



**HAL**  
open science

## Les tests de langues : une nécessité ?

Claire Tardieu

► **To cite this version:**

Claire Tardieu. Les tests de langues : une nécessité?. Le Français dans le monde. Recherches et applications, 2013, L'évaluation en didactique des langues/cultures. Continuités, tensions, ruptures, 53, pp.67-79. hal-01423819

**HAL Id: hal-01423819**

**<https://hal.science/hal-01423819>**

Submitted on 13 Jan 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.

## **Les tests de langues : une nécessité ?**

*Français dans le monde – recherches et applications* sur l'évaluation en didactique des langues

Résumé : La place croissante de l'évaluation dans le domaine scolaire de l'enseignement-apprentissage des langues pose un certain nombre de questions quant à la légitimité de la demande en tests standardisés et calibrés sur les niveaux du Cadre Européen Commun de Référence en Langue (CECRL). Aux tests psychométriques de tradition anglo-saxonne s'ajoutent aujourd'hui des tests de type actionnel qui adoptent une démarche « holistique ». Quelle est leur contribution véritable en termes de pertinence, de validité et de fiabilité (Horner 2010) ? Peut-on envisager des modes alternatifs ou complémentaires comme la grille CEF-ESTIM (Tardieu et al. 2011) et le C-Test (Reichert 2011) ?

Mots-clés : Evaluation, Certification, CECRL, CEF-ESTIM, C-Test.

### **Language Tests: A necessity?**

The increasing importance of evaluation in the field of foreign language teaching and learning at school raises a certain number of questions about the legitimacy of the demand for tests standardised and calibrated on the Common European Framework of Reference for Languages (CEFR). In addition to psychometric tests belonging to the English tradition, action-oriented tests that take a holistic approach have appeared. What is their true contribution to the world of testing, in terms of relevance, validity and reliability? (Horner 2010) Would it not be possible to profit by alternative or complementary methods such as the CEF-ESTIM grid (Tardieu & al., 2011) and the C-Test (Reichert, 2011)?

Key-words: Evaluation, Certification, CEFR, CEF-ESTIM, C-Test.

### **Introduction**

Pour commencer, interrogeons-nous sur cette réflexion de Gardner : « Quand l'évaluation se fonde progressivement dans le paysage, elle n'a plus besoin d'être distinguée des autres activités de la classe. A l'instar de ce qui se passe dans un bon apprentissage, les enseignants et les élèves sont toujours en train d'évaluer. Il n'y a plus alors nécessité d'enseigner pour l'évaluation parce que l'évaluation est omniprésente ; le besoin de tests formalisés pourrait bien disparaître » (Gardner, 1996 : 115). L'image suscitée ici est celle d'une évaluation de type fondu-enchaîné avec l'apprentissage conçu comme une action mesurable sur le terrain ; autrement dit une évaluation formative, accompagnant la progression comme son ombre, et rendant de fait caduc tout autre type de mesure distanciée. Pure utopie ou simple réalisme ? La question se pose aujourd'hui d'autant plus que des outils d'évaluation en langues distanciés se développent en réponse à une demande sociale et politique effective et pressante. La suggestion de Gardner, qui provient d'une remise en cause des tests psychométriques à l'anglo-saxonne, se justifie-t-elle encore avec l'apparition d'un nouveau type de test, de quatrième génération que l'on pourrait appeler « l'approche sociolinguistique intégrée de type actionnel » ?

En premier lieu, nous présenterons les trois premières approches de l'évaluation formalisée en montrant leurs points forts et leurs points faibles. Puis nous nous demanderons si la quatrième approche, contemporaine, pallie effectivement les défauts des précédentes. Enfin nous proposerons une définition souple de l'évaluation et des pistes possibles pour un usage modéré et utile.

## **I- Les tests et les 3 premières approches**

Spolsky (1981) divise l'histoire du *testing* en langues en trois grandes approches : pré-scientifique, structuraliste psychométrique et sociolinguistique-intégrative. (cf Reichert 38-39).

### **I-1 L'approche pré-scientifique**

L'approche pré-scientifique s'inscrit dans une logique d'examens et de juges, mais manque d'objectivité et de fiabilité, la fiabilité étant, selon Horner, « la mesure selon laquelle on retrouvera le même classement des candidats dans deux passations (réelles ou simulées) des mêmes épreuves » (Horner, 2010 : 17). Elle manque aussi de validité – à savoir, « la garantie que ce qui est effectivement testé est ce qui, dans le contexte donné, doit être évalué et que l'information recueillie donne une image exacte de la compétence des candidats en question » (Horner, 2010 : 17). L'approche pré-scientifique est celle pratiquée au Baccalauréat, par exemple, qui ne s'appuie ni sur des théories du langage ou de l'apprentissage, ni sur une méthodologie scientifique. Dès la première moitié du 20<sup>e</sup> siècle, la docimologie (en particulier la commission Carnegie de 1932) a montré qu'il faudrait 127 correcteurs pour obtenir la valeur stable d'une dissertation de philosophie, 78, pour une composition française, 28 pour un devoir d'anglais, 19 pour une version latine et 13 pour un devoir de mathématiques... (Laugier & Piéron, 1938 ; Piéron, 1963 : 23)

### **I-2 L'approche structuraliste-psychométrique**

L'approche structuraliste-psychométrique intervient pour répondre à la demande d'objectivité et de fiabilité, dans un souci d'évaluer la maîtrise du système linguistique (Mc Namara, 2000), avec en contrepartie un effet d'atomisation et de décontextualisation. On trouve trace de cette approche dans une partie de l'épreuve du baccalauréat appelée « épreuve de compétence linguistique » instaurée en 1984 (B.O.E.N. n° 27 du 7 juin 1983) et abolie en 2002 (B.O.E.N. n° 23. Note de service 2001-091 du 7 juin 2001).

