

RECHERCHES
EN DIDACTIQUE
DES LANGUES
ET DES CULTURES

Recherches en didactique des langues et des cultures

Les cahiers de l'Acedle

7-2 | 2010

Les langues tout au long de la vie : Permanences et évolutions en didactique des langues

Votre B1 est-il mon B1 ? L'interculturel dans les tests d'évaluation en Europe

Claire Tardieu



Édition électronique

URL : <http://journals.openedition.org/rdlc/2301>

DOI : 10.4000/rdlc.2301

ISSN : 1958-5772

Éditeur

ACEDLE

Référence électronique

Claire Tardieu, « Votre B1 est-il mon B1 ? L'interculturel dans les tests d'évaluation en Europe », *Recherches en didactique des langues et des cultures* [En ligne], 7-2 | 2010, mis en ligne le 01 octobre 2010, consulté le 19 avril 2019. URL : <http://journals.openedition.org/rdlc/2301> ; DOI : 10.4000/rdlc.2301

Ce document a été généré automatiquement le 19 avril 2019.



Recherches en didactique des langues et des cultures is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Votre B1 est-il mon B1 ? L'interculturel dans les tests d'évaluation en Europe

Claire Tardieu

1. Introduction

- 1 Le *Cecrl* est présenté par son éditeur comme un "*instrument pratique qui permet d'établir clairement les éléments communs à atteindre lors des étapes successives de l'apprentissage*" et également "*comme un instrument idéal pour la comparabilité des résultats de l'évaluation dans les pays du conseil de l'Europe*". Mais qu'en est-il vraiment ? Les projets européens successifs comme celui de la grille hollandaise (Dutch CEFR grid) ou de la banque européenne d'items d'ancrage (Ebafls) ont montré que l'outil européen ne suffisait pas pour établir des équivalences de niveau entre les tests européens en langues. Autrement dit, un item qui dans mon pays est réussi par un élève au niveau B1 peut l'être dans un autre pays par un élève à un niveau A2 seulement. Cette communication ne traitera pas des différences culturelles en matière d'apprentissage, mais uniquement en matière d'évaluation, dans le domaine très spécifique des tests.
- 2 Quels sont les biais observés qui entravent la possibilité d'établir une banque européenne d'items par niveaux du *Cecrl* (Vijver & Hambleton, 1996) ?
- 3 Au-delà de ces biais, est-il vraiment possible, voire souhaitable, qu'un outil existe et éradique les différences culturelles en matière d'évaluation ? Ne doit-on pas au contraire s'intéresser aux différences de valeurs et de référents accordés à des types de test qui paraissent, de prime abord, identiques (Blommaert, 2004 ; Zarate, Lévy, Kramersch, 2008) ?
- 4 Pour autant, il n'est pas forcément inapproprié d'utiliser le *Cecrl* comme outil d'estimation empirique, ouvert à tous, sans volonté d'uniformisation, sorte de plus petit dénominateur commun dans le domaine de l'enseignement-apprentissage des langues, comme on pourra le voir avec le projet Cef-estim (2009).

2. Quels sont les biais culturels en matière d'évaluation observés à travers le projet européen Ebafls ?

- 5 Van de Vijver et Hambleton (1996) ont dressé une taxonomie des biais culturels en matière d'évaluation. Le premier biais est conceptuel et dû à la non-équivalence des concepts d'un groupe culturel à l'autre. Le second est méthodologique résultant de la manière d'administrer les instruments. Le troisième est lié aux items eux-mêmes (des traductions inadéquates par exemple).

Au cours du projet Ebafls, ces différents biais ont pu être observés.

