



HAL
open science

“ Ce qui fait enjeu ” dans les évaluations linguistiques : Définitions, circulations, implications

Emmanuelle Huver

► To cite this version:

Emmanuelle Huver. “ Ce qui fait enjeu ” dans les évaluations linguistiques : Définitions, circulations, implications. Suzuki, E., Potolia, A., Cambrone-Lasne, S. (dir.). Penser la didactique du plurilinguisme et ses mutations : idéologies, politiques, dispositifs, Presses universitaires de Rennes, 2018, 978-2-7535-7553-0. hal-01817787

HAL Id: hal-01817787

<https://hal.science/hal-01817787>

Submitted on 18 Jun 2018

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

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

« Ce qui fait enjeu » dans les évaluations linguistiques : Définitions, circulations, implications

Emmanuelle HUVER

Université François Rabelais de Tours, France

Abstract

This article is a reflection on the concept of high-stakes testing and its corollary, the standardization/technification of assessment in the field of languages teaching/learning. I will both question the notion and problematize the origins and implications of its introduction into this field. To do this, I will examine the different definitions of the concept that are proposed as well as its circulation between different linguistic and cultural traditions. Finally, I will propose some ideas to conceptualize language teaching evaluation conversely in a qualitative and hermeneutic perspective.

Keywords: assessment, language teaching/learning, high stakes assessment, standardization

Résumé (allemand)

Dieser Artikel behandelt die Thematik des auf English sogenannten „high-stakes testing“ (d.h. sehr anspruchsvolle Evaluation) und insbesondere die Auswirkungen dieses Begriffs im Bereich der Sprachendidaktik. Um diese Problematik zu behandeln, werde ich verschiedene Definitionen von diesem Wort betrachten, sowie seine Geschichte und die Geschichte seiner Verbreitung, in Europa einerseits und in die Sprachendidaktik andererseits. Dies wird mich dazu führen, eine andere Konzeption der Evaluation vorzuschlagen, die sich auf hermeneutische Grundzüge bezieht.

Schlagwörter: evaluation, sprachendidaktik, high stakes testing, standardisierung, hermeneutik

.....

Introduction

Dans le domaine de l'évaluation des langues, la distinction entre évaluations « à enjeux élevés » (*high stakes*) et évaluation « à enjeux faibles » (*low stakes*) occupe actuellement une place prévalente, dans la mesure où elle accompagne et justifie une forte technicisation des certifications de langues. Si la notion existe de longue date dans la littérature anglophone s'intéressant au *language testing*, il n'en va pas de même dans la recherche francophone, où le terme est d'un emploi beaucoup plus récent. Par ailleurs, comme j'avais commencé à l'évoquer dans une publication précédente (Huver et Springer, 2011 : 282), cette dichotomie n'est jamais interrogée, notamment dans la littérature francophone, comme si elle était à la fois évidente et absolue.

Il m'a donc semblé intéressant de s'interroger plus en profondeur sur cette notion, d'une part pour la « désévidencier » (c'est-à-dire questionner l'idée que certaines évaluations seraient, en soi, à enjeux élevés) et, d'autre part, pour problématiser les origines et les implications de son introduction, non seulement pour les évaluations de langues elles-mêmes, mais aussi pour la didactologie et la didactique des langues (désormais DDL).

1. Des enjeux élevés : une évidence ?

1.1. Du risque à la prévisibilité

Le terme *enjeu* provient, étymologiquement, du champ sémantique du jeu, et signifie dans ce cadre « ce que l'on risque dans un jeu et qui doit, à la fin de la partie, revenir au gagnant »¹. Il s'étend au XVIII^e siècle au domaine économique, où *retirer son enjeu* signifie *se retirer d'une affaire*. Au XX^e siècle, il se détache complètement du champ initial du jeu, puisqu' « enjeu » désigne « ce que l'on peut gagner ou perdre dans une entreprise quelconque »².

Par ailleurs, dans le cadre de l'évaluation des politiques publiques, le terme désigne « les conséquences *prévisibles* pour l'intervention (...) afin de *dissiper les malentendus*, de *minimiser le risque* de déception quant aux résultats de l'opération (...) et d'*anticiper* les difficultés au cours de l'exercice »³.