### **I-3 L'approche sociolinguistique intégrative**

A partir des années 60, la demande est en faveur d'une approche intégrée, testant à la fois la compétence linguistique fine et les savoir-faire (CO, CE, PO, PE) malgré les difficultés en termes de coût et d'attribution de scores en productions orales ou écrites.

Le TOEFL (Test Of English as a Foreign Language) est un bon exemple de test qui associe l'approche structuro-psychométrique avec des items discrets et l'approche intégrée introduisant les *skills*, soit une évaluation des compétences de réception et de production en contexte social plausible. Ainsi, aux exercices de compétence linguistique pure s'ajoutent des items sous forme de QCM, par exemple, qui testent la capacité à comprendre un texte oral ou écrit « authentique » ou à produire un discours adapté à un type d'interlocuteur ou à une forme particulière (lettre, article, etc.) Ce type d'évaluation semble à la fois valide, fiable, et faisable au sens où « les examinateurs ne disposent que d'un temps limité, ne voient qu'un échantillon limité de performances et qu'il y a des limites au nombre et à la nature des catégories qu'ils peuvent manipuler comme critères (Horner, 2010 : 17). »

Alors pourquoi une quatrième approche ? Quels besoins nouveaux se sont-ils fait jour ?

### **I-4 Les limites de l'approche sociolinguistique intégrative**

#### **I-4-1 Limites quant à la compétence évaluée**

Une critique, celle de Bachman (1990 : 82), porte sur le fait que dans ce type de tests alliant des savoir-faire et des éléments discrets, il est impossible de déterminer la nature du lien entre les deux. Ensuite, ce type d'évaluation ne mesure que certains aspects de la compétence à communiquer langagièrement. On peut citer à cet égard la synthèse réalisée par Springer (2000) dans sa « marguerite de la communication » (reprenant l'évolution de Chomsky à

Canale et Swain en passant par Hymes et Charaudeau) et bien sûr les travaux de Oller (1976) et Bachman (1990). Oller affirme que la compétence langagière n'est pas seulement constituée d'éléments séparés. D'où l'idée d'une compétence intégrative : « The Unitary Competence Hypothesis ». Selon Oller, la production écrite ou orale n'est pas séparée de la réception, car le locuteur doit en permanence contrôler sa production pour s'assurer d'être compris. L'inverse est par ailleurs tout aussi vrai : en phase de réception d'un message, l'auditeur ou le lecteur doivent formuler des hypothèses sur la suite du message. D'autre part, les tests formalisés ne mesurent pas, en particulier, les dimensions véritablement sociologiques et pragmatiques de la communication, ni les dimensions cognitive et intellectuelle, ni encore moins la capacité de communication réelle chère à Bachman. Bachman fait en effet la distinction entre la compétence à communiquer langagièrement (*communicative language ability*) et la communication véritable (*actual communication*). La première inclut une dimension liée aux savoirs (« language knowledge dimension ») comprenant elle-même des compétences langagières organisationnelles et pragmatiques, des compétences stratégiques et des mécanismes psycho-physiologiques ; la seconde se caractérise par sa dimension « d'utilisation réelle » (Bachman, 2007 : 54). Puren souligne ce problème en écrivant :

« Alors que la recherche d'objectivité voire de « scientificité » dans l'évaluation implique une approche très analytique – avec au minimum une dissociation des différentes compétences de compréhension/expression orale/écrite –, l'évolution récente de la didactique des langues a réactivé l'idée ancienne d'une indissociabilité entre les trois objectifs fondamentaux. D'une part l'objectif actuel de compétence de communication inclut une « composante socioculturelle » (la maîtrise de la langue suppose la maîtrise des codes qui régissent son emploi en situation). L'approche cognitive, d'autre part, met l'accent sur l'importance des activités réflexives (et donc de la formation intellectuelle) dans le processus d'apprentissage linguistique. Enfin l'approche par les tâches, vers laquelle semble s'orienter actuellement l'enseignement scolaire des langues en France, ne va pas simplifier le problème puisque c'est principalement le produit final d'une activité personnelle ou en petits groupes qui devra désormais être évalué, alors que jusqu'à présent l'enseignant pouvait dans une certaine mesure observer et évaluer directement et en temps réel l'activité même d'apprentissage parce qu'elle était étroitement liée à son enseignement : comment faire la part, dans l'exposé collectif d'un groupe d'élèves, entre les compétences langagières et les compétences culturelles de chacun, la maîtrise des ressources documentaires, ou encore l'efficacité du travail de groupe ? » (Puren, 2001 : 14)

Il semble que Puren poursuive ici dans la droite ligne des travaux de Bolton (1991) qui, en retraçant l'évolution de l'objet évalué, constatait déjà voici plus de vingt ans, avec l'avènement de l'approche communicative dans l'enseignement-apprentissage des langues, qu'il est beaucoup plus facile d'évaluer des connaissances que des compétences, *a fortiori* lorsque ces dernières s'exercent dans un travail collaboratif et, conviendrait-il d'ajouter, « proche de la vie réelle ».