2.1. Le projet Ebafls

- 6 Ebafls visait à produire une banque d'items d'ancrage correspondant à un niveau linguistique défini. Cette banque devait permettre aux pays de connaître de manière empirique le niveau de langue étrangère de leurs élèves et de faire des comparaisons internationales. Le niveau de référence retenu était le niveau B1 du CECR. Les items devaient être discriminants entre ce niveau et les niveaux proches A2 et B2 et couvrir la compréhension écrite et orale en anglais, français et allemand. Il a été décidé d'emprunter des items existants, déjà utilisés dans des tests standardisés dans les pays et fournis par les pays. Ceci devant permettre de définir si le niveau B1 est le même dans chaque pays.
- 7 Pour la mesure, la méthode utilisée est celle des modèles de réponses à l'item (IRT). Cette théorie est basée sur l'item et non sur le test entier. Elle lie les caractéristiques de l'item et la personne qui y répond à la probabilité que cette personne réponde.
- 8 Le modèle utilisé est le "One Parameter Logistic Model" (Oplm). Ce modèle pose comme hypothèse que les compétences des élèves sont unidimensionnelles et que la probabilité de donner la réponse correcte augmente avec le niveau de compétence de l'élève. La difficulté est déterminée par la localisation de la réponse à l'item sur l'échelle de compétences et la discrimination par la pente.
- 9 Les paramètres de discrimination définissent dans quelle mesure une réponse à l'item est due à la compétence et ils sont utilisés pour donner un score aux réponses.

2.2. Exemples de biais conceptuel

2.2.1. Le choix de L1 ou de L2 dans les consignes et les items (Ebafls)

- 10 Un biais conceptuel réside dans le choix de la langue utilisée pour le test. En France et en Écosse, on considère que les consignes et les questions de compréhension doivent être rédigées dans la langue de scolarisation. Dans les six autres pays concernés par le projet, le choix de la langue 2 s'impose comme une évidence.
- 11 Les Français et les Écossais participant au projet ont tenté d'expliquer que la compréhension des consignes et des questions pouvait exiger un niveau B1 qui n'était pas forcément celui atteint par l'élève passant le test. Les autres participants ne semblaient

pas connaître ce dilemme et avaient pour tradition d'administrer tous les tests de langue en L2.

- 12 En ce qui concerne la France, on peut s'étonner de l'utilisation de la L1 dans les tests alors que l'utilisation de la L2 pour les consignes et la langue de la classe est recommandée dès l'école.

2.2.2. La différence de jugement

- 13 Pour lier les items aux CECRL deux méthodes ont été utilisées :
- La spécification des contenus des tests
 - La standardisation des jugements, des notations
- 14 Pour la spécification, les experts ont utilisé les paramètres de la *Dutch grid*.
- 15 Pour la standardisation, des personnes ont été formées afin que les procédures soient les mêmes dans tous les pays. Ces personnes formaient à leur tour d'autres personnes dans leur pays (majoritairement des enseignants du second degré). Il fallait aussi obtenir une compréhension commune des niveaux du CECRL (standardiser les jugements). Le but était que les juges puissent déterminer le niveau minimum demandé à un candidat pour qu'il réponde correctement à une ou plusieurs questions. En ce qui concerne les items de compréhension écrite, on a en outre demandé aux enseignants de langue d'indiquer pour cinq ou six de leurs élèves sélectionnés, s'ils étaient capables ou non de mener les tâches reportées sur une liste.
- 16 L'analyse des résultats de la sélection du niveau des items montre que les enseignants ne sont pas d'accord entre eux. Les enseignants ont du mal à estimer si les items appartiennent au niveau A2, B1 ou B2. De plus, on observe des différences dans les scores, le nombre de bonnes réponses des candidats aux différents niveaux du CECRL.
- 17 Ainsi, deux différentes méthodes de standardisation ont été utilisées, mais aucune n'a conduit à l'affirmation de l'existence d'un lien ferme entre EBAFLS et le CECRL.
- 18 Pour l'oral, il existe également des différences par rapport aux niveaux du CECRL et entre les pays.