Ainsi, au sème du risque font explicitement place ceux, inverses, de la prévisibilité, de la transparence (« dissiper les malentendus ») et du contrôle (« minimiser le risque »). En outre, si cette évolution peut être expliquée par des évolutions sociétales et économiques, (néolibéralisme, société de la transparence, du risque zéro, du contrôle, etc.), elle révèle également un glissement épistémologique fondamental : lorsque la notion de risque renvoie à une imprévisibilité/altérité irréductible du sens (*vs* un sens conçu comme transparent, au moins à l'horizon), elle suppose de fait une conception et une théorisation bien différente – *herméneutique* – de l'évaluation (pour plus de détails, *cf.* Huver, 2014).

1.2. Durcissement des catégorisations

En DDL, Madaus définit la notion d'enjeu dès 1988, de la façon suivante :

« Tests whose results are seen – rightly or wrongly – by students, teachers, administrators, parents or the general public, as being used to make important decisions that immediately and directly affect them » (Madaus, 1988 : 87).

D'autres définitions ont, par la suite, été proposées, qui ont toutes pour point commun de déplacer la distinction entre enjeux « faibles » *vs* « élevés » des *formes* de l'évaluation (i.e. les types d'activité proposés) à ses *conséquences* et ses *fonctions* (même si par ailleurs ces conséquences et ces fonctions peuvent, elles, varier selon les définitions).

[High stakes testing :] « Testing that has serious consequences for the individual taking the test. (...) The term “high stakes test” also includes most of the standardized national tests, such as the SAT, the Graduate Record Exam, the MCATs, any test that has a consequence regarding admissions, graduation, or promotion »⁴.

¹ CNRTL. URL : <http://www.cnrtl.fr/lexicographie/enjeu>, consulté le 10 janvier 2016.

² URL : <http://www.larousse.fr/dictionnaires/francais/enjeu/29621>, consulté le 10 janvier 2016.

³ URL : <http://www.inpes.sante.fr/slh/articles/390/02.htm>, consulté le 10 janvier 2016. C'est nous qui soulignons dans les citations.

⁴ URL : <http://www.nclcd.org/students-disabilities/homework-study-skills/trouble-high-stakes-testing>, consulté le 10 janvier 2016.

L'idée que cette ligne de partage existe en soi n'est, cependant, jamais discutée ni remise en question. Or, la définition proposée par Madaus repose, quant à elle (au moins partiellement⁵), sur un autre postulat, puisqu'il définit la notion d'enjeu comme un phénomène perçu (« seen as ») indépendamment d'une quelconque « vérité » (« rightly or wrongly »). En d'autres termes, il n'existerait pas, « objectivement », d'enjeux élevés ou d'enjeux faibles, indépendamment du point de vue (i.e. de l'interprétation) des individus. Si l'on va au bout de cette logique, il n'est dès lors plus possible de tracer une frontière fixe entre les deux types d'évaluation en s'appuyant sur des critères externes et *a priori* (que ceux-ci soient des critères de forme, de conséquence ou de fonction).

On assiste donc, d'une certaine manière, à une double rigidification :

- a. Rigidification dans l'acception du terme, qui passe du risque (i.e. du non complètement contrôlable), au contrôle du risque, ce qui permet de justifier le nouage alors opéré avec des formes d'évaluation très technicisées, à des fins de crédibilité (1) et d'administration de la preuve (2) :
 - (1) « Pour être crédibles, les évaluateurs, notamment s'ils interviennent de l'extérieur, ont besoin de présenter des méthodes 'techniquement' assurées et font appel à une démarche scientifique, également réclamée par les donneurs d'ordres »⁶.
 - (2) « L'objectif des tests normalisés est [de] fournir des évaluations justes, valides et fiables qui donnent des résultats probants. S'ils sont réalisés avec attention et un haut niveau d'assurance qualité, les tests normalisés peuvent éliminer les partis pris et éviter les avantages injustes en testant des informations identiques ou similaires dans les mêmes conditions de test »⁷.
- b. Rigidification de ses frontières – entre ce qui relèverait d'un enjeu élevé et ce qui n'en relèverait pas. La conséquence en est une focalisation de la recherche sur le développement d'outils (et de « bonnes pratiques »), indépendamment des interprétations (i.e. des réceptions) de ces outils.

2. Des émergences en décalé pour des projets contrastés

L'usage et la traduction de *high stakes testing* révèlent en outre des circulations terminologiques et notionnelles qui restent peu interrogées, alors que leur compréhension est fondamentale pour comprendre les dynamiques qui traversent la DDL (Zarate et Liddicoat, 2009).

