



HAL
open science

Assessment of the language skills: from formative to certificative assessment

Yves Bardière, Cristiana Cervini, Monica Masperi

► **To cite this version:**

Yves Bardière, Cristiana Cervini, Monica Masperi. Assessment of the language skills: from formative to certificative assessment. *Mediazioni. Rivista online du studi interdisciplinari su lingue e culture*, 2021, 32, pp.A1-A11. hal-03590794

HAL Id: hal-03590794

<https://hal.science/hal-03590794>

Submitted on 27 Jun 2022

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.

Évaluation des acquisitions langagières : du formatif au certificatif

Yves Bardière, Université Grenoble Alpes, Cristiana Cervini, Università di Bologna, Monica Masperi, Université Grenoble Alpes

Citation: Bardière, Yves, Cristiana Cervini, Monica Masperi (2021) "Évaluation des acquisitions langagières : du formatif au certificatif", in Monica Masperi, Cristiana Cervini, Yves Bardière (eds.) *Évaluation des acquisitions langagières : du formatif au certificatif*, *mediAzioni* 32: A1-A11, <http://www.mediazioni.sitlec.unibo.it>, ISSN 1974-4382.

Depuis l'émergence, dans les années 50, de l'évaluation et du *testing* en tant que disciplines scientifiques, les courants dominants en didactique des langues-cultures se sont toujours reflétés sur les formes que les tests en langues ont assumées : le *discrete-point testing* (Lado 1961) a été inspiré par le courant structuraliste et les approches comportementalistes en psychologie et en sciences cognitives. Ensuite le *testing* communicatif (Morrow 1979) a mis en avant l'unité textuelle, écrite et orale, la capacité de transmettre un message compréhensible au-delà de son acceptabilité sur le plan formel, et l'évaluation des habiletés en réception et en production. Ces dernières années, la communauté scientifique a mis la focale sur des concepts clés qui suscitent nombre d'interrogations et alimentent des études empiriques, tels que l'authenticité situationnelle et interactionnelle dans le *task-based testing* ou encore l'évaluation à visée formative ou diagnostique, pour ne citer que quelques fondamentaux (Hamp-Lyons 2016).

Aujourd'hui les candidats présentent de plus en plus des profils langagiers plurilingues, sont en mobilité internationale et ont recours systématiquement aux technologies numériques, et cela quel que soit le contexte formatif et évaluatif (école, université, centre de certification, centre de langues, etc.). En parallèle,

les tests et certifications constituent un réel enjeu pour les candidats et les institutions et font émerger la nécessité de concevoir et de déployer des dispositifs valides, fiables et équitables.

Le colloque international EVAL19, qui s'est tenu en octobre 2019 à l'Université Grenoble Alpes avec le concours de deux autres Institutions (Université de Bologne et Université de Nice), a abordé des questions qui interpellent actuellement la recherche sur l'évaluation en langues, à travers des réflexions théoriques et des retours sur expériences portant sur la validité des dispositifs, l'évaluation des approches intercompréhensives et plurilingues ou encore l'évaluation des interactions langagières. Les quatorze contributions issues de ce colloque et réunies dans ce volume tentent d'apporter un éclairage sur les rapports entre les différents types d'évaluation (formative, sommative, certificative), leur rôle dans le processus d'apprentissage/enseignement des langues et leur impact potentiel sur les acteurs de terrain (candidats, enseignants, évaluateurs, institutions).

Cet ouvrage rend compte de recherches menées sur plusieurs langues, dans différents contextes de formation : si les langues indo-européennes, telles que l'**anglais**, l'**italien** et l'**allemand**, fournissent l'essentiel du corpus d'exemples proposés, les langues dites MoDiMes (Moins Diffusées et Moins Enseignées) comme le **japonais** et le **chinois** servent également d'illustration et reflètent la **volonté plurilingue** qui anime l'ensemble de ce recueil. Notons également que les contributions s'inspirent de nombreuses théories de référence et approches méthodologiques, comme la psychométrie, la linguistique énonciative ou la mimésis, pour n'en citer que quelques-unes. N'oublions pas enfin qu'elles s'inscrivent toutes dans une perspective plus évaluative que formative, comme le veut l'intitulé même du colloque dont est issue cette publication.

La définition de l'évaluation, telle que la propose **Tim McNamara** lors de sa conférence plénière¹, permet de brosser le cadre général dans lequel s'inscrivent les différents articles répertoriés dans ce numéro : « Assessment is a process of *observing* evidence from performance, *interpreting* it, so that we can *reach conclusions* about what individuals know or can do and in order to *make decisions* based on these conclusions » (souligné par l'auteur).

