituts privés pour une tâche que l'institution aurait peut-être pu monter elle-même, mais pour laquelle elle passe des marchés d'un gros montant financier. Le jour où l'on demandera aux candidats aux certifications de les payer eux-mêmes, comme cela arrive déjà dans certaines universités, l'externalisation de l'évaluation aura un surcoût financier pour les individus comparativement à une situation antérieure de gratuité.

- 23 Les langues étant probablement la matière scolaire apprise la plus spontanément en autodidacte en dehors de l'école (plus que les maths ou la physique par exemple), la certification des compétences en langues recèle un potentiel économique d'autant plus intéressant que les heures consacrées aux Langues Vivantes Étrangères diminuent dans l'institution scolaire. L'alignement sur les mêmes outils d'évaluation et les mêmes niveaux met en concurrence de légitimité l'école et les certificateurs privés, de Cambridge au Portfolio en passant par le *Zertifikat Deutsch*. Le dernier texte publié sur le site ministériel de *Eduscol* en mai 2010 reprend d'ailleurs le passage du BO 2006 évoqué ci-dessus en une formulation nettement plus ferme : « la certification sera conçue en étroite relation avec un organisme international reconnu ». Le BO est passé de la vague référence au pays aux organismes concrets, et leurs sigles, DELF & DALF, DIALANG et CLES, Test-DAF et ESOL, TOEFEL et TOEIC, DELE (espagnol) et PLIDA (italien), prolifèrent à présent dans l'enseignement public.
- 24 L'innovation du CECRL accompagne efficacement l'évolution du monde éducatif vers des professeurs moins nombreux, veillant à ce que les élèves ou étudiant/e/s apprennent en relative autodidaxie, et un marché économique de la certification externe, chargé d'accréditer le résultat du travail des deux premières catégories de personnes. Mais l'axiome du plus grand sérieux de l'évaluation externe n'est remis en question que par un nombre très restreint de collègues, qui s'effraie des montants en jeu (Comerford, 2009). Pour citer un exemple, l'avis d'attribution du marché des certifications des seuls élèves des classes de Seconde de lycées français était de 2 millions 600 000 euros pour l'ESOL de Cambridge et de 3 millions 500 000 euros pour l'Institut Cervantes pour l'année 2010-2011, soit un seul niveau de classe et seulement deux des diverses langues concernées¹⁴.
- 25 L'application du CECRL bute sur l'écart inévitable entre le bel idéal et les aspérités de la réalité : certaines distorsions entre les deux ont déjà été énoncées, comme l'absence de valorisation des compétences plurilingues, actionnelles (collaboratives) et contextuelles, le CECRL ne proposant pas de descripteurs à leur sujet (Huver, 2009 : 80). Ou encore, tout positivant qu'il se conçoive, le *Cadre* formule cependant des évaluations négatives pour certains seuils, par exemple des restrictions de type « peut communiquer lors de tâches simples et habi-

tuelles **ne demandant qu'un** échange d'informations simple et direct sur des sujets familiers et habituels » pour A2 (*Cadre*, p. 25) ou « a un contrôle **limité** de quelques structures syntaxiques et formes grammaticales simples appartenant à un répertoire mémorisé » pour A1 (*Cadre*, p. 28). Les enseignants, surtout, soulignent le grand écart entre l'exigence d'évaluation par rapport aux critères « objectifs » proposés par les descripteurs et ce qui se fait en classe (*Langues Modernes*, 2008 : 28). L'évaluation de compétences comme la médiation, qui paraît plus comportementale que langagière (*Langues Modernes*, 2008 : 60), ou qui reflètent l'habitus et la culture familiale (aisance sociolinguistique) pose des problèmes éthiques aux collègues qui ont l'impression d'évaluer le caractère ou le milieu social de leurs élèves (*Langues Modernes*, 2008 : 60). Même si les consignes manquent encore pour ces évaluations délicates du point de vue éthique, la récente évolution des concours de recrutements des enseignants montre, avec l'introduction de l'épreuve de « Agir en fonctionnaire de l'État éthique et responsable », une appétence grandissante de l'institution à évaluer la qualité morale des actants, qui risque de garder une actualité à ce problème.

