

QUE SAIT-ON DE L'ÉVALUATION EN LANGUE VIVANTE ÉTRANGÈRE ?

Emmanuelle HUVER

Professeure, Université de Tours

A. Éléments de préambule

Cette note propose un état des lieux de la recherche sur l'évaluation en didactique des langues (désormais DDL) en général, et non pas (ou en tout cas pas prioritairement) sur l'évaluation des langues dans l'Éducation nationale française (désormais EN) en particulier. Des focalisations sur les recherches menées sur le terrain éducatif français et/ou sur les implications des recherches pour le système éducatif français seront néanmoins opérées, lorsque cela sera possible et pertinent.

À des fins de concision, je ne présenterai pas les jalons historiques des recherches sur l'évaluation en langues (*cf.* pour cela Huver et Springer, 2011 ; IGEN 2007 ; Tardieu, 2013). Je proposerai donc plutôt un panorama des recherches contemporaines afin de dégager quelques grandes tendances dominantes (plutôt que des points de consensus), ainsi que des « points aveugles » de la recherche (en d'autres termes, les problématiques qui y sont moins fréquemment traitées et qui pourraient, de fait, constituer des perspectives pour la recherche, pour l'enseignement et pour la formation des enseignants). Au vu des limites imparties à cette note, je m'en tiendrai d'une part essentiellement à la question de l'évaluation des apprentissages et, d'autre part, je ne développerai pas les contenus des travaux que je citerai : les références bibliographiques permettront au lecteur de se reporter directement aux recherches empiriques mentionnées.

Enfin, je travaille moi-même dans le domaine de l'évaluation en langues depuis environ une vingtaine d'années, dans une perspective visant plus spécifiquement à interroger la place et le statut qu'y occupe la diversité (notamment – mais pas uniquement – la diversité linguistique et culturelle). J'ai choisi¹ de travailler cette problématique à partir des apports croisés de la phénoménologie (qui s'intéresse à la manière dont les personnes concernées – dont le chercheur – comprennent les situations et leur donnent sens) et de l'herméneutique (qui travaille à partir de la pluralité des interprétations, en y incluant celles du chercheur). Cette précision me semble nécessaire pour la compréhension de l'état des lieux qui suit : mes choix de travail impliquent en effet que tout état des lieux (comme celui qui est demandé ici aux experts de cette conférence de consensus) constitue une interprétation d'une situation, telle que perçue par tel « expert » particulier, à partir de ses expériences, de ses projets, de ses convictions. Autrement dit, je considère que mon choix de positionnement, ainsi que l'histoire et le projet qui le fondent, colorent ma perception du domaine – comme d'autres orientations / projets / histoires la coloreraient également inévitablement, mais d'une autre manière. Affirmer le caractère situé de cet état des lieux relève donc ici d'un choix éthique, *a fortiori* dans le climat actuel où le questionnement sur les finalités et les valeurs de l'école est fortement engagé et où l'évaluation (au sens large) joue un rôle central, au-delà même de la thématique de l'enseignement des langues vivantes qui occupe la présente note.

¹ Plusieurs publications exposent mon positionnement ainsi que les raisons de ce choix (par exemple : Huver, 2014a ; Huver, 2018, ainsi que Castellotti, 2017, pour une synthèse générale de ces orientations).

B. Quelques tendances dominantes dans la recherche

Les recherches sur l'évaluation en DDL sont peu nombreuses, *a fortiori* les recherches ancrées dans le système scolaire français. En outre, elles s'inspirent très fortement des travaux développés en sciences de l'éducation (désormais SDE) sur l'évaluation en général et sur l'évaluation scolaire en particulier (sans toutefois que les modalités, les enjeux et les questions que posent ces importations soient réellement problématisés). Parmi les recherches sur des situations scolaires, celles qui portent sur les matières « linguistiques » sont essentiellement consacrées à l'évaluation de la langue de scolarisation (en l'occurrence, le français : cf. par exemple les travaux de L. Allal ou encore IFE 2014) et beaucoup moins à l'évaluation des langues dites « vivantes » ou « étrangères » (GFEN 2002). Enfin, si des travaux visant à documenter et à inventorier les pratiques scolaires d'évaluation et de notation ont été développés de manière générique (cf. par exemple le rapport IGEN 2013 ou encore Merle 2018), il n'en va pas de même pour les langues vivantes en particulier (le rapport IGEN 2007 constitue plus un ensemble de préconisations qu'un état des lieux des pratiques).