Gardner a raison sur ce point, et Bachman tout autant : les tests formalisés ne certifient pas la compétence réelle à communiquer langagièrement.

#### I-4-2 Limites quant à la fiabilité des items

L'expérimentation menée au sein du projet européen EBAFLS (Building a European Bank of Anchor Items For Language Skills) a montré que des items corrélés aux niveaux du CECRL par des experts de chaque pays et ayant fait l'objet d'une validation psychométrique ne fonctionnaient pas de la même manière d'un pays à l'autre, avec des élèves de huit pays européens censés être au même niveau du CECRL et qu'il était par conséquent impossible, pour l'instant tout au moins, de créer une banque d'items d'ancrage (Tardieu, 2010). Comment expliquer les causes de l'échec de ce projet mené de 2005 à 2007 ?

La taxonomie de Van de Vijver et Hambleton (1996) établit trois biais possibles, à savoir :

- un premier biais conceptuel, soit la non-équivalence des concepts d'un groupe culturel à l'autre ;
- un second, méthodologique, résultant de la manière d'administrer les instruments ;

- un troisième, lié aux items eux-mêmes (des traductions inadéquates par exemple).

Au cours du projet EBAFLS, ces différents biais ont pu être observés (voir Tardieu 2010).

Concernant le biais lié aux items, une différence essentielle est apparue à l'issue du projet EBAFLS. Les analyses des résultats ont révélé un important fonctionnement différentiel à l'item (DIF) rendant impossible l'établissement d'échelles concomitantes de pays à pays. En d'autres termes, un même item ne se comporte pas de la même manière selon le pays. Ce qui revient à dire que les compétences d'un élève au niveau B1 dans un pays ne sont pas équivalentes à celles d'un élève au niveau B1 dans un autre pays. Cette différence de fonctionnement de l'item s'exprime à la fois en termes de difficulté et en termes de discrimination.

En matière de biais lié à l'administration du test, on s'est aperçu que le type de consignes n'était pas le même selon les pays. Certains adressant des consignes concernant l'ensemble du test, d'autres des consignes par tâches, ou pour chaque item.

Le choix de la langue de passation relève aussi d'un biais lié à l'administration du test, comme l'a montré le pilote effectué en utilisant les deux circonstances. Un protocole fut établi en compréhension de l'écrit pour le français, et administré dans deux pays. Ce pilote a fait apparaître que les scores étaient plus élevés lorsque le test était administré en L1. Néanmoins, il fut décidé d'administrer le test en L2, en raison de l'objectif même du projet : la création d'une banque d'items communs.

On peut se demander si le choix de la langue d'administration du test ne relève pas également d'un biais conceptuel, certains pays ne se posant pas la question, d'autres y accordant une grande importance.

Enfin, des différences de jugement ont été observées au moment de la spécification et de la standardisation des items, en amont de la passation, et également en aval par le retour des professeurs ayant fait passer les items dans les différents pays.

Ainsi rien ne permet d'affirmer l'existence d'un lien ferme entre EBAFLS et le CECRL. On peut légitimement se demander comment les tests contemporains de type actionnel parviennent à lier leurs composantes aux paramètres du CECRL.

## **II- L'approche sociolinguistique intégrative de type actionnel**

### **II-1 les fondements**

Pour compléter la classification de Spolsky, une quatrième approche, non psychométrique, devrait être ajoutée : « l'approche sociolinguistique intégrative de type actionnel ». Cette approche s'appuie sur une conception socio-pragmatique de la communication appliquée à l'enseignement dont on trouve les fondements chez Nunan (1989), Widdowson (2002[1990]), R. Ellis (2003), entre autres. Ces chercheurs ont posé les bases de l'approche actionnelle, dans la poursuite de l'approche communicative : on ne communique pas seulement pour agir *sur* autrui mais aussi et surtout pour agir *avec* autrui. L'accent est mis sur l'utilisation de la L2 aux fins d'accomplir des tâches sociales qui vont de la simple déclinaison d'identité à la discussion de haut vol dans des domaines de spécialité. La nouveauté de cette approche est de mettre l'accent sur l'ancrage dans la réalité de la communication, et de concevoir les activités de réception, de production, d'interaction et de médiation de manière holistique, concourant toutes à la réalisation d'un projet. Ce type d'enseignement appelle un nouveau type d'évaluation, plus en phase avec des scénarii réels, et davantage apte à mesurer des compétences indépendamment de tout contexte d'apprentissage.