- 26 Pour les pédagogues ayant mis leurs états d'âme de côté, le plus gros souci est la traductibilité des niveaux dans la réalité scolaire : ils indiquent par exemple avoir une vision claire des niveaux A1 et C, mais en être réduits à des approximations pour les intermédiaires. La difficulté de distinguer entre C1 et C2 est relevée avec constance, qu'il s'agisse de remarquer que nombre de locuteurs de langue maternelle n'atteignent pas les descripteurs C dans diverses tâches, ou de supposer que la cote C2 n'a aucun intérêt éducatif et a été rajoutée pour compléter le paradigme en une déclinaison binaire des 3 niveaux-lettres. Pour différencier les compétences moyennes après quelques années de scolarité, on voit apparaître un B1+ s'intercalant entre B1 et B2. Enfin, se pose, suivant l'endroit du CECRL où l'on puise son inspiration évaluatrice, la question de la nécessité de réussir la tâche à 100%, si l'on prend pour modèle la tâche de discrimination auditive du pilote d'avion évoquée page 5 de l'introduction, qui paraîtra seule légitime à ceux qui prennent l'avion. Ou bien l'on privilégiera la réponse plus modérée et partielle, si l'on s'inspire des spécifications de ALTE, qui considère que le niveau est atteint si 80% des items sont atteints (*Cadre*, p. 175).

- 27 Toutes ces difficultés sont les difficultés induites par une échelle dont les différents barreaux ne sont pas équidistants : la distance entre A1 et A2 est plus réduite qu'entre B1 et B2, elle-même plus réduite qu'entre B2 et C1. Il ne s'agit pas non plus d'une progression géométrique, rien de tel n'ayant été formalisé par des recherches linguistico-métriques. L'application des niveaux chiffres-lettres à un monde scolaire qui n'a pas été pensé pour cela mène les collègues à bricoler des échelles de correspondances suivant les classes et les niveaux, comme le rapporte la revue des *Langues Modernes* (2008 : 48 & 49), afin d'éviter que les élèves ne ratent exprès leur évaluation diagnostique de début d'année, afin d'être mis dans un groupe plus faible qui leur assurera de meilleures « notes » :

	A1	A2	B1	B2
Classe de Seconde	5/20	10/20	20/20	? (peu vraisemblable)
Classe de Première	3/20	7/20	18/20	20/20

- 28 Ces échelles sont alors concertées dans l'établissement, et représentent l'avatar fortuit d'un modèle d'évaluation en contradiction avec la pratique évaluatrice de tout un système scolaire. La question posée par un élève lors de la répartition en groupes de compétences en début d'année et rapportée en titre d'un article des *Langues Modernes* « Alors, m'sieur, on est dans le groupe des nuls ? » (2008 : 37) montre que l'évaluation par les compétences et le profil langagier est interprétée à la lumière des pratiques évaluatrices habituelles de la société en mathématiques, à la Bourse ou aux Jeux Olympiques, séparant les meilleurs des moins bons. L'évaluation positive du CECRL peine alors à s'implanter dans les esprits et l'effet de valorisation-élève n'est pas au rendez-vous que lui donnaient ses promoteurs.
- 29 Le changement de type d'évaluation contribue également à l'effacement de la formation intellectuelle au profit de la formation pratique. Dans l'enthousiasme généré par une évaluation devenue concrète, comme la fameuse carte postale que l'on est capable d'écrire du pays étranger, on perd la conscience que toutes ces tâches, même multipliées par l'envie d'expérimenter d'un/e enseignant/e imaginaire/ve, ne sont jamais que la juxtaposition de petites compétences opérationnelles. La mise en œuvre du CECRL s'inscrit fort bien dans une évolution qui va du modèle éducatif humaniste vers une formation

pratique, évolution largement entamée par certaines méthodes de langues qui l'ont précédé. Ce type d'évaluation borne le champ des possibles en délimitant le faisable par la mesure de ce qui a été fait : c'est l'inconvénient du parcours d'accumulation de compétences qui remplace ce qui était auparavant accès au savoir, contenant seul tous les possibles. La production du savoir demande du temps, des formes de cohérence complexes, pour que l'individu se l'intériorise et qu'il devienne en lui une seconde nature. Un savoir intellectuel sur la langue, le pays et sa culture peut être jugé inatteignable pour l'enseignement massifié de nos sociétés modernes, et son prix devient apparemment exorbitant pour les budgets publics du troisième millénaire. Mais seul le savoir libère vraiment sans se démoder, et dépasse l'inventaire à la Prévert des compétences de ceci ou de cela, liste agrégée de mini-qualifications qui ne font qu'adapter l'apprenant aux voyages qu'il réalisera, aux objets qu'il consommera et à la profession dans laquelle il travaillera. L'observation technocratique oublie les tâches complexes qui caractérisent l'accès mature de la pensée au monde. Dans la dispersion inutile, mais réelle des descripteurs, des échelles et des spécifications, c'est le chronomètre qui est devenu la raison d'être de la course, et le coureur est bien loin des préoccupations.