Autre élément marquant de cet état des lieux : la plupart des recherches actuelles sur l'évaluation en DDL se positionnent par rapport aux préconisations institutionnelles, notamment celles du CECRL, en général pour réfléchir à des modalités de mise en œuvre de ces préconisations.

1. Critique de l'évaluation « sanction » - Des SDE à la DDL

On peut notamment relever la circulation, des SDE vers la DDL, des notions suivantes : évaluation formative / sommative ; produits / processus ; connaissances / compétences ; démarche/ résultats. Les définitions (et les délimitations respectives) de ces notions sont très diverses et ne sont pas consensuelles au sein même des SDE. Néanmoins, les recherches en DDL ont largement intégré ces distinctions : notamment, l'idée d'une pluralité des fonctions de l'évaluation et des démarches qui en résultent est désormais largement admise et courante en DDL (cf. par exemple : Lussier 1992 ; Springer coord., 2001 ; Huver et Springer, 2011 ; Veltcheff et Hilton, 2003 ; Goullier, 2005, ainsi que différents manuels de langue ou encore le tableau proposé dans le CECRL (point 9.3, p.139), qui introduisent cette distinction).

Cette diversification des visées, fonctions et démarches d'évaluation a ouvert en DDL une réflexion autour de la critique d'une évaluation « sanction » et en faveur d'une évaluation dite « positive » (Merle, 2018), c'est-à-dire une démarche :

- qui ne se limite pas à des évaluations notées et qui ne procède pas par focalisation sur les « fautes » et/ou par soustraction de points ;
- qui « rassure » les élèves et favorise l'apprentissage en mettant en évidence leurs réussites et leurs progrès, sur la base notamment de descripteurs et de critères (Ifé, 2014).

La visée sommative de l'évaluation est alors complétée par une visée plus formative, dans laquelle l'évaluation est essentiellement pensée comme une aide à l'apprentissage : en DDL, cela a engendré des travaux visant à développer des critères et des descripteurs (Veltcheff et Hilton, 2003) ainsi que des outils d'évaluation formative (Lussier, 1992) et d'auto/co-évaluation. Quelques travaux, plus rares, se sont attachés non pas tant aux outils qu'aux pratiques d'évaluation, pour tenter d'en

évaluer les effets (cf. par exemple Manoïlov et Tardieu (2015), même si les situations étudiées ne sont pas des situations scolaires mais des tandems français-anglais à l'université).

2. Des contenus « élargis »

Complémentaire à cela, un nombre important de recherches s'attachent à développer des outils d'évaluation dans une perspective méthodologique particulière de manière à assurer une cohérence entre enseignement et évaluation (et, réciproquement, à faire de l'évaluation un levier pour agir sur l'enseignement²). Dans la recherche de langue anglaise, ces recherches concernent le développement d'évaluations « basées sur les tâches » (*task based assessment* – cf. par exemple les travaux de Bachman et Palmer, 2010 – dans une perspective par ailleurs très fortement influencée par la psycholinguistique). Dans la recherche de langue française, elles concernent le développement d'évaluations « communicatives » (Lussier, 1992 ; Bolton, 1987) puis, après la publication du CECRL, d'évaluations « actionnelles » (Cervini et Jouannaud, 2015 ; Bento, 2012 ; Piccardo, 2012 ; Piccardo *et al.*, 2011 ; Bourguignon *et al.*, 2005 ; Tardieu, 2013, par exemple). Les outils créés prennent ici la forme d'évaluations par projets, par tâches, voire par scénarios (cf. par exemple des certifications comme le diplôme de compétences en langues (DCL) ou le certificat de compétences en langues de l'enseignement supérieur (CLES)).

Le point commun de toutes ces recherches est de préconiser (et de développer) des outils d'évaluation plus « authentiques » (*i. e.* proches des usages communicatifs « réels » de la langue). Il en a résulté un élargissement de la portée et des contenus de l'évaluation, puisqu'il s'agit alors d'évaluer des compétences (et non plus seulement des connaissances) communicatives (et non plus seulement linguistiques).

3. Influence du CECRL : constats et conséquences

Depuis les années 1970 et pour l'Europe en tout cas, les recherches en DDL (dont celles sur l'évaluation) sont difficilement séparables des travaux émanant du Conseil de l'Europe et des politiques linguistiques éducatives qui en résultent³.