Cette définition de l'évaluation s'applique aussi bien à la conception des contenus des tests qu'aux choix à opérer pour respecter les principes de validité, de fiabilité et d'équité dans le *testing*, ou encore à l'identification de profils d'apprenants.

Les quatre actions qui, selon Tim McNamara, sous-tendent tout processus d'évaluation, sont régies par des relations causales : toute *prise de décision* repose sur des *conclusions*, elles-mêmes subordonnées à l'*interprétation* d'éléments que livre l'*observation*. Notre conférencier invité consacre l'essentiel de son intervention au dernier constituant de cette chaîne de causation ("*make decisions*"), et plus précisément aux principes de **justice** et d'**équité**, qui doivent prévaloir dans *toute prise de décision* et s'établir au bénéfice du candidat. Tim McNamara introduit une distinction essentielle entre *fairness*, critère interne ayant trait à la qualité intrinsèque du test (le test livre-t-il des informations justes, fiables, valides sur les compétences de l'individu ?) et *justice*, critère externe (quelle utilisation sociale et politique est faite du test ?).

C'est dans cet esprit d'équité que se situe l'analyse d'**Anna Cardinaletti**, qui cible sa recherche sur les candidats en situation de handicap en Italie, en particulier les **étudiants atteints de troubles sensoriels (sourds) et d'apprentissage (TA)**. L'admission aux cours universitaires en Italie étant soumise à la réussite d'un test d'anglais de niveau B1, autocorrectif et en ligne, et d'une épreuve en langue italienne, l'auteure s'interroge sur l'**équité** des tâches proposées dans les tests : certaines typologies d'exercice sont-elles à même d'évaluer les compétences des étudiants en situation de handicap de manière équitable ? Les

¹ Podcast accessible à cette adresse :

<https://eval19.sciencesconf.org/resource/page/id/8/#mynamara>

Ce volume propose une transcription de la conférence de Tim McNamara, assortie de quelques modifications mineures destinées à en faciliter la lecture.

résultats obtenus à ces deux tests sont-ils attribuables au format des tests ou à la situation de handicap ? Les conclusions de l'auteure font apparaître la nécessité de modifier le contenu et le format des tests linguistiques en tenant compte de leur accessibilité. Elles montrent que les tests en L1 devraient être modelés sur ceux de L2 pour garantir une évaluation de toutes les compétences linguistiques et donner une image aussi complète que possible de chacune d'entre elles.

La contribution de **James Purpura**, autre conférencier invité à EVAL19, permet de compléter le cadre général dans lequel s'inscrivent les articles ou série d'articles répertoriés ci-dessous, plus particulièrement ceux consacrés à la certification CLES. Comme le stipule sans ambiguïté le titre, *A Rationale for Using a Scenario-Based Assessment to Measure Competency-Based, Situated Second and Foreign Language Proficiency*, l'analyse de James Purpura porte sur l'évaluation scénarisée (*scenario-based assessment or SBA*), dont se réclame ouvertement le CLES. L'article adopte pour une bonne part une perspective historique, qui retrace les différentes conceptualisations de la compétence en langue seconde/étrangère depuis les années 1960 jusqu'à nos jours. Les notions de contexte, d'authenticité, d'intégration des compétences, d'intentionnalité, de médiation, de co-construction et d'activités collaboratives pour résoudre un problème, incontournables dans une évaluation de type SBA, sont passées au crible de l'analyse. L'auteur plaide en faveur d'une approche orientée vers l'apprentissage (*a learning-oriented approach to assessment or LOA*) dans la conception d'une évaluation de type SBA. Il recommande notamment la prise en compte du contexte (*contextual dimension*), de la culture (*cultural dimension*) et de la compétence (*proficiency dimension*). Des définitions très éclairantes de la compétence sont également proposées.

Plusieurs articles sont ensuite dévolus au Système d'Évaluation en Langues à visée Formative (SELF). **Triscia Biagiotti**, **Cristiana Cervini** et **Monica Masperi** proposent une analyse à fois quantitative et qualitative de la version italienne de SELF en se fondant sur une expérience de validation à large échelle menée sur un échantillon de 1499 candidats. Elles commencent par rappeler, à l'instar de Tim McNamara, que la **compréhension de l'oral** et plus globalement la

compétence communicative ne sont pas directement mesurables et mettent de ce fait en avant une approche psychométrique fondée sur la théorie classique des tests et la théorie de la réponse à l'item pour étudier le comportement des items qui testent cette habileté. Ce sont donc les trois premiers éléments qui rentrent en ligne de compte dans la définition de *assessment* (cf. *supra*) qui sont plus particulièrement sollicités mais aussi le quatrième, SELF ayant également pour vocation de déterminer des groupes de niveaux en orientant l'apprenant vers le niveau à acquérir.