2.3. Exemple de biais liés à l'administration des instruments

2.3.1. Les consignes en L1 ou L2 (EBAFLS)

- 19 Concernant l'impact du choix de la langue pour l'administration du test, il fut décidé de faire un pilote afin de mesurer d'éventuels écarts dans les scores de réussite. Un protocole fut établi en compréhension de l'écrit pour le français, et administré dans deux pays, la Suède et l'Écosse. Les résultats montrent en premier que le fait de traduire tous les items dans la langue cible induit des réponses qui n'avaient pas été prévues par les concepteurs du test original dans les consignes de correction ou leur traduction. Autrement dit, le fait de changer la langue des consignes et des items a modifié le caractère même des items d'une manière imprévisible.
- 20 En second lieu, il apparaît que les scores ne sont pas les mêmes selon que le test est administré en L1 ou en L2., la balance ayant tendance à privilégier l'administration du test en L1. Les items choisis pour le pilote étaient de trois sortes : réponses courtes, questions à choix multiple, phrases lacunaires.

Tableau 1 : Résultats du pilote L1-L2

Scores de réussite par pays (Suède et Écosse) et par condition (L1/L2 ou L2)						Analyse des résultats	
						<i>Items de Compréhension écrite</i>	
	Item	L1 Sw	L2 Sw	L1/Sc	L2 Sc		
	1	.761	.943	.899	.965		•→ Items 1 à 8 : réponses courtes
	2	.572	.465	.643	.519		
	3	.874	.956	.747	.949		Suède : L1 < L2 (3 items sur 8 mieux réussis)
	4	.906	.811	.902	.892		
	5	.918	.937	.918	.921		Écosse : L1 = L2 (4 items pour chaque condition)
	6	.906	.950	.963	.959		
	7	.786	.884	.826	.908		•→ Items 9 et 10 : QCM
	8	.830	.736	.884	.759		
	9	.516	.497	.570	.554		Suède et Écosse : L1 > L2 (2 items sur 2 pour chaque pays)
	10	.472	.371	.274	.152		•→ Items 11 à 19 : lacunaires
	11	.925	.654	.802	.658		
	12	.692	.459	.787	.557		Suède et Écosse : L1 > L2 (9 items sur 9 pour la Suède, 8 sur 9 pour l'Écosse).
	13	1.000	.792	.973	.598		
	14	.717	.308	.216	.079		
	15	.862	.333	.405	.180		
	16	.799	.428	.838	.421		
	17	.755	.509	.698	.434		

	18	.484	.465	.415	.459				
	19	.566	.421	.448	.348				

- 21 Il apparaît néanmoins que la réussite des élèves suédois et écossais varie selon la langue d'administration du test en fonction du type d'item. Dans le cas des réponses courtes, l'administration en L2 est favorable aux élèves suédois et équitable pour les élèves écossais pourtant habitués aux consignes en L1, ce qui relativise la conclusion privilégiant l'administration en L1 (Il faudrait faire un pilote sur tous les types d'items).

2.3.2. La différence de niveaux selon les pays

- 22 Les langues ne sont pas forcément enseignées au même moment du cursus ni avec les mêmes objectifs. Ainsi, pour contrecarrer ce biais, la collecte de données en compréhension écrite qui s'est déroulée entre mars et juin 2006 et a permis de récupérer 11712 livrets, s'est effectuée auprès d'élèves d'âges différents : les élèves avaient entre 16 et 18 ans sauf pour l'épreuve d'anglais en Allemagne et au Luxembourg où les élèves avaient entre 14 et 16 ans. En fait, les écoles et les élèves étaient choisis sur la base de leur niveau B1 présumé, tout en assurant une diversité d'écoles.
- 23 Quant au test de français, il a été passé par de jeunes écossais de classe de première (et non de fin de collège ou de seconde) qui n'ont que deux ans de français.