Pour ce qui concerne l'évaluation, cette influence est sensible à différents niveaux :

- **au niveau de la conception des outils d'évaluation des apprentissages**

Ainsi, les évaluations en langue actuelles :

- s'alignent sur les niveaux du CECRL ;
- visent à développer une perspective actionnelle ;
- recourent aux descripteurs du CECRL (« formulé[s] de manière positive » (CECRL, p. 30)) pour développer une évaluation positive et non centrée sur un locuteur idéal natif ;
- et, quoique plus marginalement, intègrent l'auto-évaluation et/ou une perspective plurilingue (*Portfolio Européen des Langues*, dont une déclinaison a d'ailleurs été réalisée pour le collège (Castellotti *et al.*, 2007). Cf. également Castellotti, 2006).

² Dans la recherche, la cohérence et les effets d'influence réciproque entre évaluation et enseignement sont réfléchis à partir des notions de validité et de *washback effect* (cf. inventaire de ces travaux dans Huver, 2018 et dans Doucet, 2001).

³ Cf. par exemple, pour le domaine scolaire plus particulièrement : *Langues modernes* 2012, Abdelgaber et Medioni coord. (2012). Cf. également Tardieu (2006).

L'articulation au CECRL est également avérée sur le terrain scolaire : ainsi, le rapport IGEN 2007 consacre plus de la moitié de ses pages à l'articulation entre CECRL et nouveaux programmes de LV. Cela se traduit dans les instructions officielles et les programmes par : un alignement des niveaux de langue attendus sur les niveaux du CECRL ; l'inclusion de modalités d'évaluation dites « par compétences », y compris pour les évaluations certificatives (baccalauréat notamment) ; une évolution des référentiels et des critères d'évaluation.

- **au niveau de la recherche**

Les recherches actuelles sur l'évaluation sont elles-mêmes très influencées par les préconisations du CECRL et s'inscrivent pratiquement uniquement dans cette perspective. Cette tendance s'explique par la conjonction d'un ensemble de facteurs plus généraux, qui dépassent le seul CECRL :

- le fait que la DDL se considère essentiellement comme une discipline d'intervention (devant donc produire des outils et des dispositifs didactiques) ;
- la demande institutionnelle orientée vers la création d'outils et de pratiques efficaces et généralisables (les « bonnes pratiques ») ;
- la situation actuelle de sous-financement de la recherche publique en France, les appels d'offre institutionnels constituant en effet des sources de financement essentielles pour les recherches, tout en influençant aussi inévitablement la recherche elle-même.

Ceci a pour conséquence que la définition des finalités et des « besoins » revient de plus en plus aux institutions, la recherche étant alors (auto)envisagée dans son rôle d'accompagnement et de mise en œuvre d'orientations pensées au moins en partie en dehors d'elle (Castellotti, Huver et Leconte, 2017 : pour un exemple, cf. la récente amplification du CECRL qui a essentiellement consisté en un travail ingénierique de conception de nouveaux descripteurs sans (ré)interroger les enjeux et la situation historique au fondement de la version initiale du CECRL).

- **au niveau de l'évaluation des systèmes éducatifs**

Par la constitution de banques d'indicateurs échelonnés, le CECRL accompagne et rend possible une extension de la portée de l'évaluation. Il ne s'agit alors plus seulement d'évaluer les apprentissages mais aussi les systèmes éducatifs dans le cadre d'évaluations standardisées (PISA, *SurveyLang*), à des fins de comparaisons intra- et inter-nationales, avec, souvent, un objectif de pilotage par les résultats.

Cette visée de l'évaluation a été très largement discutée et problématisée par la recherche en sciences humaines (Cassin, 2014 ; Martuccelli, 2010 ; Meirieu, 2018 par exemple). Complémentairement à ces critiques d'ordre politique, on peut également se demander dans quelle mesure l'appréhension de l'évaluation par le biais de grilles et de descripteurs étalonnés et standardisés ne participe pas plus largement au renforcement de représentations relevant d'un « imaginaire du contrôle » (Ardoino et Berger, 1986) et d'un « nouveau scientisme » (Aflalo, 2009), peut-être encore plus pour les LV que pour d'autres disciplines scolaires où ces descripteurs n'existent pas, et ce, *a fortiori* depuis la parution du *Volume Compagnon* du CECRL, qui confirme (et donc renforce) cette logique (Huver, 2018b).

C. Points aveugles et perspectives de recherche

La prévalence de ces orientations et de ces thématiques de recherche conduit également à ce que soient minorées d'autres orientations, thématiques et pistes possibles pour la recherche, mais aussi pour l'enseignement et la formation des enseignants. J'en relèverai ici trois principales.