2.1. Aux États-Unis

La notion de *high stakes testing* s'inscrit, historiquement, dans les réflexions états-uniennes en matière de politique éducative. Les objectifs en sont les suivants (considérés comme indissociables) :

- Améliorer les apprentissages des enfants issus de familles défavorisées (essentiellement scolarisés dans des *public schools*) ;

⁵ Porte par ailleurs assez vite refermée (si tant est qu'elle ait jamais été sérieusement ouverte), tant du point de vue des thématiques abordées (les outils, massivement) que de celui des démarches et des orientations de recherche adoptées, qui restent fondées sur l'idée d'une univocité du sens (cf. HUVER, 2014c, partie 2.4.).

⁶ URL : <http://bbf.enssib.fr/consulter/bbf-2005-05-0063-001>, consulté le 10 janvier 2016.

⁷ URL : http://www.ets.org/fr/understanding_testing/purpose, consulté le 10 janvier 2016.

- Proposer un système d'évaluation plus économique ;
- Inscire l'éducation dans un programme global de défense nationale.

C'est ainsi que naît le *National Defense Education Act* (NDEA) en 1958, qui accorde une place importante aux tests standardisés pour atteindre ce triple objectif. Le dispositif est généralisé dans les années 1970 via le « SAT test score » puis étendu en 1983, suite à la publication du rapport *A nation at risk*, commandé par l'administration Reagan (National Commission on Excellence in Education, 1983). Ce rapport prolonge en effet l'idée du lien éducation/défense nationale introduite par le NDEA en le situant dans le contexte de compétition globale de l'économie mondiale. C'est dans ce cadre précis que la notion de *high stakes test* est introduite, pour réaffirmer l'importance des tests standardisés dans une politique éducative qui se veut non seulement méritocratique mais aussi concurrentielle⁸ (Johnson et Johnson, 2009 ; Watzke, 2003).

En DDL, si Madaus conceptualise la notion dans les années 80, celle-ci tire ses origines d'une réflexion bien plus ancienne. Spolsky (1995), dans son histoire des tests de langue aux États-Unis, situe ainsi le début de la réflexion sur ce sujet dans les années 1920. Il montre de manière fine (sur plusieurs chapitres) la forte influence exercée par la psychologie et plus particulièrement par la psychométrie, ce qui se manifeste notamment par l'importation de méthodologies issues de l'analyse statistique (travaux de Spearman) et d'outils issus des recherches sur les tests mentaux (Galton, Binet, etc.). Cette notion est, depuis, très présente et largement conceptualisée par le biais des nombreux travaux sur le « washback effect » (i.e. « l'effet retour » des évaluations) et la validité (cf. par exemple : Messick, 1996 ; Cheng, Watanabe et Curtis (dir.), 2004 ; Solórzano, 2008, etc.).

2.2. En France

En France, le terme *enjeu fort* n'a fait son apparition que plus récemment (années 1990 / 2000), tant dans le domaine des politiques éducatives (introduction des tests standardisés internationaux de type PISA) que dans celui de la DDL (publication du *Cadre européen commun de référence pour les langues* (désormais CECRL)).

Par ailleurs, la notion n'est pas réellement définie, si ce n'est en reprenant les travaux anglophones. Tout porte donc à croire que le terme y est utilisé dans son acception commune ou en le considérant comme une « simple » traduction du terme anglais. Or, en la matière, les circulations terminologiques ne sont jamais de « simples » traductions⁹ et il importe donc d'y revenir plus précisément.

La réflexion de la place de l'évaluation dans les politiques éducatives date des travaux en docimologie et tient notamment à l'engagement d'H. Piéron. Les réflexions docimologiques tirent en effet leurs racines d'une réflexion sur la réforme du système scolaire français et sont généalogiquement liées aux mesures mises en œuvre par Napoléon I^{er} (les concours, notamment) pour permettre à la bourgeoisie (vs l'aristocratie) d'accéder à certains postes

⁸ Ce qui est ensuite prolongé et renforcé en 2002 sous l'administration G.-W. Bush par le programme *No child left behind*, qui généralise les tests et sanctionne les écoles n'ayant pas atteint les objectifs visés (celles-ci doivent par exemple aider les élèves qui le souhaitent à changer d'école et payer les frais de transports liés à leur nouveau lieu de scolarisation).