Pour reprendre la taxonomie d'Alderson (2002), il s'agit ici de faire la distinction entre le test de type « achievement » relié à un apprentissage en amont et le test de type « proficiency », indépendant d'une formation et tourné vers l'avenir. Ainsi fera-t-on mieux la distinction entre l'évaluation du savoir (ou du niveau) qui est « l'évaluation de l'atteinte d'objectifs spécifiques », « centrée sur le cours », correspondant à « une vue de l'intérieur » et

l'évaluation de la capacité (mise en œuvre de la compétence ou performance), c'est-à-dire « de ce que l'on peut faire ou de ce que l'on sait en rapport avec son application au monde réel » (CECRL, 2001 :139), l'évaluation de la capacité correspondant « à une vue de l'extérieur ». Certains certificats, comme le *Deutsches Sprachdiplom* (certificat d'allemand) ou le Certificat en Langue de l'Enseignement Supérieur (désormais CLES), se réclament de cette approche.

## **II-2 Les propositions de l'institution**

### **II-2-1- Développer les certifications**

L'article 3 du décret n°2005-1011 du 22 août 2005 relatif à l'organisation de l'enseignement des langues vivantes étrangères dans l'enseignement scolaire français prévoit la mise en place de certifications des connaissances et compétences acquises en langues vivantes étrangères.<sup>1</sup> Ce qui signifie que les élèves devraient pouvoir se voir offrir à différents moments du cursus définis par l'institution des évaluations en langues indépendantes de leurs cours de langues, validant leurs performances selon les échelles du CECRL.

Depuis le décret de 2005, au sein même de l'institution, des efforts sont effectivement menés pour « crédibiliser » les diplômes. L'examen de langue pour l'obtention du Diplôme National du Brevet (DNB – B.O. n° 22 du 7 juin 2007) qui s'est adaptée à l'évaluation de 5 compétences (compréhension de l'oral, expression orale en continu, interaction orale, compréhension écrite et expression écrite) donne lieu à la délivrance non d'un certificat à proprement parler, mais d'un diplôme, où la langue étrangère est affectée d'un coefficient 1 et validée au niveau A2 du CECRL. Dans ce cas cependant, l'évaluation n'est pas conçue par un Institut extérieur du pays de la langue étudiée, mais par les professeurs eux-mêmes ou par une équipe de professeurs.

Il en va de même pour les nouvelles épreuves du Baccalauréat entrées en vigueur à la session 2013, avec une validation à B2 (LV1) ou B1(LV2). On devine les limites de ces diplômes délivrés de l'intérieur, tout au moins en termes de validité et de fiabilité.

Un type de certification indépendant aurait plusieurs avantages : non seulement il posséderait une valeur non limitée au territoire français et à ses traditions éducatives, mais il permettrait aussi à terme de « libérer » les professeurs d'une partie de leur tâche (qui les oblige à être à la fois formateur et évaluateur-certificateur) et de redonner ses lettres de noblesse à leur mission d'évaluation formative propre à accompagner l'apprentissage, selon la vision de Gardner. On pourrait ainsi d'un côté évaluer pour accompagner l'apprentissage en fonction des objectifs requis à chaque niveau, et de l'autre proposer des tests en langue déconnectés de l'apprentissage. Une telle proposition qui se dessine dans l'environnement institutionnel paraît bien adaptée au contexte contemporain.

A titre d'exemple, une évaluation en allemand du Goethe Institute indexée sur les niveaux du CECRL est proposée aux élèves volontaires en fin de Troisième ou de Seconde depuis 2006 sur la base du volontariat. Les épreuves portent sur la compréhension écrite, la compréhension orale (écoute d'un document sonore et réponses par écrit à un questionnaire) et l'expression écrite. Les épreuves orales prennent la forme d'entretiens visant à évaluer l'expression orale (la capacité à converser sur des sujets familiers pour le niveau A2 par exemple). La certification est acquise à un niveau précis du CECRL quand les candidats ont obtenu ce niveau dans chaque activité langagière. Ils reçoivent alors le diplôme DSD (*Deutsches Sprachdiplom*) avec le relevé de leurs notes par activité langagière. Tous les élèves qui ont réussi au moins une activité se voient remettre une attestation signée par l'autorité académique.

### **II-2-2 Les limites de ces certificats**

---

<sup>1</sup> JO du 25 août 2005, Titre 1, articles 3 à 7.

Le projet EBAFLS ayant montré la difficulté de standardisation des items, on peut en premier lieu douter de la valeur de ces certificats ou diplômes en termes d'attestation de niveaux.

Ensuite, il apparaît que les composantes de la compétence à communiquer langagièrement sont évaluées de manière séparée et non véritablement holistique.

Enfin, ces évaluations sont sans doute moins valides et moins fiables que des évaluations de type approche sociolinguistique intégrée (avec volet psychométrique).

D'autres certificats, comme le DCL ou le CLES, sont également proposés, voire requis à des niveaux supérieurs du cursus. Offrent-ils une pertinence plus grande ? On prendra ici l'exemple du CLES2 dont l'obtention est en principe requise à la sortie du Master (en France en tout cas).