1. Usages évaluatifs

La centration des recherches en évaluation sur la création d'outils et la mise en œuvre de préconisations laisse en friche tout un pan de réflexion sur les pratiques et les représentations (ce que j'appelle « les usages »⁴) des enseignants, des élèves, des parents, sur les liens entre ces usages et les apprentissages, que ce soit dans une perspective cognitive (construction des apprentissages), psychologique (effets sur la motivation par exemple) ou sociologique (effets sur la reproduction sociale par exemple).

La question des **représentations** de l'évaluation est certes évoquée en SDE (Veslin et Veslin, 1992 ; Merle, 1998 et 2007 ; Barlow, 2003, etc.), ou en psychologie sociale (Butera *et al.*, 2011), usuellement pour mettre en évidence leur caractère négatif. Elle est en revanche plus rarement abordée en DDL. On peut néanmoins recenser les travaux de Shohamy (2001) sur les « effets destructeurs » de l'évaluation, ceux de Barbé (1990) et Chardenet (1999) qui évoquent les représentations négatives attachées à l'évaluation, ainsi que certains de mes travaux du début des années 2010 qui complexifient ce constat en mettant en évidence la forte ambivalence de ces représentations.

De même, les **pratiques** des enseignants et des élèves sont très peu connues et documentées pour les LV (sur les pratiques d'évaluation scolaire en général : *cf.* Merle, 2018). Il serait donc intéressant de développer des recherches sur **leurs usages évaluatifs et leurs vécus de l'évaluation**, dans une perspective compréhensive. D'autres travaux (Tardieu, 2010 ; Normand-Marconnet, 2012, mais aussi Aslan, en cours) s'intéressent à la dimension (inter)culturelle de ces usages, ouvrant une voie encore peu explorée vers des recherches à orientation anthropologique. Ceci serait d'autant plus intéressant pour l'EN qu'il n'existe actuellement aucune recherche sur les usages évaluatifs relatifs aux différentes LV enseignées. Or, on peut se demander si « on évalue de la même façon » en anglais, en allemand, en chinois, en russe ou en arabe, alors ces langues ont des systèmes linguistiques différents, des histoires différentes, des traditions d'enseignement différentes et qu'elles véhiculent des représentations et des rapports différents à la norme.

2. Diversité linguistique et culturelle ; contacts de langues

La thématique de la diversité linguistique et culturelle n'est pas totalement absente des recherches sur l'évaluation en langues (*cf.* par exemple, Huver, 2010 ; Huver et Springer, 2011 ; Chardenet, 2011 ; Partie 3 de Huver et Ljalikova, dir., 2013). Mais du fait des orientations dominantes précédemment évoquées, elle est essentiellement travaillée sous l'angle de l'élaboration de dispositifs et d'outils pour faire suite à des demandes institutionnelles (pour des exemples, *cf.* Lenz et Berthele, 2010 ; Candelier coord., 2007 ; Castellotti *et al.*, 2007).

⁴ J'entends ici par « usage », à la fois les pratiques et les représentations, considérées comme indissolublement liées.

Peu de travaux, en revanche, problématisent la pluralité linguistique et culturelle en tant qu'enjeu qui traverse et travaille les problématiques évaluatives (cf. toutefois Chardenet, 1999 et 2011, sur les fondements langagiers de l'activité évaluative). Or, les travaux en sociolinguistique, en acquisition et en DDL sur les contacts de langues montrent que le « mélange des langues » est courant dans les usages plurilingues quotidiens et dans les classes de langue. Cependant, peu de travaux existent sur les implications de ce constat pour l'évaluation (*i. e.* sur les rapports entre norme et variation dans les usages évaluatifs). Ceci présente pourtant des enjeux forts pour l'EN, dans la mesure où ces problématiques sont fortement agissantes dans les situations scolaires. Les élèves sont en effet en contact avec d'autres variétés des langues que les variétés normées, scolaires, de référence (vocabulaire, accent etc.), du fait :

- de la diversité de leurs expériences et de leurs répertoires linguistiques et culturels ;
- de leurs usages langagiers, usages d'appropriation et usages évaluatifs, influencés par l'en dehors de la classe (réseaux sociaux, mobilités, langues familiales, etc.).