2.3.3. La perception du test par les professeurs

- 24 En compréhension orale, un seul test a été conduit pour chaque langue en prenant soin qu'il contienne des items de tous les pays qui en ont fournis. La sélection des items et des élèves était similaire à celle du pré-test d'écrit. Il a été administré début 2007. Il y a eu 3721 réponses. Les élèves et enseignants ont également rempli un questionnaire. Un consensus s'est établi sur les points suivants : Tous les enseignants, dans les trois langues, s'accordaient pour dire que la qualité des enregistrements n'était pas bonne. La plupart pensaient qu'il aurait fallu passer l'enregistrement deux fois plutôt qu'une et certains, que les instructions n'étaient pas claires et le vocabulaire quelque fois trop difficile. Pour l'anglais, les remarques n'ont porté que sur le son.
- 25 En revanche, on remarque des disparités importantes concernant l'appréciation de la difficulté. Les enseignants suédois ont trouvé le test d'allemand trop rapide, ne laissant pas assez de temps pour répondre, alors que les espagnols l'ont trouvé trop lent. Les enseignants ont trouvé les textes de français trop difficiles à comprendre (trop rapides, vocabulaire trop difficile). Les espagnols l'ont trouvé trop lent.
- 26 Les élèves ont fait, de manière notable, les mêmes commentaires que leurs enseignants (mauvaise qualité du son, trop rapide, vocabulaire trop difficile, trop lent...).

2.4. Exemple de biais liés aux items

2.4.1. La différence de DIF (Ebafls)

- 27 Une différence essentielle s'est faite jour à l'issue du projet Ebafls. A l'exception des items d'anglais en compréhension orale, les analyses révèlent beaucoup de DIF (fonctionnement différentiel à l'item) au point qu'il est impossible d'établir des échelles concomitantes de pays à pays. Ce qui revient à dire que le même item ne fonctionne pas de la même manière selon le pays. Ou encore qu'un élève au niveau B1 dans un pays n'est pas équivalent à un élève au niveau B1 dans un autre pays.

2.4.2. En termes de difficulté de l'item

- 28 L'analyse des résultats de la pré-enquête montre des exemples où il existe un fonctionnement différentiel d'item dû à la difficulté de l'item. La difficulté est déterminée par la localisation de la réponse à l'item sur l'échelle de compétences. Il existe des items pour lesquels on constate de grosses différences de probabilités de réussite entre les pays, pour des élèves de niveau de compétences comparable. Ceci peut s'expliquer par le type d'items auquel l'élève est habitué.
- 29 Par exemple, les élèves écossais réussissent en général mieux les items de type réponse courte que leurs homologues suédois qui eux réussissent mieux les items lacunaires (voir tableau 1).

2.4.3. En termes de discrimination

- 30 Il existe aussi des items montrant un fonctionnement différentiel de l'item en termes de discrimination. Les paramètres de discrimination définissent dans quelle mesure une réponse à l'item est due à la compétence et ils sont utilisés pour donner un score aux réponses. Donc cela veut dire que les mêmes réponses à un même item ont été notées différemment. Par conséquent, les scores obtenus pour un bloc d'items ne sont pas immédiatement comparables.
- 31 Finalement, ce qui ressort de ce projet, c'est que le noyau commun n'existe peut-être pas : on obtient des items qui ont des paramètres propres à chaque pays.
- 32 Ceci revient-il à dire que la communication interculturelle dans le domaine de l'évaluation reste nécessairement du domaine de la juxtaposition des pratiques et des références ? On peut se référer ici à Porcher et Zarate pour qui :
- L'obstacle majeur à la compréhension culturelle et la communication interculturelle est l'organisation structurée que tout membre d'une communauté intériorise et incorpore et qui lui permet de se situer et d'agir à l'intérieur des différents champs sociaux dans lesquels il opère (Porcher & Zarate, 1986 in Coste, 1994 : 129).
- 33 Les auteurs concluant à un "habitus" bourdieusien, "une véritable grammaire des comportements culturels" (*ibid.*). Ici, il s'agirait d'un "habitus", d'"une véritable grammaire des comportements évaluatifs".