L'évaluation de la **compétence lexicale** occupe une place centrale dans l'article de **Rui Yan, Yijie Zhang et Mariarosaria Gianninoto**. En raison des spécificités du système d'écriture chinois (écriture non alphabétique) et de l'absence de racines lexicales communes avec les langues européennes, l'appréciation de la compétence lexicale est en effet essentielle dans l'évaluation en **chinois** langue étrangère (CLE). Après avoir présenté les référentiels CLE, les auteures entreprennent de définir cette compétence et discutent des choix méthodologiques opérés. Des exemples concrets illustrent les orientations didactiques adoptées et mises en œuvre.

Dans leur enquête, **Tomoko Higashi et Chieko Shiota** s'intéressent, quant à elles, au **fonctionnement cognitif** des candidats qui réalisent les **courtes tâches d'écriture** proposées par le test SELF **japonais**. Ces dernières consistent à sélectionner un ou plusieurs mots pour compléter des énoncés dans un texte. Par le biais d'un protocole de verbalisation à voix haute (*think-aloud protocol*), ces auteures passent au crible l'activité de six étudiants en japonais de niveau hétérogène (A1 à B1) en situation d'exécution de la tâche. Leurs analyses portent sur les processus et les stratégies mobilisés pour lire et comprendre le texte écrit et sur l'activation des connaissances **grammaticales** et **lexicales** nécessaires pour mener à bien la tâche.

Se fonder sur une approche interdisciplinaire pour affiner les outils d'analyse des données que livre SELF a constitué le défi qu'une équipe de chercheurs et enseignants en didactique des langues et en statistiques de l'Université Grenoble Alpes a tenté de relever. Ainsi **Vincent Brault, Frédérique Letué, Marie-José**

Martinez, Sylvain Coulange, Marie-Pierre Jouannaud et Anne-Cécile Perret ont-ils entrepris d'aller au-delà des niveaux par habiletés que SELF propose, pour tenter d'établir le **profil de compétences** des étudiants en fonction de leurs schémas de réponses, en croisant la classification de SELF avec des algorithmes dits de "*co-clustering*", définis pour organiser et regrouper les données collectées selon des caractéristiques communes. Pour mener à bien cette étude, cette équipe interdisciplinaire a développé un logiciel expérimental qui ouvre de nouvelles perspectives de recherche sur l'évaluation formative en langues.

Viennent ensuite toute une série d'articles sur le **Certificat de compétences en langues de l'Enseignement supérieur (CLES)**. Ces articles se penchent plus spécifiquement sur l'interaction orale au niveau B2. **Catherine Felce** rappelle que le caractère interactif et co-construit des échanges rend l'évaluation individuelle de l'interaction orale particulièrement complexe et introduit, en s'inspirant de l'analyse de Chafe (1976), la notion de *repackaging*, pour désigner la recomposition d'un message au sein de son tour de parole. Selon cette auteure, les critères mis en avant dans le Cadre européen laissent de côté la dimension co-construite des échanges et les ajustements mutuels des discours produits. Elle illustre son propos théorique par un corpus d'exemples concrets prélevés à partir de productions en allemand.

Pascale Manoïlov fonde son analyse sur une tâche menée en anglais par des élèves de fin de collège (niveaux A1+ à B1) dans le cadre d'une recherche commanditée par la Direction de l'Évaluation, de la Prospective et de la Performance (DEPP). Son article apparaît étroitement lié à celui de Catherine Felce puisqu'il souligne lui aussi la nature co-construite du discours-en-interaction et la complexité à laquelle se trouvent confrontés les enseignants qui doivent utiliser le Cadre pour évaluer la compétence pragmatique de leurs élèves. Il en est même le prolongement dans la mesure où il propose à partir des critères spécifiés dans le CECRL une grille exploitable pour évaluer la compétence d'interaction orale.

C'est également vers l'interaction orale que s'oriente l'étude menée par **Yves Bardière**. Ce dernier décrit les mécanismes de l'immersion mimétique qui

stimule, selon lui, la mimésis comportementale et *in fine* la performance pragmatique et linguistique du candidat. Il essaie de montrer en quoi une approche de type *scenario-based language assessment* (SBLA) telle que le CLES, est particulièrement bien adaptée à la mise en place de ce processus d'immersion. Son argumentation théorique se fonde entre autres sur un sujet CLES B2 anglais disponible en ligne.