⁹ Cf. par exemple VERONIQUE (2009) sur la circulation du terme *applied linguistics*, devenu *linguistique appliquée* en français, alors que les deux termes dessinent des territoires, des orientations et des traditions disciplinaires très contrastés.

d'État¹⁰. Par ailleurs, la réflexion d'H. Piéron s'inscrit très explicitement dans une orientation politique socialiste (il participe activement au plan Langevin-Wallon). Ainsi, même si ses préconisations peuvent, dans leur forme standardisée et objectivée, être rapprochées des évaluations standardisées états-uniennes, elles ne s'inscrivent pas du tout dans le même projet politique. Par ailleurs, pour des raisons qu'il n'est pas possible de développer ici (poids des concours, tradition de la dissertation, etc.) et malgré les préconisations de Piéron, la standardisation des évaluations (et, avec elle, la notion d'enjeux forts) est restée en France très minoritaire – voire inexistante – dans le système éducatif¹¹ et ce, jusque très récemment.

Dans le domaine de la DDL, les quelques travaux existant dans les années 1980 et au début des années 1990 sur l'évaluation sont essentiellement d'ordre méthodologique, puisqu'ils visent à adapter les évaluations au tournant communicatif (Lussier, 1992 par exemple). L'élaboration du DELF répond à ce même objectif méthodologique, ainsi qu'à des visées d'institutionnalisation et donc de légitimation du domaine du français langue étrangère en tant que discipline (Porcher, 1986). La standardisation des évaluations ne semble ainsi pas être une préoccupation première et fait même l'objet de critiques, en partie d'ailleurs issues des réflexions anglophones sur ce sujet (Bolton, 1991).

Le virage vers la standardisation des évaluations date d'une quinzaine à une vingtaine d'années, soit bien plus tardivement que dans l'aire anglophone. La question des « enjeux élevés » y joue un rôle clé, via le *CECRL*, puisqu'elle est en partie à l'origine de l'élaboration de ce document (c'était en tout cas la principale demande des organismes certificateurs : Coste, 2013) et que, par ailleurs, la manière dont celui-ci a ensuite été investi a largement contribué à la diffusion de la notion (Huver, 2014b).

Dans ces dynamiques de circulation d'une aire linguistico-scientifique à une autre, il est cependant étonnant que les critiques (pourtant fortes : Shohamy, 2001 ; Johnson *et al.*, 2008) n'aient pas circulé avec la même vigueur ou que, lorsqu'elles existent, elles portent essentiellement sur le lien de cause à effet entre néo-libéralisme et omniprésence de l'évaluation.

Ainsi, si l'introduction du terme *évaluation à enjeux élevés* est plus récente en France et témoigne essentiellement d'une importation non interrogée (i.e. coupée des débats de société et des projets socio-politiques qui en sont à l'origine), on peut se demander ce que permet son soudain succès, alors que les traditions locales (concernant les conceptions de l'école, de l'enseignement, de l'évaluation, mais aussi de la DDL) semblaient a priori assez peu y prédisposer.

3. Circulations : quelques hypothèses

3.1. Le néo-libéralisme à l'œuvre ?

La diffusion de la notion d'*évaluation à enjeux élevés* tient largement aux actions d'évaluation des politiques publiques mises en œuvre depuis une bonne dizaine d'année, si bien que l'appareillage technique qui l'accompagne est fréquemment dénoncé comme levier de

¹⁰ Cette généalogie se tisse au travers des travaux du professeur Laugier dans les années 30, qui en mettant en évidence les disparités dans les modalités de notation dans les épreuves de concours ouvre la porte aux réflexions docimologiques développées ensuite par H. Piéron.

¹¹ La recherche d'objectivité ayant pris d'autres voies, notamment celle de la double correction et du recours à des jurys, par exemple.

l'introduction de la pensée néo-libérale dans toutes les sphères de la société (cf. Huver, 2014c). Si ce constat a certes une indéniable validité, le postulat sur lequel il se fonde doit cependant être relativisé. Estimer que l'évaluation – notamment dans sa version hautement technicisée (grilles, standards, etc.) – est une conséquence et un vecteur du néo-libéralisme revient en effet à considérer que :

- L'usage de l'évaluation à des fins politiques serait un phénomène récent, voire inédit. Or, Chardenet (1999 : 61) situe aux années 60 l'entrée du mot *évaluation* dans les discours éducatifs, notamment dans le domaine de la didactique des mathématiques. Les réflexions de Piéron sur le sujet datent de la même période ;
- L'évaluation ne pourrait être mise qu'au service d'un projet socio-économique néolibéral. Or, Piéron inscrit sa réflexion sur la docimologie dans un projet de politique éducative explicitement socialiste (cf. *supra*)¹².