### **II-3 Le CLES**

Le certificat en langue de l'enseignement supérieur, niveau 2 n'est pas exempt d'intérêt : il entend proposer une évaluation qualitative de la compétence de communication langagière conçue dans sa globalité. Il se déroule pendant 3 heures, suivant une phase de compréhension de l'oral prenant appui sur des documents audio ou vidéo authentiques d'environ 800 mots au total, puis d'une phase de compréhension de l'écrit (environ 2000 mots à traiter en compréhension globale et sélective), puis une production écrite contextualisée de 300 mots, et enfin une phase d'interaction orale de 10 mn en présence d'un examinateur, le tout articulé autour d'une thématique commune et les différentes phases fonctionnant en synergie (B.O. n°20 du 17 mai 2007).

Ce certificat a le mérite de solliciter les multiples facettes de la compétence à communiquer langagièrement (CO, CE, PE, IO) en une tentative d'évaluation holistique autour d'une thématique suivie devant être traitée de façon multimodale à travers les différents supports. L'innovation majeure consiste à introduire l'interaction orale entre deux candidats, au lieu de celle entre l'examineur et le candidat, interaction qui repose sur une conception de la tâche pouvant se rapprocher des types définis par Prabhu ou Pica, Kanagy et Falodun. Prabhu (1987) distingue trois grands types de tâches fondés sur le genre d'activité cognitive impliquée (déficit d'information, déficit de raisonnement, déficit d'opinion (Ellis, op.cit., 214). Pica, Kanagy et Falodun (1993) opèrent une classification psycholinguistique fondée sur des catégories interactionnelles qui envisagent la relation entre les interactants – par exemple, qui détient l'information nécessaire à l'accomplissement de la tâche ? Ou encore, la tâche requiert-elle que les participants parviennent à un accord final (cf. Ellis, op.cit., 215) ? Mais dans quelle mesure les sujets du CLES proposés au niveau 2 empruntent-ils véritablement à de tels fondements théoriques ?

Tout d'abord on constate que même si la thématique est la même tout au long de la passation du test, et même si l'interaction orale requiert que l'on recherche un consensus par rapport à un problème évoqué tout au long de l'épreuve, les différentes composantes de la compétence à communiquer langagièrement n'en sont pas moins évaluées séparément, comme dans l'approche antérieure non actionnelle. Et cela est d'autant plus vrai que les performances ne sont pas compensables (ce qui cause d'ailleurs un taux d'échec assez élevé au CLES2 – plus de 50%). Et peut-il en être autrement ? L'intérêt réel du CLES nous semble résider plutôt dans le fait qu'il peut modéliser l'apprentissage en amont d'une manière plus satisfaisante d'un point de vue didactique que les tests habituels, ce qui est un atout majeur, par rapport à d'autres certificats.

Mais si ces tests distancés, non psychométriques, ne testent pas la compétence langagière d'une manière plus globale que les tests de l'approche précédente, sont-ils vraiment utiles en tant que tels ? Ne peut-on se contenter, comme le préconise Gardner, d'une évaluation au plus près des progrès de l'élève ? Tout dépend en fait de l'information que l'on recherche.

### **III- Quelles perspectives ?**

#### **III-1 Evaluer la performance du système**

En termes institutionnels, le test formalisé semble avoir sa place pour évaluer la performance du système. Ainsi, les évaluations PISA ou celles menées par la Division de la l'Evaluation de la Prospective et de la Performance (DEPP) créée en 1987<sup>2</sup> renseignent utilement les professionnels de l'enseignement et les politiques sur la pertinence du système.

En termes d'évaluation ou de certification individuelle, d'autres modalités nous semblent pouvoir être envisagées.

#### **III-2 Evaluer les performances individuelles**

##### **III-2-1 Estimer le niveau d'un cours**

La première modalité consiste à valider le niveau individuel par l'estimation du niveau d'un cours. C'est ce que font déjà, d'une certaine manière, de nombreux instituts d'évaluation (Cambridge Institute, Goethe Institute, CIEP, etc.), en proposant des thématiques différentes selon le niveau de la formation et du test qui la clôt. Par exemple, à A2, le candidat est interrogé sur sa famille, ses hobbies, etc., en accord avec les descripteurs de niveaux du CECRL. Au niveau B1 on lui demande de raconter un film ou un livre. La grille CEF-ESTIM (Tardieu et al., 2011) est un outil permettant aux professeurs de toute langue et enseignant à tous les niveaux du cursus de situer le niveau de leur cours en relation avec les descripteurs du CECRL. Ce faisant, il devient possible de valider le niveau des élèves en fonction du niveau estimé du cours ou de la formation en question.