3. Doit-on absolument chercher à éradiquer ces différences ?

3.1. Exemple de recherche de consensus

- 34 Ces projets européens mettent certes à jour des différences de conceptions et de pratiques dans le domaine de l'évaluation selon les pays. Dans le projet de la "Dutch CEFR grid", lors de l'estimation empirique du niveau minimum de compétence requis pour réussir tel item, le consensus était obtenu plus facilement de la part des experts de tradition pragmatique, se référant à un type d'idéologies conçues comme des "phénomènes ou pratiques matérielles" (Blommaert 2004 : 161) que de la part des experts de tradition textuelle, se référant aux idéologies en tant que "phénomènes idéationnels, d'abord cognitifs" (*Ibid.*). Les certitudes des uns et des autres étant remises en cause et avec elles une part de leur identité.
- 35 Ces experts étant au nombre de cinq (un anglais, un finlandais, un néerlandais, un allemand, une espagnole catalane et une française), il était facile de rechercher l'accord de la majorité, mais telle ne fut pas la démarche adoptée. L'estimation empirique s'est faite après discussions jusqu'à obtention d'un consensus. Maddalena de Carlo utilise les deux notions *idem* et *ipse* avancées par Paul Ricoeur pour désigner les deux faces de l'identité : *idem* renvoyant à la permanence du soi à travers le temps, *ipse* à la faculté de rencontrer autrui et de s'en trouver modifié (1990 : 14). Ce travail de standardisation des items s'est donc opéré grâce à une modification des identités culturelles en matière d'appréciation du niveau requis par un élève pour accomplir une tâche de compréhension orale ou écrite avec succès. Ainsi, l'avis des « experts » proposé par le module d'entraînement en ligne demeure-t-il forcément relatif aux contingences spécifiques de la situation créée par le projet.

3.2. Le consensus est-il toujours souhaitable ?

- 36 Le fondement même du projet Ebafls est précisément la recherche du consensus relatif et limité à un certain nombre d'items. C'est même la préservation de l'identité culturelle des tests propre à chaque pays qui motive la création d'une banque d'items d'ancrage. Grâce à un pool d'items communs, chaque pays pourra conserver ses items habituels. Pourtant, l'expérience montre que cela ne fonctionne pas. Ne doit-on pas entrer alors dans cette zone interculturelle où naissent et prospèrent les paradoxes ?
- 37 On peut citer en exemple le cas de la France qui, à cause de l'interaction culturelle provoquée par le projet Ebafls, vient de décider d'administrer une partie des nouveaux items du prochain bilan de fin de troisième en anglais, allemand et espagnol dans la langue-cible, même si cela ne sera pas sans incidence sur les résultats.
- 38 Il ne s'agit pas ici de rechercher un consensus, d'universaliser un procédé, mais plutôt d'intégrer des paramètres divers, voire divergents au sein d'une démarche qui se veut moins dogmatique, plus intégrative, au final, plus complexe et plus simple à la fois.

4. Des outils à usage souple : l'exemple de Cef-Estim

4.1. Complexité et simplicité

39 Le *Cecrl* qui entend d'une certaine manière unifier la symbolique du langage, à la fois comme représentation, comme action et comme pouvoir, pour reprendre des termes utilisés par Kramersch (2009), peut se laisser interpréter comme permettant de modéliser la complexité, de la simplifier sans la dénaturer. Pour autant, il ne peut satisfaire à toutes les ambitions qu'on voudrait lui prêter. En fait, chacun sait qu'il est très difficile de lier les niveaux du *Cecrl* avec des tests ou des activités de manière scientifique. Et ceci pour les raisons suivantes :

- Le *Cecrl* n'est pas un outil
- les descripteurs ne sont pas rigoureux
- le terme "activité langagière" est confondu avec celui d'"activité d'apprentissage", comme le souligne Portine, ce qui conduit à "l'aplatissement des pseudo-activités langagières sur les tâches à accomplir" et "au refoulement de la composante cognitive" [...] (2008 : 251-2).

Faut-il pour autant rejeter complètement le *Cecrl* et les descriptions qu'il propose ?