Différentes questions mériteraient donc d'être posées :

- Quel traitement des dimensions inter- et translinguistiques des répertoires des élèves⁵ ?
- Quelles représentations (pour les enseignants et les élèves) des contacts de langues : une perception négative (des « fautes ») ? ; positive (des « erreurs ») ? à favoriser (en tant que stratégies plurilingues et occasions pour réfléchir aux usages plurilingues) ? Les travaux novateurs de M. Marquilló Larruy (2003 et 2013) sur l'interprétation de l'erreur dans une perspective plurilingue devraient de ce point de vue donner lieu à des prolongements, en l'occurrence sur le terrain éducatif français.
- Quelles relations entre usages évaluatifs et insécurité linguistique ? On peut supposer qu'une conception très normative de la langue pèse fortement sur les usages évaluatifs, ce qui contribue à une insécurisation linguistique des élèves mais aussi des enseignants, qui peut gêner voire empêcher l'apprentissage ou, tout du moins, la prise de risque.

Ces questions sont d'autant plus importantes pour l'EN que :

- le rapport particulier entretenu au français et à la diversité linguistique en France est vecteur de rapports particuliers aux langues, à la norme et à la correction linguistique en général, qui jouent sans doute sur les usages évaluatifs. En outre, le territoire français présente des situations de diversité linguistique et culturelle « intense » (DOM, langues régionales, migrations, etc.), avec des enjeux identitaires vifs, sur lesquels il semble pour le moins difficile de faire l'impasse. Il serait donc nécessaire de développer des réflexions situées sur les questions d'évaluation en langues dans les départements d'outre-mer, dans certains dispositifs éducatifs bilingues (en langue régionale par exemple), ou encore pour certaines démarches travaillant avec la pluralité linguistique. Réciproquement, ces situations ont souvent constitué des « incubateurs » de réflexion productifs et originaux, quoique peu connus, pour les questions d'évaluation (*cf.* par exemple la notion de polynomie pour l'enseignement/apprentissage du corse, qui aboutissent à des formes de relativisation de la notion d'erreur par acceptation de plusieurs variantes).

⁵ En lien également avec le français en tant que langue de scolarisation (vs un enseignement compartimenté des langues et, plus généralement, des disciplines scolaires).

- envisager l'évaluation sous l'angle des contacts de langues et de l'insécurité linguistique permet de déplacer la réflexion des méthodologies d'enseignement aux enjeux sociolinguistiques et interculturels de l'apprentissage. Ce déplacement touche également les interprétations des enquêtes internationales (type *Surveylang*) en déplaçant le focus de la préconisation de « nouvelles » méthodes d'enseignement et d'évaluation vers le rapport des enseignants, des parents et des élèves de France à la norme linguistique.

3. Dimensions politiques et épistémologiques de l'évaluation

Depuis les années 1990, les recherches anglophones (ou d'inspiration nord-américaine) ont développé des travaux sur l'évaluation en langues inscrits dans le courant dit « critique », qui visent à mettre en évidence les idéologies sous-jacentes et les effets pervers des évaluations standardisées, certificatives (Shohamy, 2001, Mc Namara, 2010, McNamara et Roever, 2006 par exemple) ; d'autres, dans une perspective historicisante (Spolsky, 1995) ou philosophique (Fulcher, 2014), ont pour objectif de mettre en évidence le caractère socialement construit et historiquement situé de ces évaluations et, ce faisant, les choix qui ont fondé leur conception et leur diffusion. Ces orientations sont en revanche peu représentées dans les travaux de langue française, même si elles commencent à émerger (par exemple, Zarate *et al.*, 2004 ; Coste, 2013 ; Anquetil *et al.*, 2017 ; Huver, 2014b et 2018a et b). La centration des recherches sur le développement d'outils d'évaluation induit, par un phénomène de vases communicants, une plus faible place pour des réflexions questionnant la demande institutionnelle et réfléchissant, sur le long terme ou autrement que dans les termes institutionnellement formulés, les enjeux liés à l'évaluation en langues, au-delà notamment d'une orientation communicative/actionnelle – ce qui serait pourtant nécessaire, étant donné les mutations rapides de la société et du système éducatif dans un environnement mondialisé.

Enfin, une réflexion devrait également être engagée sur les enjeux et les implications (politiques, épistémologiques, éthiques) liés au recours croissant à l'automatisation, aux algorithmes et à l'intelligence artificielle dans l'évaluation, *a fortiori* suite au développement de différents outils d'évaluation automatisés, à distance ou impliquant de très grands groupes (Moocs, évaluation de positionnement, etc. ; cf. par exemple Bourdet et Teutsch (2000), Cervini et Jouannaud (2015) ou encore les outils développés par les services « recherche et développement » de quelques structures d'enseignement ou organismes certificateurs).