Claire Tardieu s'intéresse au CLES B2 dans son format désormais bi-niveau, un niveau B1 pouvant être décerné à un candidat qui passe une épreuve initialement prévue pour le niveau B2. Elle se demande si un sujet CLES conçu pour le niveau B2 permet de valider un niveau B1. La grille CEF-ESTIM constitue selon elle un outil d'estimation empirique pertinent pour fournir simultanément des informations sur ces deux niveaux.

Le Diplôme de Compétence en Langue (DCL) auquel **Maria Rosaria Gianninoto** et **Yan-Zhen Chen** font référence, présente de fortes similarités avec le CLES. Ceci n'a rien de surprenant, Claude Springer ayant participé à la conception et au pilotage de l'un comme de l'autre. Comme cet auteur le souligne lui-même, « le CLES a bénéficié de la réflexion [...] menée pour le DCL » (Springer 2002)². Il précise également dans la foulée que « l'évaluation de la compétence (au singulier) se fait [...] globalement et dans l'action à travers un scénario d'action que l'on retrouve aujourd'hui dans une vision maximaliste de l'approche actionnelle ». Il s'agit bien là de la visée holistique, actionnelle, scénarisée qui caractérise aussi le CLES. On comprend dès lors pourquoi, dans leur article ayant trait au chinois, Maria Rosaria Gianninoto et Yan-Zhen Chen considèrent que le DCL chinois fasse figure d'exception. Dans le contexte actuel du développement de l'offre de certification pour le chinois langue étrangère en France, le DCL représente en effet la seule certification de chinois qui soit adossée au CECRL et adopte explicitement une perspective actionnelle. Les auteurs mettent notamment en relief les éléments innovants du DCL, comme le

² Introduction par Claude Springer à son article (2002), "Effets du plurilinguisme sur les compétences à l'université : quelles compétences en langues pour les étudiants ?" in *Colloque, de l'Académie suisse des sciences humaines et sociales, Langues et production du savoir*, Université de la Suisse italienne, Lugano, 14 juin 2002, 19-32.

recours important aux ressources authentiques, l'interrelation entre les activités langagières et la sollicitation des candidats à mettre en œuvre des « stratégies compensatoires » en cas de difficultés de compréhension écrite et orale.

Aisha Nasimi, Giulia Peri, Adriano Gelo, Sabrina Machetti et Paola Masillo situent leur analyse dans le contexte italien. Ils constatent qu'un nombre croissant de candidats aux profils plurilingues variés se présentent aux épreuves de certification en italien L2. Ils concluent à la nécessité d'adopter un nouveau construit qui rende compte des caractéristiques des répertoires linguistiques plurilingues et des besoins communicatifs de ces candidats. Cette situation oblige les concepteurs et les institutions à s'interroger sur le format d'une certification plus adaptée à ces nouveaux profils. Les candidats qui passent la Certification d'italien langue étrangère (CILS) constituent un cas intéressant : le niveau A2 est nécessaire pour l'obtention d'un permis de séjour longue durée et le niveau B1 pour l'obtention de la nationalité italienne.

La contribution d'**Elena Ballarin** et **Paolo Nitti** s'inscrit elle aussi dans le contexte italien mais ne cible pas une certification particulière. Elle se situe au croisement entre didactique de l'italien L2 et linguistique textuelle. Les auteurs partent du constat que l'évaluation de la compétence textuelle est insuffisamment explorée dans deux contextes d'apprentissage : celui de l'alphabétisation et celui de l'italien académique. Ils décrivent le processus qu'ils ont suivi pour concevoir des descripteurs visant à évaluer la compétence textuelle en italien et l'expérimentation qui en est faite en classe avec un groupe d'étudiants en mobilité internationale.

La part relativement importante consacrée au CLES et à SELF participe de l'originalité de cette publication. En effet, à la différence des grands tests monolingues financés par des organismes privés, SELF et CLES affichent clairement leur vocation plurilingue et leur mission de service public. De plus, tous les auteurs concernés par ces deux dispositifs en sont également des acteurs importants : directeur scientifique, membre du comité scientifique, concepteur, évaluateur, formateur d'évaluateur, etc. Ils possèdent donc une connaissance intime du contexte dans lequel s'inscrit leur analyse. La réflexion

théorique s'appuie ainsi sur un vécu bien réel et des illustrations très concrètes. Cette réflexion dépasse le seul cadre de CLES et de SELF et manifeste une visée suffisamment généralisante pour recueillir l'attention de tous ceux qui s'intéressent aux questions d'évaluation.