##### **III-2-2 Le C-Test**

La seconde modalité consiste à utiliser un test simple, fiable, valide, faisable et très économique, le C-Test. En fait, le C-Test est une forme améliorée du test de closure. Il s'agit de fournir un texte écrit dans lequel on efface la deuxième moitié du mot tous les 2 mots, tout en laissant intacte la première et la dernière phrase du texte. Chaque mot tronqué vaut pour un item d'évaluation. Un C-Test est généralement constitué de 4 à 6 textes et chaque texte contient 20 à 25 items. Ce qui fait en tout environ 80 à 100 items par C-Test (Grotjahn, 2006, Reichert, 2011 : 40). Le candidat doit s'efforcer de reconstituer le texte au mieux. (A titre indicatif un locuteur natif réalise une performance de 90% en général)

Dans sa thèse, Reichert explique que l'on s'est aperçu que les items avaient une forte corrélation entre eux, et donc que la compétence linguistique pouvait être unitaire conformément à la théorie développée par Oller. (Reichert 41). Mais surtout, Reichert a montré le haut degré de corrélation entre les résultats au C-Test et ceux obtenus par les mêmes étudiants au Test de Connaissance du Français (TCF), de type TOEFL amélioré comprenant l'évaluation en compréhension écrite et orale, en expression écrite et orale, et en compétence linguistique. Ce qui revient à dire qu'un test très facile à créer et très peu coûteux à faire passer fournit une information aussi fiable que d'autres tests beaucoup plus lourds sur la compétence langagière. Bien évidemment le C-Test ne fournit pas de renseignements précis sur chacune des compétences, ni sur les connaissances, et ne peut donc remplacer d'autres tests dont tel est l'objet, mais il peut tout à fait valider un niveau général.

#### **III-3 Évaluation et certification**

##### **III-3-1 Fonction sociale de l'évaluation**

---

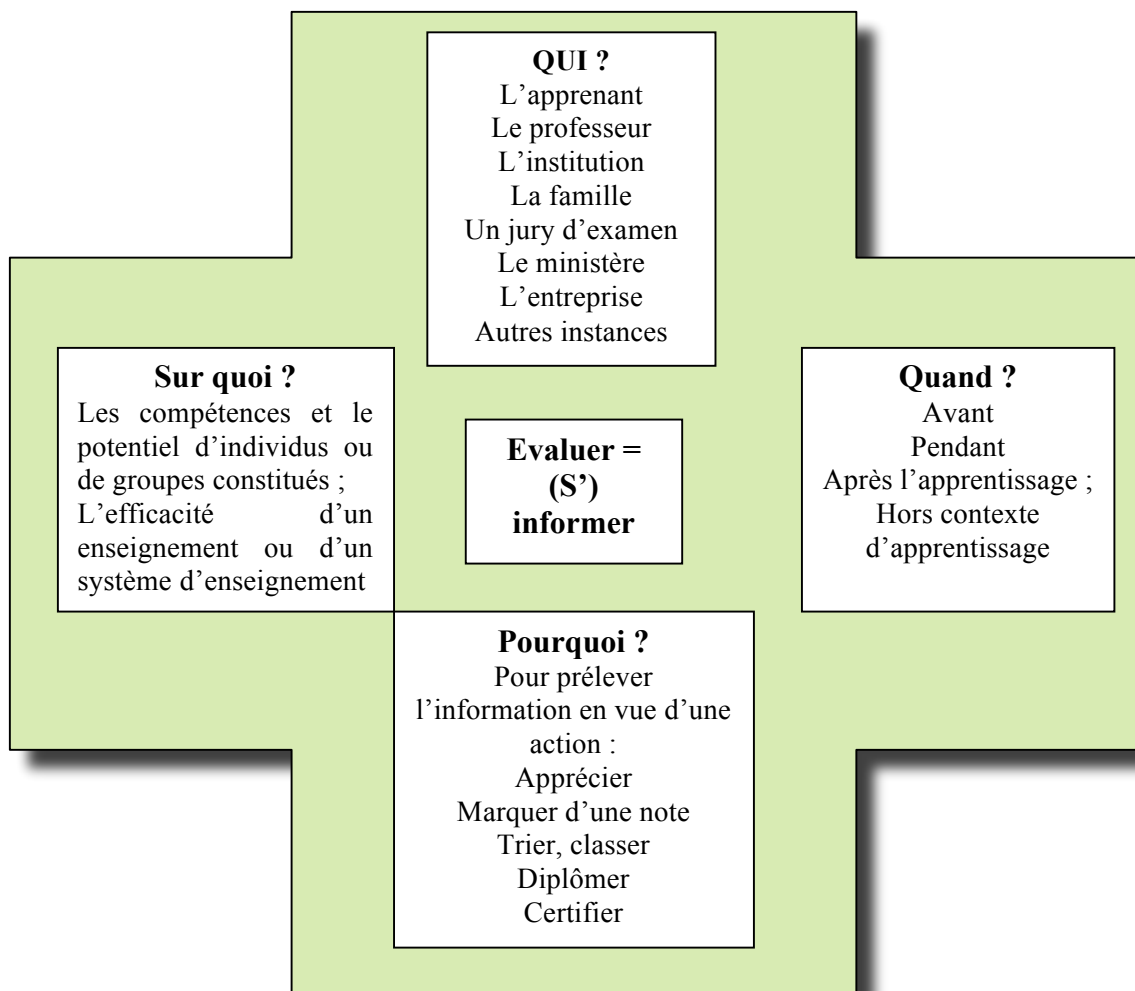
<sup>2</sup> La DEP fut créée en 1987 sous ce nom mais existait précédemment sous diverses appellations en tant que service statistique du ministère. C'est en 1975/1976 seulement qu'est conçue et réalisée la première évaluation d'élèves en France. Par la suite, la DEP est devenue DPD (Direction de la programmation et du développement) en 1997, a repris le nom de DEP en 2003. C'est en 2006 que le deuxième P fut ajouté avec la restructuration du ministère en avril, passant sous la coupe du nouveau Secrétariat général.