Par conséquent, le lien de causalité fréquemment établi entre évaluation et néo-libéralisme, s'il a ses zones de pertinence, suppose également d'être interrogé et approfondi, au regard d'autres interprétations.

3.2. Influence des organismes certificateurs : un conflit d'intérêts ?

La circulation d'évaluations à enjeux élevés peut également être considérée, dans le domaine de la DDL plus particulièrement, comme liée à l'influence grandissante des organismes développeurs de tests et des associations qui les regroupent (CIEP, ETS, Eurocentres, ALTE, etc.). Ceux-ci ont très largement influencé la configuration finale du CECRL, puisque c'est pour répondre aux besoins exprimés par eux que les niveaux communs de référence ont finalement occupé une place centrale dans le CECRL lui-même (Coste, 2013), ce qui a eu pour conséquence, par la suite, d'indexer l'ensemble de la DDL sur ces niveaux (Huver, 2014b). Or, ce faisant, ces organismes certificateurs ont également très largement contribué à créer un marché présentant des besoins qu'eux seuls seraient en mesure de satisfaire (i.e. des évaluations fiables, valides, standardisées, étalonnées, etc.). Dans le domaine économique, c'est ce que l'on peut appeler un conflit d'intérêts.

Mais surtout, ce qu'il y a d'intéressant pour la problématique qui m'occupe ici, c'est que la manière dont y est traitée la notion d'enjeu élevé y fait office de pivot argumentatif décisif. Prenons l'exemple – classique – des tests de langue utilisés dans le cadre des politiques migratoires¹³ : ils sont catégorisés comme « à enjeux forts » puisque du niveau évalué (lui-même indexé sur les niveaux communs de référence) dépend l'obtention (ou non) de l'autorisation de résidence ou de la nationalité. Dans le raisonnement évoqué ci-dessus, les évaluations *doivent* être objectives *du fait de ces enjeux* et leur conception doit *donc* être technicisée *pour garantir* cette objectivité. Or, seuls les organismes certificateurs ont les moyens logistiques, matériels et techniques d'élaborer ce type de test. Par conséquent, c'est bien l'argument des enjeux forts qui crée une demande de certification tout en permettant aux

¹² Cf. également l'histoire des tests en URSS, qui met en évidence le même phénomène d'usage de tests standardisés à des fins d'égalitarisme et, *in fine*, d'émancipation de la classe ouvrière, les autres formes d'évaluation étant considérées comme des moyens de perpétuer la domination de la classe bourgeoise (HUTEAU et LAUTREY, 1997).

¹³ C'est-à-dire où l'évaluation sert à certifier un niveau de langue dans la perspective d'obtenir un permis de séjour ou encore la nationalité.

organismes certificateurs de légitimer – et par ce biais d'imposer – une technicité et, en même temps, de capter, voire de monopoliser, la réponse à cette demande¹⁴.

3.3. Circulation des techniques, effacement des histoires, déresponsabilisation des acteurs

On l'a vu plus haut, le terme d' « *enjeu élevé* » apparaît, dans la littérature anglophone, comme une *notion* de recherche (avec ses débats, ses lignes de fracture, ses controverses, etc.), alors qu'il semble pris, en France – y compris dans la recherche francophone en DDL –, comme un terme usuel et transparent, une évidence qu'il n'est donc pas besoin de problématiser et d'interroger, une « simple traduction ».

Cette migration terminologique non interrogée aboutit à effacer le caractère construit et relatif (*vs* absolu) de la notion (*cf. supra*). Mais aussi, et surtout, elle efface les enjeux et projets politiques et épistémologiques qui ont initialement sous-tendu l'introduction et le développement de la notion ainsi que les mises en œuvre qui en découlent, pour ne retenir, justement, que ces mises en œuvre techniques (standardisation, étalonnage, technicisation, procédures issues des statistiques et de la psychométrie).