4.2. L'exemple du projet Cef-Estim

40 Ce projet (2008-9) fait partie du programme 2008-2011 du Celv (Centre Européen des Langues Vivantes) *Empowering Language Professionals*. Il se donne pour objet de disséminer la grille hollandaise (*Dutch Cefr grid*) parmi les professeurs de langue en Europe. La grille hollandaise, conçue pour les fabricants de tests et d'examens, nécessite une adaptation à un univers plus large, moins consensuel, plus complexe qui est celui des pratiques quotidiennes en Europe.

41 La grille Cef-Estim vise à doter tous les professeurs qui le souhaitent d'un outil pour estimer empiriquement la fourchette de niveau de leurs activités de classe en relation avec les descripteurs du *Cecrl*. La grille est accessible sous deux formats : un format Word pour usage personnel, et un format Access pour la création de banque d'activités référencées de manière souple au sein d'un établissement, d'un groupe d'établissements, voire d'une académie. Cette vision souple du *Cecrl* nous rappelle la double fonction de cet outil européen : à la fois transnational, universalisant et local et particularisant.

42 Cela peut se traduire concrètement par trois exemples :

- la grille Cef-Estim est utilisable partiellement (on n'est pas obligé de tout remplir) ;
- les niveaux ne sont pas conçus comme des points précis mais plutôt comme des fourchettes ;
- les échantillons d'activités qui accompagnent la grille ne sont pas des modèles et préservent la grande diversité des pratiques d'enseignement ;

43 Ainsi le projet Cef-Estim s'est-il donné pour but d'offrir à un public plus large d'enseignants de langues en Europe un outil permettant de cibler plus précisément, mais de manière souple, les activités ordinaires de la classe en relation avec les descripteurs de niveaux du *Cecrl* en termes de textes, activités langagières et compétences. Les échantillons fournis par des professeurs d'anglais de cinq pays européens différents (Arménie, Bulgarie, Espagne, France et Pologne) sont significatifs de la variété d'utilisations possibles. La grille Cef-Estim s'éloigne ainsi grandement de l'injonction

dogmatique qui a pu parfois émaner des partisans du Cefcl pour contribuer, même brièvement, à la formation des professeurs de langues tout au long de la vie.

5. Conclusion

- 44 Au terme de notre étude il apparaît que le Cefcl n'offre pas l'appui nécessaire pour harmoniser les pratiques d'évaluation et de certification en Europe. Mais ceci n'est pas une donnée nouvelle. Si l'on disposait d'un véritable outil pour lier scientifiquement les tests et les niveaux de compétences en langue, il serait effectivement possible de créer une banque d'items d'ancrage et de gommer, d'une certaine manière, l'impact des différences de culture d'évaluation selon les pays.
- 45 Mais cette inadéquation du Cefcl aux ambitions qu'il suscite n'est-elle pas au final la garantie contre une forme non souhaitée d'uniformisation culturelle et ne favorise-t-elle pas des démarches interculturelles plus riches ? Pour prendre véritablement sens, les nomenclatures des situations langagières doivent nécessairement s'incarner dans des contextes locaux, des situations d'apprentissage ou d'évaluation spécifiques et demeurer ainsi interprétables à l'infini.

BIBLIOGRAPHIE

- Alderson, C. *et al.* (2006). "Analysing Tests of Reading and Listening in Relation to the Common European Framework of Reference : The experience of the Dutch Grid Construct Project", in *Language Assessment Quarterly*, Volume 3, number 1.
- Blommaert, J. (2004). *Discourse*. Cambridge : Cambridge University Press.
- Building a European Bank of Anchor items for assessing Language Skills (EBAFLS) (2004-2007) https://www.up2europe.eu/european/projects/ebafls-building-a-european-bank-of-anchor-items-for-foreign-language-skills_122469.html
- Cadre Européen Commun de Référence pour les Langues : apprendre, enseigner, évaluer* (2001). Conseil de l'Europe. Paris : Didier.
- CEF-ESTIM grid project (2008-9). <http://cefestim.ecml.at/>
- Coste, D. (1994). "Dépendant de la culture et non-dépendant de la culture. Stéréotypes et prototypes ". In D. Coste (dir.) *Vingt ans dans l'évolution de la didactique des langues (1968-1988)*, Langues et apprentissage des langues. Paris : Credif, Hatier/Didier, pp. 117-127.
- De Carlo, M. (1998). *L'Interculturel*. Paris : Clé international.
- Demorgon, J. (2004). *Complexités des cultures et de l'interculturel*. Paris : Editions Economica-Anthropos.
- Dutch CEFR grid project (2003-4). <http://www.lancs.ac.uk/fss/projects/grid/>
- Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford : Oxford University Press.