- 30 L'évaluation est tenue à tort pour une technique politiquement neutre : la prolifération des usages évaluatifs depuis une vingtaine d'années tend à les inscrire dans le sens commun comme une nécessité qui n'est plus discutée. Quand s'y ajoute l'effet de sérieux donné par les certifications par organismes extérieurs éblouissant de leur prestige financier une école sous-financée et méprisée socialement, le souhaitable devient le meilleur score obtenu à la certification, y compris a contrario avec l'épisode de l'évaluation en langues au niveau A2 du Diplôme National du Brevet (DNB) de 2009, validé pour tous les élèves a posteriori en raison des difficultés de mise en pratique. L'inefficacité de cette première application peut être tenue pour une difficulté initiale, on peut y voir à l'inverse la révélation d'un effet pervers de l'évaluation façon CECRL qui distribue des « assignats scolaires » ne correspondant plus à une vraie qualification. La déclinaison dirigiste, sous des dehors patelins, du CECRL en France, sa préférence pour l'évaluation de compétences « court-termistes », et son évolution vers une certification externe à l'école rendent plus

nette la distorsion entre un certain idéal pédagogique et l'expérience scolaire.

@BIBLIOGRAPHIE

Corpus

AMBASSADE DE FRANCE à Abuja
« Pourquoi évaluer et certifier son niveau en langue française », consulté le 13/05/10 : http://www.ambafrance-ng.org/france_nigeria/spip.php?article739

Bulletin Officiel n° 23 du 8 juin 2006
« Enseignement des langues vivantes », téléchargeable à : <http://www.education.gouv.fr/bo/2006/23/MENE0601048C.htm>

Cadre européen commun de référence pour les langues. Apprendre, enseigner, évaluer, Division des politiques linguistiques Strasbourg, Conseil de l'Europe & éditions Didier, Paris, 2005.

« Le Cadre européen : où en sommes-nous ? » dans *Les langues Modernes*, n° 2/2008.

FRANC-PARLER, « Évaluer avec le CECR » sur le site Franc-parler, consulté le 20/05/09 : http://www.francparler.org/dossiers/cecer_evaluer.htm

GOULLIER, Francis, 2007, « Faut-il “appliquer” le Cadre Européen Commun de Référence pour les Langues (CECRL) ? », Conférence du 26 septembre 2007 au CRDP de Bourgogne, téléchargée le 21/04/09 : <http://crdp.ac-dijon.fr/Faut-il-appliquer-le-Cadre.html>

GOULLIER, Francis, 2008, « C.E.C.R., évaluation, Portfolio : quelle cohérence ? », Conférence du 14 mai 2008 à l'Université Stendhal de Grenoble, vidéo téléchargée sur le site du CUEF pour usage personnel à : <http://les-conferences-du-cuef.neopodcasts.com/content/1971.m4v>

INSPECTION GENERALE DE L'EDUCATION NATIONALE, 2007, « L'évaluation en langue vivante : état des lieux et perspectives d'évolution », *Rapport à monsieur le ministre de l'Education nationale, de l'Enseignement supérieur et de la Recherche*, n° 2007/009/janvier 2007 : www.education.gouv.fr/l-evaluation-en-langue-vivante-etat-des-lieux-et-perspectives-d-evolution.html

Littérature secondaire

COMERFORD, Richard, 2009, « Alerte ! L'Éducation Nationale est tombée dans l'escarcelle des marchands de certifications ! » dans ABDELGABER, Sylvie & MEDIONI, Maria-Alice, (dir.), *Enseigner les langues vivantes avec le Cadre Européen*. Hors-Série numérique n° 18 des *Cahiers Pédagogiques*, téléchargeable sur le site du CRAP-Cahiers pédagogiques, p. 162-164.