Au-delà de la forme technicisée (parce que standardisée) des évaluations, il y a donc là une forme de technicisation moins voyante, plus implicite, mais d'autant plus pernicieuse, qui consiste à dissocier les techniques des histoires, principes, projets épistémologiques et politiques qui les fondent et les animent (voir Castellotti et Nishiyama, 2011 pour un exemple similaire à propos du *CECRL*). Or, ce postulat d'autonomie de la technique est très largement contestable et a d'ailleurs été très largement contesté, notamment en philosophie de la technique (voir Heidegger, 1953 tout particulièrement), de même que celle-ci a également contesté la préséance de la technique sur les usages par les individus (voir par exemple, les notions de dispositif *vs* processus de subjectivation développées par Agamben (2007) à la suite de Foucault).

Si le terme *enjeu élevé* s'est imposé sans discussion ni problématisation dans la recherche sur les évaluations de langues, c'est plus largement parce que la représentation de la technique comme indépendante de dimensions épistémologiques, politiques, etc. (ce que j'ai appelé ailleurs – Huver, 2014c – une conception « techniciste » de la DDL) est largement dominante dans ce domaine (à des degrés divers, qu'il faudrait préciser) et insuffisamment explorée. Or, elle a des effets à différents niveaux, dont l'évaluation (et la question des enjeux élevés entre autres) constitue un indéniable miroir grossissant. Par exemple, le fait de situer l'éthique de l'évaluation dans la technicisation et la standardisation des évaluations (voir par exemple le site d'ETS Global) génère des processus d'exclusion et de déresponsabilisation :

- Exclusion des enseignants de la conception des évaluations à enjeux élevés, au motif, justement, de ces enjeux élevés (Huver et Springer, 2011 : 114) ;
- Déresponsabilisation des acteurs et des instances décisionnaires : dès lors que l'éthique et l'équité sont une affaire technique, la responsabilité du processus d'évaluation jusqu'à la prise de décision, qui en fait constitutivement partie (de Ketele, 2010), est déplacée sur la technique elle-même et non sur les personnes qui la mettent

¹⁴ Ce que MADAUS énonçait d'ailleurs dès 1988 (High-stakes test « *transfers control over the curriculum to the agency which sets or controls the exam* »).

en œuvre¹⁵, comme si l'humain était, au final « gouverné par les instruments » (Lascoumes et Le Galès, 2005) et, en l'occurrence, « gouverné par les nombres » (Desrosières, 2008). Ainsi, si l'on reprend l'exemple de l'utilisation des évaluations standardisées dans le cadre des politiques migratoires : plutôt que l'exactitude des niveaux qui serait à garantir, n'est-ce pas la politique migratoire elle-même et notamment son postulat d'antériorité de la maîtrise de la langue sur l'intégration, qui seraient à questionner ?

4. Perspective : face à la technicisation de la DDL, l'herméneutique

Que l'on ne se méprenne pas. Mon objectif ici n'est pas de m'intéresser à « l'impact » des évaluations standardisées ou à enjeux élevés (Mc Millan, 2005 ; Mons, 2009). Il n'est (surtout) pas non plus de prolonger le débat « pour » ou « contre » l'évaluation ou l'évaluation standardisée : on voit bien les limites de la critique socio-économique (évoquées rapidement *supra*) ainsi que des travaux visant à comparer les évaluations internes *vs* externes (Monseur et Demeuse, 2005) ou encore à enjeux élevés *vs* faibles (Stobart et Eggen (dir.), 2012), qui restent uniquement centrées sur les outils. Mon propos est plutôt de *conceptualiser* l'évaluation « autrement », notamment en tirant toutes les conséquences du fait que toute évaluation, aussi standardisée soit-elle, suppose l'implication et la subjectivation d'un (ou de plusieurs) individu(s) et que la quête d'une évaluation objective (i.e. vidée de l'implication des individus qui la mettent en œuvre) est un non-sens.