Conclusion

Quelques tendances dominantes ressortent de la recherche sur l'évaluation en langues autour de : l'évaluation positive ; l'élargissement de la portée de l'évaluation des connaissances aux compétences et des dimensions linguistiques aux dimensions communicatives ; une conception de l'évaluation (et notamment l'auto-évaluation) comme soutien à un apprentissage réflexif. La forte influence du CECR est également une tendance contemporaine marquante. Cependant, ces tendances dominantes révèlent en creux des « points aveugles », peu explorés alors qu'ils présentent des enjeux forts pour la DDL en général et pour les LV à l'école en particulier.

Au-delà de ces constats, l'évaluation renvoie à des problématiques sur lesquelles il y a peu de consensus, y compris au sein de la recherche : ainsi, une définition commune, de même que les tendances dominantes sus-citées y provoquent des débats. Ceux-ci renvoient plus largement à la question des finalités de l'enseignement/apprentissage des langues, aux valeurs et convictions des

personnes impliquées, mais aussi aux conceptions de la recherche en DDL (plus ou moins « appliquée », *i. e.* prioritairement portée sur l'élaboration d'outils et de dispositifs, vs « fondamentale » et théorisante); aux disciplines convoquées (SDE, sociologie, psychologie, neurosciences, etc.); aux théorisations des notions fondant le domaine (théorisations des notions de langue, d'enseignement, d'apprentissage, d'évaluation, etc.).

Ainsi, les choix effectués ne sont pas anodins, non seulement pour l'évaluation scolaire et les usages évaluatifs en classe, mais aussi pour les recherches qui seront initiées et soutenues et, enfin, pour la formation des enseignants, en constante tension dialectique entre formation technique à un usage conforme des outils dans une visée d'efficacité, et formation réflexive sur les représentations (de la langue et de l'évaluation), les valeurs et les projets individuels et collectifs au fondement de ces usages.

Bibliographie

Abdelgaber, S. et Medioni, M.-A., coord. (2012). Enseigner les langues vivantes avec le *Cadre européen. Cahiers pédagogiques*, hors-série numérique n°8.

Aflalo, A. (2009). L'évaluation : un nouveau scientisme. *Cités* n° 37, p. 79-89. <https://www.cairn.info/revue-cites-2009-1-page-79.htm>.

Anquetil, M., Derivry, M. et Gohard-Radenkovic, A. (2017). En finir avec le Je contraint et réifié dans l'objet PEL pour une didactique de la biographie langagière comme processus relationnel. *TDFLE*, n 70, *La pensée CECRL*, <http://revue-tdfle.fr/les-numeros/numero-70/23-en-finir-avec-le-je-contraint-et-reifie-dans-l-objet-pel-pour-une-didactique-de-la-biographie-langagiere-comme-processus-relationnel>.

Ardoino, J. et Berger, G. (1986). L'évaluation comme interprétation. *POUR*, n°107, pp.120-127.

Aslan, A. (en cours). *L'altérité et la diversité au défi des évaluations certificatives : réflexions à partir du DELF: histoire, usages, réceptions*, Thèse de doctorat, sous la direction d'E. Huver, <http://www.theses.fr/s174431>.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice : developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.

Barbé, G. (1990). Se former à l'évaluation ou l'évaluation pour se former ? *Études de linguistique appliquée*, n°80, pp.119-131.

Barlow, M. (2003). *L'évaluation scolaire, mythes et réalités*. Paris : ESF.

Bento, M. (2012). L'évaluation des langues dans une perspective actionnelle. *Les Langues modernes*, n°1, p.12-17.

Bolton, S. (1987). *Évaluation de la compétence communicative en langue étrangère*. Paris : Hatier.

Bourdet, J.-F. et Teutsch, P. (2000). Définition d'un profil d'apprenant en situation d'auto-évaluation. *Alsic*, Vol. 3, n°1, <http://journals.openedition.org/alsic/1658>.

Bourguignon, C. Delahaye, P. et Vicher, A. (2005). L'évaluation de la compétence en langue : un objectif commun pour des publics différents. *Études de linguistique appliquée (ÉLA)*, 2005/4, n° 140. pp. 459-473.

Butera, F., Buchs, C. et Darnon, C. (2011). *L'évaluation une menace ?* Paris : PUF.

Candelier, M. (coord.) (2007). *CARAP – Cadre de référence pour les approches plurielles des langues et des cultures*, Conseil de l'Europe, Graz.

<http://carap.ecml.at/LinkClick.aspx?fileticket=%2ff86MKs659s%3d&tabid=425&language=fr-FR>.

Cassin, B. (2014). *Derrière les grilles. Sortons du tout-évaluation*. Paris : Fayard / 1001 nuits.