DE KETELE, Jean-Marie / GERARD, François-Marie, 2005, « La validation des épreuves d'évaluation selon l'ap-

proche par les compétences » dans *Mesure et évaluation en éducation*, vol. 28, n° 3, p. 1-26, consulté en août 2010 : http://www.bief.be/docs/publications/validcomp_070221.pdf

FRATH, Pierre, 2006, « Le portfolio européen des langues et le cadre européen commun de référence : entre normalisation institutionnelle et responsabilité individuelle », 12 pages téléchargeables à <http://www.aplv-languesmodernes.org>

HUVER, Emmanuelle, 2009, « Le CECRL : Au service de l'évaluation ? », dans *Education et Sociétés plurilingues* n° 27/ décembre 2009, p. 71-82.

JÜTTEMANN, Gerd, 2004, *Psychologie als Humanwissenschaft. Ein Handbuch*, Göttingen, Vandenhoeck & Ruprecht.

TÖNSHOFF, Wolfgang, 2003, « Referenzrahmen: Zwischen Ansprüchen und

Erwartungen » in BAUSCH, Karl-Richard et al., (Hg.), *Der Gemeinsame Referenzrahmen für Sprachen in der Diskussion*, Arbeitspapiere der 22. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts, Tübingen, Narr, p. 180-191.

WALTER, Paul, 1999, « Die Vermessung des Menschen: Messtheoretische und methodologische Grundlagen psychologischen Testens », in GRUBITZSCH, Siegfried (Hg.) *Testtheorie – Testpraxis: psychologische Tests und Prüfverfahren im kritischen Überblick*, Eschborn, Klotz, p. 98-127.

WINTER, Felix, 2008, *Leistungsbewertung. Eine neue Lernkultur braucht einen anderen Umgang mit Schülerleistungen*, Hohengehren, Schneider Verlag.

@NOTES

1 L'auteure de ces lignes n'a pas l'occasion d'utiliser le CECRL dans son quotidien de professeure de linguistique à l'Université de Strasbourg.

2 C'est ainsi que Yannick Lefranc parle de « droit des parleurs », pas du « droit de la langue » dans « Le "Français Langue d'Intégration" et le "droit à la langue du pays d'accueil". Remarques sur un révélateur et un "politiseur" de débat » in Huck Dominique et Potriquet Ghislain (dir.), *Émergence des notions de droits linguistiques et droits à la langue. Les apports d'une approche historique*, Actes de la journée d'études du 23 mai 2012 du GEPE / EA 1339 Université de Strasbourg, p. 77-118.

3 Le BO n° 2 du 14 janvier 2010 donne en annexe une telle fiche de « Profil linguistique » pour un candidat au BEP.

4 Voir la contribution de Dominique Huck à cette même journée sur le CECRL et l'évaluation.

- 5 « leur » s'applique aux niveaux et descripteurs, outils-clés de l'évaluation par le CECRL.
- 6 Site de l'organisation internationale de la Francophonie, donc (para)-institutionnel.
- 7 L'édumétrie est la forme spécifique d'évaluation pour les situations scolaires, que les politiques de testing de l'OCDE (PISA) confondent par exemple avec la psychométrie, qui est fort différente.
- 8 Le problème de l'échelle d'intervalle et la dépendance de la qualité d'échelle des choix théoriques sont soulevés notamment par Gerd Jüttemann (2004) et Paul Walter (1999).
- 9 Définitions du Petit Larousse.
- 10 C'est moi qui souligne.
- 11 La pertinence du test est son caractère approprié à l'objectif visé.
- 12 La fiabilité implique que le test soit reproductible sans différence de résultats par d'autres personnes en d'autres lieux. Reproductibilité qui semble au moins aléatoire vu les consignes d'adaptation et de flexibilité données par le *Cadre*.
- 13 La validité implique que le test mesure bien ce qu'il doit / dit mesurer.
- 14 Annonces du BO du 30 juillet 2010, BO n° 146C.