De Ketele ajoute une dimension importante, au cœur de notre problématique, à savoir, sa fonction sociale. Selon lui, évaluer c'est confronter un ensemble d'informations à un ensemble de critères (référentiel) en vue de prendre une décision (De Ketele, 1980). De quelle décision parle-t-on ici ? S'agit-il d'une décision personnelle de l'apprenant, de celle du professeur ou du formateur évaluateur, d'un jury d'examen, d'un employeur ? Par définition, une évaluation distanciée de type certification se situe hors de la sphère familière de la classe, et implique une prise de décision qui échappe au sujet évalué. Plus le besoin d'un regard distancié est affirmé, plus le sujet semble devoir être mis à distance de la prise de décision, avec des effets potentiellement nuisibles pour lui.

Si l'on veut vraiment privilégier l'aspect individuel et social de l'évaluation, il conviendrait d'afficher clairement une éthique. A cet égard, cette seconde réflexion de Gardner paraît appropriée : « Evaluer, c'est recueillir des données sur les compétences et les potentiels des individus dans le double objectif de leur en faire retour utilement et de procurer des informations indispensables à la communauté environnante (Gardner, 1996 : 114) Ajoutons cependant que si l'évaluation doit s'avérer utile et profitable à l'individu et à la communauté, elle devrait avant tout permettre la mesure positive du potentiel de chacun pour le bénéfice de tous.

En définitive, évaluer, c'est (s') informer à des degrés divers, dans des domaines divers, avec des intentions diverses émanant de personnes ou d'instances également diverses. On peut résumer cet éventail de combinaisons possibles par le schéma suivant :



*Schéma de l'évaluation située*

Il paraît ainsi important de situer toute forme d'évaluation ou de certification et de définir en amont les paramètres qui lui donneront à la fois forme et sens.

La proposition de Gardner citée en entrée en matière constitue une option pragmatique qui pourrait suffire si l'on utilisait en parallèle des référentiels d'apprentissage. Le temps que l'on consacre à évaluer serait utilisé à concevoir des activités d'apprentissage reliées à ces référentiels. D'une certaine manière, c'est ce que le CECRL a tenté de faire en proposant des descripteurs de niveaux de compétences. La difficulté à relier directement les compétences individuelles à ces descripteurs nous engage à imaginer une troisième voie qui est la suivante : prévoir des sessions d'apprentissage (en classe ou en situation réelle d'apprentissage) constitués de projets et d'activités reliés à ces descripteurs, en une sorte de balisage initial et offrir en parallèle des occasions de tester son niveau, à certains moments du cursus, indépendamment de l'apprentissage.

## Conclusion

Dans ces conditions, la proposition de Gardner n'est plus utopique mais s'inscrit au contraire dans un réalisme pragmatique. Les informations recueillies sur chaque apprenant lui sont retournées sur le mode de son choix (appréciation, note, couleur, diplôme, etc.) : elles lui appartiennent de droit en tant que sujet ; et la société quant à elle peut se satisfaire de l'information pourvue par le référentiel correspondant à la session d'apprentissage.

Si l'on veut absolument répondre à la demande du monde de l'entreprise, on pourrait effectivement offrir la possibilité à certains moments du cursus d'obtenir une certification individuelle par le biais d'un C-Test (peu coûteux) ou d'une évaluation-certification de type CLES ou Cambridge Exams, plus coûteuse et pas forcément plus fiable –sans oublier d'indiquer une date de péremption...

## Références

- Alderson, C., Figueras, N., Kuijper, H., Nolte, G., Takala, S., Tardieu, C. (2006), "Analysing Tests of Reading and Listening in Relation to the *Common European Framework of Reference*, The experience of the Dutch CEFR Construct Project", in *Language assessment Quarterly*, Volume 3, n°1. p. 3-30.
- Bachman, L.F. (1990), *Fundamental Considerations in Language Testing*, Oxford, UK , Oxford University Press.
- (2002), "Some reflections on task-based language performance assessment", *Language Testing*, Vol.19, n°4, p. 453-476.
  - (2007), "What is the construct ? The dialectic of abilities and contexts in defining constructs in language assessment, in Fox J. & al. (Eds.), *Language testing reconsidered*, Ottawa, CA, University of Ottawa Press, p. 41-72.
- Conseil de l'Europe (2001), *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*, 1<sup>e</sup> éd. 1996, 2<sup>e</sup> éd. corr. 1998, Paris, Didier.
- ESOL Examinations, University of Cambridge (2004), *Samples of oral production, illustrating for English, the levels of the Common European Frame of Reference*, Strasbourg, Council of Europe.
- Canale, M. & Swain, M. (1980), "Theoretical bases of communicative approaches to second language teaching and testing", *Applied Linguistics*, Volume 1, 1-47.
- Ellis, R. (2003), *Task-based Language Learning and Teaching*, Oxford : Oxford University Press.
- Gardner, H. (1996), *Les intelligences multiples*, Paris, Retz.
- Gille, E & Noijons, J. (2010), "Issues in constructing an international item bank: Differences in item behaviour across countries", *Revue française de linguistique appliqué*, vol. 15, n°1.