Castellotti, V. (2017). *Pour une didactique de l'appropriation. Diversité, compréhension, relation*. Paris : Didier.

Castellotti, V. (2006). Éducation bi-plurilingue et évaluation : le rôle des portfolios. *Le français dans le monde*, n° 348, pp. 18-19.

Castellotti, V., Coste, D., Moore, D. et Tagliante, C. (2007). *Portfolio européen des langues – Collège*. Paris : Didier/CIEP.

Castellotti, V., Huver, E. et Leconte, F. (2017). Demande institutionnelle et responsabilité des chercheurs : langues, insertions, pluralité des parcours et des perceptions. In : Beacco, J.-C., Krumm,

H.-J., Little, D. et Thalgott, P. (eds.), *L'intégration linguistique des migrants adultes. Les enseignements de la recherche*, De Gruyter Mouton.

<https://www.degruyter.com/downloadpdf/books/9783110477498/9783110477498-058/9783110477498-058.pdf>.

Cervini, C. et Jouannaud, M.-P. (2015). Ouvertures et tensions liées à la conception d'un système d'évaluation en langues, numérique, multilingue et en ligne, dans une perspective communicative et actionnelle. *Alsic*, vol. 18, n° 2, <http://journals.openedition.org/alsic/2821>.

Chardenet, P. (2011). L'évaluation, un domaine de connaissance complexe entre théorie et pratiques sociales. Blanchet P. et Chardenet P. (dir.), *Guide pour la recherche en didactique des langues et des cultures*. Paris : Editions des archives contemporaines.

Chardenet, P. (1999). *De l'action évaluative à l'acte d'évaluation. Approche théorique et pratique communicationnelle*. Paris : L'Harmattan.

Coste, D. (2013) Petit retour à Rüsclikon, pp.140-148. In : Huver, E. et Ljalikova, A., (coord.), 2013, « L'évaluation en didactique des langues et des cultures : continuités, tensions, ruptures », *Recherches et applications – Le français dans le monde*, n°53.

Doucet P. (2001). Pour un test utile. *ASp*, n°34, <http://journals.openedition.org/asp/1696>.

Fulcher, G. (2014). Philosophy and language testing. In: Kunnan, A.-J. dir., *The Companion to language assessment*, John Wiley and sons ed.

GFEN (2002). *Réussir en langues*. Lyon : Chronique sociale.

Goullier, F. (2005). *Les outils du Conseil de l'Europe en classe de langue*. Paris : Didier.

Huver, E. (2018a). « Ce qui fait enjeu » dans les évaluations linguistiques : Définitions, circulations, implications. In Suzuki et al. coord., *Penser la didactique du plurilinguisme et ses mutations : idéologies, politiques, dispositifs*. Rennes : PUR. <https://hal.archives-ouvertes.fr/hal-01817787>.

Huver, E. (2018b). Peut-on (encore) penser à partir du CECRL ? Perspectives critiques sur la version amplifiée. *Mélanges. CRAPEL. Éclectisme en didactique des langues : hommage à Francis Carton*, n°38. <http://www.atilf.fr/spip.php?article4222>

Huve,r E. (2014a). *De la subjectivité en évaluation à une didactique des langues diversitaire. Pluralité, altérité, relation, réflexivité*. Habilitation à diriger des recherches, université de Tours, NB : Consultable sur HAL : <https://hal.archives-ouvertes.fr/tel-01211154v1>.

Huver, E. (2014b). CECRL et évaluation : interprétations plurielles et logiques contradictoires. *Cahiers du GEPE*. <http://www.cahiersdugepe.fr/index.php?id=2652>.

Huver, E. (2010). J'ai pas pu corriger parce que c'est le mélange des langues - L'évaluation de la compétence plurilingue entre résistances et tensions. *Cahiers de l'ACEDLE*, n° 8, http://acedle.org/IMG/pdf/Huver_Cahiers-Acedle_7-2.pdf.

Huver, E. et Ljalikova, A., (coord.) (2013). L'évaluation en didactique des langues et des cultures : continuités, tensions, ruptures. *Recherches et applications – Le français dans le monde*, n°53.

Huver, E. et Springer, C. (2011). *L'évaluation en langues*, Paris : Didier.

Ifé (2014). *Evaluer pour (mieux) faire apprendre*. Dossier de veille de l'Ifé, n°94, <http://veille-et-analyses.ens-lyon.fr/DA-Veille/94-septembre-2014.pdf>.