Un volume consacré à un domaine aussi large que celui de l'évaluation des acquisitions langagières ne saurait prétendre à l'exhaustivité tant sur le plan quantitatif que qualitatif. Cet ouvrage ne déroge pas à ce principe. Pour conclure ce propos liminaire, il est tout au plus possible, dans le cadre de cette introduction générale, de mettre en évidence quelques domaines transversaux aux différents articles. Parmi les notions passées au crible de l'analyse figurent notamment celles d'**équité** (Tim McNamara, Anna Cardinaletti), de **co-construction** du discours en interaction, de **plurilinguisme** (Nasimi *et al.*). Si les différentes compétences langagières sont toutes évoquées au gré des analyses, on s'aperçoit que l'**interaction orale** donne lieu à des développements plus importants (Bardière, Felce, Manoïlov, Tardieu). De même, la **pragmatique** fait l'objet d'une analyse récurrente et plus ou moins approfondie selon le cas (Ballarin *et al.*, Bardière, Felce, Gianninoto *et al.*, Manoïlov, Tardieu). Les **modalités verbales, suprasegmentales**, ainsi que les **modalités non verbales**, en particulier kinésiques, le comportement des interlocuteurs, leurs "attitudes" (Springer 2017) sont aussi explorés à diverses reprises (Bardière, Felce, Higashi *et al.*, Manoïlov). Une approche **psychométrique** de l'évaluation est proposée dans les articles consacrés à SELF (Biagiotti *et al.* ; Briault *et al.*), mais elle transparaît aussi dans la contribution de Cardinaletti et, de manière plus indirecte, dans les analyses faisant référence aux descripteurs du CECRL (Felce, Tardieu, McNamara), si l'on considère avec Longuet et Springer que « Le projet d'échelle de North et Schneider (1998), adopté par le Conseil de l'Europe, s'inscrit clairement dans l'optique scientifique positiviste et psychométrique » (2018 : 35) ou encore que « Les nouveaux descripteurs du CECR/VC 2018 sont le résultat d'une intégration de descripteurs déjà validés psychométriquement et proposés par différents organismes de formation et de certification » (*ibid.*: 7).

Cette présentation nécessairement sélective ne saurait rendre compte de la richesse du contenu des pages qui vont suivre. Nous espérons simplement qu'elle aura su aiguïser la curiosité intellectuelle du lecteur qui, par choix ou par nécessité, se trouve confronté à la complexité de l'action d'évaluer au sens étymologique du terme (*esvaluer*), à savoir *extraire la valeur* de ce que l'on observe, ou éprouve un intérêt quelconque pour ce domaine. Nous tenons également à remercier les conférenciers invités qui ont aimablement apporté leur concours, les auteurs et tous ceux et celles qui de près ou de loin ont contribué à la réalisation de cet ouvrage. Nous saluons notamment la mémoire de notre collègue Rosanne Zanini, qui nous a récemment quittés, pour son engagement dans l'organisation du colloque EVAL19, la mise en ligne du site et la collecte des contributions scientifiques.

Références bibliographiques

Hamp-Lyons L. (2016) "Purpose of assessment", in D. Tsagari et J. Banerjee (eds.), *Handbook of Second Language Assessment*, Boston/Berlin: Walter de Gruyter Inc.

Lado R. (1961) *Language testing: The construction and use of foreign language tests: A teacher's book*, Bristol, Inghaterra: Longmans, Green and Company.

Longuet F. et C. Springer (2018) *Autour du CECR - Volume complémentaire (2018) : médiation et collaboration. Une didactique de la relation écologique et sociosémiotique*, Editions des archives contemporaines, France, doi : <https://doi.org/10.17184/eac.9782813004055>

Morrow K. (1979) "Communicative language testing: revolution of evolution?", in Brumfit, C.K. et K. Johnson (eds.), *The Communicative Approach to Language Teaching*, Oxford: Oxford University Press, 143-159.

Springer, C. (2017) "De la tâche simple à la tâche complexe (Haziniango na haziro). Quelles évolutions pédagogiques et évaluatives ?", *Évaluation des compétences en français. Enjeux actuels en contextes plurilingues et*

pluriculturels,
01709031

<https://hal.archives-ouvertes.fr/hal-01709031/document>