- Grotjahn, R., Klein-Braley, C. & Raatz, U. (2002), "C-Tests: an overview", in J.A. Coleman, R. Grotjahn & Raatz, U. (Eds.), *University Language Testing and the C-Test*, Bochum, Germany, ASK-Verlag, p. 93-114.
- Hymes, D. (1972), « On Communicative Competence », in J.B. Brice and J. Holmes (Eds.), *Sociolinguistics*, Reading, Penguin Modern Linguistics.
- Laugier, H., Piéron, H Mme H. Piéron, Toulouse, E & Weinberg, D. (1934), *Etudes docimologiques sur la perfection des examens et concours*, Paris, Conservatoire des arts et métiers.
- Laugier, H. & Weinberg, D. (1938), *Recherches sur la solidarité et l'indépendance des aptitudes intellectuelles d'après les notes des examens écrits au baccalauréat*, Paris, Chantenay.
- Mac Namara, T. (2000), *Language Testing*, Oxford, UK, Oxford University Press.
- Piéron, H. (1963), *Examens et docimologie*, Paris, PUF.
- Nunan, D. (1989), *Designing Tasks for Communicative Classroom*, Cambridge, Cambridge University Press.
- Oller, J.W. (1976), "Evidence for a general language proficiency factor: an expectancy grammar", *Die Neueren Sprachen*, n°75, p.165-174.
- Quivy, M. & Tardieu, C. (2002), *Glossaire de didactique de l'anglais*, Paris, Ellipses.
- Reichert, M. (2011), *The validity of the C-Test Revisited: Findings from a Multilingual Environment*, Thèse de Doctorat en psychologie de l'université du Luxembourg, sous la direction de R. Martin.
- Scriven, M. (1967), « The methodology of evaluation », in *Perspectives of curriculum evaluation*. Chicago, Rand Mc Nally, p.39-83.
- Springer, C. (2000), « Les multiples facettes du concept de « compétence » : que peut-on en retenir pour l'enseignement et l'apprentissage des langues ? », dans *Cognition, langue et culture, éléments de théorisation didactique*, Actes de la journée d'étude du 17 mars 2000, GEPEP, Université ParisVII - Denis Diderot.
- Tardieu, C. (2005), *L'épreuve de didactique aux concours. L'évaluation en langues*, Paris, Ellipses, Collection CAPES/Agrégation langues.
- (2009), « Corriger ou évaluer », dans *La correction dans l'enseignement des langues de spécialité, Les Cahiers de l'APLIUT, Pédagogie et recherche*, vol. 28, n°3, p. 9-25.
  - (2010), « Votre B1 est-il mon B1 ? L'interculturel dans les tests d'évaluation en Europe », dans *Recherches en didactique des langues - Les langues tout au long de la vie, Cahiers de l'ACEDLE*, vol. 7, n°2, 2010, p. 225-239. Version complète disponible en ligne : <http://acedle.org/spip.php?article3103>
- Tardieu, C., Raili, H., Reichert, M., Lehmann, M. (2011), *CEF-ESTIM*. Disponible en ligne : <http://cefestim.ecml.at/>
- Tyler, R. (1950), *Basic principles of curriculum and instruction*, Chicago, University Press.
- Van de Vijver, F.J.R., et Hambleton, F.K. (1996), "Translating Tests: Some practical Guidelines". *European Psychologist*, Vol. 1. n°2, 89-99.
- Widdowson, H. G. (2002), *Aspects of language Teaching*, Oxford University Press (5<sup>th</sup> impression, 1<sup>st</sup> published 1990).

#### Textes officiels:

- B.O.E.N. n° 27 du 7 juin 1983 (Baccalauréat).
- B.O.E.N. n° 23. Note de service 2001-091 du 7 juin 2001(Baccalauréat)
- B.O.E.N. n° 20 du 17 mai 2007 (CLES).
- B.O. n° 22 du 7 juin 2007 (DNB).
- JO du 25 août 2005, Titre 1, articles 3 à 7 (Certifications en langues étrangères dans le cadre institutionnel)

Sitographie:

Building a European Bank of Anchor Items for Foreign Language Skills (EBAFLS), (2007), [http://www.cito.com/research\\_and\\_development/participation\\_international\\_research/ebafls.aspx](http://www.cito.com/research_and_development/participation_international_research/ebafls.aspx)

Claire Tardieu est Professeur des universités à la Sorbonne Nouvelle Paris3 (PRES Sorbonne-Paris-Cité) où elle enseigne la didactique de l'anglais dans sa double dimension de recherche et d'enseignement. Elle appartient à l'EA 4398 PRISMES – SESYLIA et a publié un certains nombres d'articles et d'ouvrages dans son domaine. En tant qu'experte pour la France auprès de la DEPP au Ministère de l'Éducation Nationale, elle participe aux évaluations-bilan, et a pris part à deux projets internationaux : *Dutch CEFR Grid* et EBAFLS et coordonné le projet CEF-ESTIM au Centre Européen des Langues Vivantes de Graz.