- Eurydice (2002). *Foreign language teaching in schools in Europe*. Brussels : Eurydice.
- SurveyLang (2008). *Inception report for the European Survey on Language Competences*. CITO.
- Kramsch, C. (2002). *Language Acquisition and Language Socialization. Ecological Perspectives*. London : Continuum.
- Kramsch, C. (2009). *The symbolic dimensions of the Intercultural* (à paraître)
- Portine, H. (2008). « Activités langagières, énonciation et cognition. La centration sur les apprentissages ». *Cahiers de l'ACEDLE*, vol. 5. p. 233-254. Disponible en ligne : <https://acedle.org/old/spip.php?rubrique56>
- Ricoeur, P. (1990). *Soi même comme un autre*. Paris : Seuil.
- Tardieu, C. (2006). "The Dutch Grid, a Practical Tool for Evaluating CEFR Levels", in *New Standpoints*, May-June. pp. 3-4.
- van de Vijver, F.J.R, et Hambleton, F.K. (1996). Translating Tests : Some practical Guidelines. *European Psychologist*, vol. 1. n° 2. pp. 89-99.
- Zarate, G., D. Lévy, C. Kramsch (dir). 2008. *Précis du plurilinguisme et du pluriculturalisme*. Paris : Editions des archives contemporaines.

RÉSUMÉS

Le Cefl (Cadre Européen Commun de Référence pour les Langues) est souvent considéré comme un outil idéal de comparabilité des résultats de l'évaluation dans différents pays du Conseil de l'Europe. Cet article traitera des conclusions d'un projet européen montrant qu'il n'en est peut-être rien. Le projet Ebafls (Building a European Bank of Anchor Items for Foreign Language Skills) soulevait la question suivante : "Votre B1 est-il mon B1 ?" La réponse à cette question pourrait consister à dire que les différences culturelles affectent aussi le champ de l'évaluation.

The CEFR (Common European Framework of Reference for Languages) has often been considered an ideal tool to compare the results of assessment in different countries of the Council of Europe. This article will deal with the outcomes of a European project showing that this may not be the case. The Ebafls project (Building a European Bank of Anchor Items for Foreign Language Skills) raised the following question: "Is your B1 my B1?" The answer to this question might be that cultural differences also affect the field of assessment.

INDEX

Mots-clés : CECRL, Ebafls, évaluation, B1, différence culturelle, Cef-Estim

Keywords : CEFR, Ebafls, assessment, B1, cultural difference, Cef-Estim

AUTEUR

CLAIRE TARDIEU

Claire Tardieu est professeur des universités à l'université Sorbonne Nouvelle Paris 3 (EA 4398 PRISMES – EDEAGE), Elle enseigne la didactique de l'anglais en Master Recherche et dans le parcours « Métiers de l'enseignement » et a rédigé plusieurs articles et ouvrages sur la didactique

de l'anglais. Elle participe à des travaux ministériels dans le domaine de l'évaluation et, à ce titre, a pris part à deux projets internationaux : Dutch CEFR Grid et Ebafls. Elle coordonne actuellement le projet Cef-Estim au Centre Européen des Langues Vivantes de Graz.

Courriel : claire.tardieu[at]univ-paris3.fr

Adresse : Université Sorbonne Nouvelle – Paris 3, 5 rue de l'école de médecine, 75006 Paris, France.