IGEN (2007), *L'évaluation en langue vivante : état des lieux et perspectives d'évolution*. Rapport - n°2007-009, Ministère de l'Education Nationale.

<http://media.education.gouv.fr/file/45/2/4452.pdf>.

IGEN (2013). *La notation et l'évaluation des élèves éclairées par des comparaisons internationales*. Rapport n 2013-072, Ministère de l'Éducation Nationale.

Lado, R. (1961). *Language testing: the construction and use of foreign language tests*. Bristol: Longmans.

Lenz, P. & Berthele, R. (2010). *Assessment in Plurilingual and Intercultural Education*. Conseil de l'Europe, Strasbourg.

http://www.coe.int/t/dg4/linguistic/ListDocs_Geneva2010.asp#TopOfPage.

Les Langues modernes, n°1/2012 : « Évaluer avec le CECRL »

Lussier, D. (1992). *Évaluer les apprentissages dans une approche communicative*. Paris : Hachette, Collection Autoformation.

Mc Namara, T. (2010). The use of language tests in the service of policy: issues of validity. *Revue française de Linguistique Appliquée*. XV-1, pp. 7-23.

Mc Namara, T. et Roever, C. (2006). *Language Testing: The Social Dimension*, Oxford (UK): Blackwell Publishing.

Manoïlov, P. et Tardieu, C. (2015). Rétroaction corrective et reprise dans le contexte de l'apprentissage en tandems Français-Anglais à l'université. *Recherches en didactique des langues et des cultures*, n°12-3, <http://journals.openedition.org/rdlc/986>.

Marquilló Larruy, M. (2013). Dans les valises du plurilinguisme : quelle conception des normes et des erreurs ? In : Huver, E. et Ljalikova, A., (coord.), (2013). *L'évaluation en didactique des langues et des cultures : continuités, tensions, ruptures. Recherches et applications – Le français dans le monde*, n°53, pp113-125.

Marquilló Larruy, M. (2003). *L'interprétation de l'erreur*. Paris : Clé international.

Martuccelli, D. (2010). Critique de la philosophie de l'évaluation. *Cahiers internationaux de sociologie*, n° 128-129, pp.27-52.

Meirieu, P. (2018). Philippe Meirieu, l'heure de la riposte. *Le café pédagogique*, <http://www.cafepedagogique.net/lexpresso/Pages/2018/08/29082018Article636711250818462828.aspx>

Merle, P. (2018). *Les pratiques d'évaluation scolaire. Historique, difficultés, perspectives*. Paris : PUF.

Merle, P. (2007) *Les Notes – Secrets de fabrication*. Paris : PUF.

Merle, P. (1998) *Sociologie de l'évaluation scolaire*. Paris : PUF.

Normand-Marconnet, N. (2012). L'auto-évaluation comme facteur d'évolution du rapport à l'apprentissage des langues. *Recherches en didactique des langues et des cultures*, n°9-1, <http://journals.openedition.org/rdlc/2476>.

Piccardo, E. (2012). Multidimensionality of assessment in the Common European Framework of Reference for languages. *Cahiers de l'ILOB*, n° 4, p. 37-54.

Piccardo, E., Berchoud, M., Cignatta, T., Mentz, O & Pamula, M. (2011). *Parcours d'évaluation d'apprentissage et d'enseignement à travers le CECRL*. Graz, CELV

Shohamy, E., (2001). *The power of Tests, A critical perspective on the uses of language tests*. Essex: Pearson Education.

Spolsky, B. (1995). *Measured word. The development of objective language testing*. Oxford: Oxford University Press.

Springer, C. Coord. (2001). Evaluation et certifications en langue. *Les Langues modernes*, n°2.

Tardieu, C. (2013). Testing et certification. *Recherches en didactique des langues et des cultures*, n°10-2, <http://journals.openedition.org/rdlc/1655>.

Tardieu, C. (2010). Votre B1 est-il mon B1 ? L'interculturel dans les tests d'évaluation en Europe. *Recherches en didactique des langues et des culture*, 7-2, <http://journals.openedition.org/rdlc/2301>.

Tardieu, C. (2006). L'évaluation en langues : quelles perspectives ? *Recherches en didactique des langues et des cultures*, n°2, http://acedle.org/old/IMG/pdf/Tardieu-C_cah2.pdf.

Veltcheff, C. et Hilton, S. (2003). *L'évaluation en FLE*. Paris : Hachette Français Langue Etrangère.

Veslin, O. et Veslin, J. (1992). *Corriger des copies : évaluer pour former*. Paris : Hachette.