En d'autres termes, il s'agit de développer l'idée que toute évaluation est *intrinsèquement qualitative et (donc) herméneutique* (cf. également Pierozak *et al.* 2013, dans une perspective similaire à propos de la sociolinguistique). Si cette réflexion a déjà été partiellement engagée dans les sciences de l'éducation (Arduino et Berger, 1986 ; Vial, 2001), c'est encore loin d'être le cas en DDL. Ainsi, plus qu'une critique socio-économique ou une réflexion sur les techniques, c'est une réflexion épistémologique qu'il faut également développer. Et parce que l'herméneutique pose, fondamentalement, la question de l'histoire et de la réception, cela peut prendre plusieurs voies, que je ne peux qu'évoquer rapidement, à ce stade conclusif de l'article :

- *Voie de l'histoire* : en s'intéressant à l'histoire de l'évaluation des langues, du point de vue des circulations entre différentes aires linguistico-scientifiques (ce que j'ai tenté de faire ici), mais aussi entre différentes disciplines (statistiques – Desrosières, 1993 ; psychométrie – Spolsky, 1995, etc.) ;
- *Voie de la réception* : en s'intéressant, en l'occurrence, non pas aux enjeux eux-mêmes, mais à « *ce qui fait enjeu* » (pour qui ? pour quoi ? quand ? etc.), c'est-à-dire à la réception des enjeux par différents types d'acteurs (enseignants, apprenants, décideurs, etc.) ou, autrement dit, aux modalités d'interprétation et d'appropriation de différents types d'évaluation, au-delà de cette hétéro-catégorisation d'enjeu « fort », à mettre en lien, notamment, avec les histoires, les valeurs, les postures, les projets des acteurs concernés ;
- *Voie de l'épistémologie* : d'une part pour l'évaluation, en s'intéressant de plus près à la problématique de la quantification comme politique de la preuve (Desrosières, 1993). D'autre part, aux différents niveaux de la DDL, en approfondissant les implications d'une conception herméneutique (*vs* pragmatiste) du sens, de l'enseignement, de l'apprentissage... mais aussi de la langue.

¹⁵ Ce que DESROSIERES (1993 : 79) appelle des processus de désingularisation. Cf. également les travaux d'AGAMBEN évoqués *supra*.

Bibliographie

- AGAMBEN Giorgio, 2007, *Qu'est-ce qu'un dispositif ?*, Paris, Payot.
- ARDOINO Jacques et BERGER Guy, 1986, « L'évaluation comme interprétation ». *POUR*, n°107, p. 120-127.
- BOLTON Sybille, 1991, *Evaluation de la compétence communicative en langue étrangère*, Paris, Hatier.
- CASTELLOTTI Véronique et NISHIYAMA Jean (dir.), 2011, « Contextualiser le CECR ? », *Recherches et applications – Le français dans le monde*, n°50, p. 11-18.
- CHARDENET Patrick, 1999, *De l'action évaluative à l'acte d'évaluation. Approche théorique et pratique communicationnelle*. Paris : L'Harmattan.
- CHENG Liying, WATANABE Yoshinori et CURTIS Andy (dir.), 2004, *Washback in language testing: Research contexts and methods*, London, Lawrence Erlbaum.
- COSTE Daniel, 2013, « Petit retour à Rüsclikon », *Recherches et applications – Le français dans le monde*, n°53, p. 140-148.
- DAVIES Alan, BROWN Annie, ELDER Cathie, HILL Kathryn, LUMLEY Tom et MCNAMARA Tim, 1999, *Dictionary of Language Testing*, Cambridge, Cambridge University Press.
- DESROSIERES Alain, 2008, *Gouverner par les nombres : L'Argument statistique II*, Paris, Presses de l'École des Mines.
- DESROSIERES Alain, 1993, *La Politique des grands nombres – Histoire de la raison statistique*, Paris, La Découverte.
- HEIDEGGER Martin, 1953, *La question de la technique*, Paris, Gallimard.
- HUTEAU Michel et LAUTREY Jacques, 1997, *Les tests d'intelligence*, Paris, La Découverte.
- HUVER Emmanuelle, 2014, « Les inaccessibles en évaluation. Impensé ? Impasse ? Ferments ? », in Cécile Goï, Emmanuelle Huver et Elatiana Razafimandimbimanana (dir.), « Inaccessibles, altérités, pluralités : trois notions pour questionner les langues et les cultures en éducation », *Glottopol*, n°23.
URL : http://glottopol.univ-rouen.fr/telecharger/numero_23/gpl23_03huver.pdf
- HUVER Emmanuelle, 2014b, « CECR et évaluation : interprétations plurielles et logiques contradictoires », *Cahiers du GEPE*.
URL : <http://www.cahiersdugepe.fr/index2652.php>
- HUVER Emmanuelle, 2014, *De la subjectivité en évaluation à une didactique des langues diversitaire. Pluralité, altérité, relation, réflexivité*, Note de synthèse d'habilitation à diriger des recherches, Université F. Rabelais (Tours).
- HUVER Emmanuelle et SPRINGER Claude, 2011, *L'évaluation en langues*, Paris, Didier.
- JOHNSON Dale D. et JOHNSON Bonnie, 2009, « High Stakes Testing ».
URL : <http://www.education.com/reference/article/high-stakes-testing1/>
- JOHNSON Dale D., JOHNSON Bonnie, FARENGA Stephen J., et NESS Daniel (2008). *Stop high-stakes testing: An appeal to America's conscience*. Lanham : MD, Rowman et Littlefield.
- KETELE DE Jean-Marie, 2010, « Ne pas se tromper d'évaluation », *Revue française de linguistique appliquée*, vol. XV, p. 25-36.
- KUNNAN Antony-John, 2012, « High-Stakes Language Testing », *The Encyclopedia of Applied Linguistics*.
- LASCOUMES Pierre et LE GALES Patrick. dir., 2005, *Gouverner par les instruments*, Paris, Presses de Sciences Po.
- LUSSIER Denise, 1992, *Evaluer les apprentissages dans une approche communicative*, Paris, Hachette.
- MCMILLAN James H., 2005, *The impact of High - Stakes Test Results on Teacher's Instructional and classroom assessment Practices*, Virginia Commonwealth University, Metropolitan Educational Research Consortium.
URL : <http://www.eric.ed.gov/PDFS/ED490648.pdf>
- MADAUS George F., 1988, « The influence of testing on the curriculum », in Laurel N. Tanner (ed.), *Critical issues in Curriculum, 87th yearbook of the national society for the study of education (Part 1)*, Chicago, University of Chicago Press.
- MESSICK Samuel, 1996, « Validity and washback in language testing », *Language Testing*, 13/4, p. 241-256.