@RESUMES

Français

L'institution scolaire française présente le CECRL comme un changement de paradigme pour les dimensions « apprendre, enseigner, évaluer » de l'apprentissage des langues dans le contexte scolaire. La première partie de cette contribution expose quelques arguments des textes officiels (BO, rapports d'inspection, le CECRL lui-même) selon lesquels la nouvelle forme d'évaluation par l'intermédiaire du CECRL serait une « bonne » évaluation, en tout cas meilleure que les précédentes formules ; ce panorama s'accompagne d'un examen de la validité de ces arguments. La seconde partie contraste les termes d'« évaluation » et de « certification », qui s'entrecroisent constamment dans le *Cadre*. Elle émet l'hypothèse que la grandissante synonymie des deux reflète une externalisation croissante de l'apprentissage des langues. La troisième partie se penche sur la mise en pratique de l'évaluation à la façon du CECRL en France, distinguant entre des difficultés ponctuelles, solubles *a priori*, et des conséquences inhérentes à la logique de la nouvelle évaluation, dénommés ici « effets pervers ».

English

The French education system considers the *Common European Framework of Reference for Languages* as a change of paradigm in the school context as far as “teaching, learning and assessment” are concerned. The first part of this article will discuss some arguments presented in official documents (Official Bulletin, reports by inspectors of education, the CEFRL itself) according to which assessment through the use of the *Framework* would be a “good” assessment, or at least better than the previous ones. The validity of the arguments is examined. The second part contrasts the words ‘assessment’ and ‘certification’, the growing synonymy of which reveals that assessment is no longer reserved for use in classrooms and is more and more often externalized. The last part examines the practical implementation of assessment through the CEFRL in French schools and points out certain difficulties. Some of these are one-off and can be solved, but others are inherent to the *Framework* itself and can be considered as ‘adverse effects’.

Deutsch

Der Gemeinsame europäische Referenzrahmen für Sprachen und die Leistungsbeurteilung: zwischen Qualitätsanspruch und ärgerlicher Realität. Die neueren bildungspolitischen Richtlinien für Fremdsprachunterricht in Frankreich stellen den *Gemeinsamen europäischen Referenzrahmen* als einen Paradigmenwechsel vor, der das Lernen, Lehren und Beurteilen im Fremdsprachunterricht vollkommen erneuern würde. In einem ersten Teil befasst sich der Beitrag mit den Argumenten, mit denen der Fortschritt in der Leistungsbeurteilung durch Richtlinien und Berichte aus dem Erziehungsministerium gepriesen wird, und unterzieht diese Argumente einer kritischen Betrachtung. Im zweiten Teil wird die semantische Bandbreite dieser Leistungsbeurteilung untersucht, die sehr stark zur Zertifizierung neigt: die immer größere Deckungsgleichheit zwischen beiden könnte eine schulische Entwicklung abbilden, die einen Teil des fremdsprachlichen Erwerbs außerhalb der Institution vergibt. Schließlich fokussiert der Beitrag die Anwendung der Leistungsbeurteilung nach den Prinzipien des *Referenzrahmens* im französischen Fremdsprachunterricht: hierbei wird unterschieden zwischen praktischen Anfangsproblemen, die lösbar wären, und der Kohärenzschwierigkeit zwischen der neuen Evaluierungsart und der schulischen Praxis, die die Gefahr der Ineffektivität als Nebenwirkung birgt.

@INDEX

Mots-clés

évaluation en langues vivantes étrangères, externalisation de l'évaluation, harmonisation européenne des notes, marché de la certification, pointillisme de l'évaluation, praticabilité scolaire du CECRL

Keywords

assessment in foreign language teaching, certification as market value, classroom practicability of CEFR, outsourcing of school assessment, pernicky assessment procedure, standardization of testing at European level

Schlagwortindex

Europäische Notenharmenisierung, Fremdvergabe der Leistungsbeurteilung, Kleinkrämerische Evaluierungspraxis, Leistungsbeurteilung im Fremdsprachenunterricht, Schultauglichkeit des Europäischen Referenzrahmens, Zertifizierungsmarkt

@AUTEUR

Odile Schneider-Mizony

Professeure de linguistique allemande à l'université de Strasbourg. Recherches portant principalement sur la grammaticographie de l'allemand (histoire et période moderne) et l'enseignement de l'allemand langue étrangère (DaF) dans l'histoire et la période contemporaine.

IDREF : <https://www.idref.fr/158160428>

ORCID : <http://orcid.org/0000-0002-5194-0866>

ISNI : <http://www.isni.org/00000000367778768>

BNF : <https://data.bnf.fr/fr/16579721>