- MONS Nathalie, 2009, *Les effets théoriques et réels de l'évaluation standardisée*.
URL : http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/111FR.pdf
- MONSEUR C. et DEMEUSE Marc, 2005, « Les évaluations externes permettent-elles une régulation efficace ? », in Marc Demeuse et al. (dir.), *Vers une école juste et efficace*, Bruxelles, De Boeck, p. 489-517.
- NATIONAL COMMISSION ON EXCELLENCE IN EDUCATION., 1983, *A nation at risk*.
URL : <http://www2.ed.gov/pubs/NatAtRisk/index.html>
- PIERON Henri, 1963, *Examens et docimologie*, Paris, PUF.
- PIEROZAK Isabelle, DE ROBILLARD Didier, RAZAFIMANDIMBIMANANA Elatiana et DEBONO Marc, 2013, « Vers une sociolinguistique française qualitative ? Perspectives historiques critiques sur des processus de reconnaissance », *Recherches qualitatives*, n°32, p. 107-131.
URL : http://revue.recherche-qualitative.qc.ca/edition_reguliere/numero-32-1/rq-32-1-Pierozak-et-al.pdf
- PORCHER Louis, 1986, « Priorités institutionnelles », *Etudes de linguistique appliquée*, p. 75-92.
- SHOHAMY Elana, 2001, *The power of tests. A critical perspective on the uses of language tests*, Essex, Pearson Education.
- SOLÓRZANO R.-W., 2008, « High-Stakes Testing : Issues, Implications, and Remedies for English Language Learners », *Review of educational research*.
- SPOLSKY Bernard, 1995, *Measured words. The development of objective language testing*, Oxford, Oxford University Press.
- STOBART G. et EGGEN T. (dir.), 2012, « High-stakes testing - value, fairness and consequences », *Assessment in Education: Principles, Policy et Practice*, 19(1), p. 1-6.
- VERONIQUE Georges-Daniel, 2009, « La linguistique appliquée et la didactique des langues et des cultures : une polémique française au cœur d'un débat international », in Geneviève Zarate et Anthony Liddicoat (dir.), *Recherches et applications – Le français dans le monde*, n°46, p. 42-52.
- VIAL Michel, 2001, *Se former pour évaluer, Se donner une problématique et élaborer des concepts*, Bruxelles, De Boeck.
- WATZKE J. L., 2003, *Lasting Change in Foreign Language Education. A Historical Case for Change in national policy*, Westport, Praeger.
- ZARATE Geneviève et LIDDICOAT Anthony (dir.), 2009, « La circulation internationale des idées en DDL », *Recherches et applications – Le français dans le monde*, n°46.