Évaluer pour mieux enseigner avec SELF FLE

Perceptions croisées

Cristiana Cervini(1), Yannick Hamon(2), Monica Masperi(3)

(1) Università di Bologna, DIT, (2) Università Ca' Foscari, DSLLC, (3) Université Grenoble Alpes, Lidilem

Résumé : L'objectif de notre intervention est de rendre compte de l'expérimentation du dispositif d'évaluation SELF FLE conçu dans le cadre du projet IDEFI Innovalangues. Cette expérimentation menée à l'université Ca' Foscari de Venise en février 2019 concerne environ 100 étudiants répartis en deux groupes correspondant à deux cursus différents (Département d'économie et Département d'études de l'Asie et de l'Afrique Méditerranéenne). Notre intention est de dégager, à partir des résultats obtenus, un lien fort entre l'évaluation à visée diagnostique et formative et la prise en compte, dans les pratiques d'enseignement/apprentissage, de principes clés pouvant guider l'innovation, en tenant compte non seulement du seul niveau de compétence en langue des candidats, mais aussi des leurs profils langagiers (culture d'appartenance, vécu langagier, biographie). Nous entendons analyser les résultats obtenus dans le contexte universitaire italien en les comparant aux résultats obtenus lors des expérimentations effectuées en France, en dégageant des éléments saillants qui permettent d'envisager des innovations dans les pratiques.

Mots-clés : évaluation formative, positionnement en langues, SELF FLE, Université Ca' Foscari (Venise), programme IDEFI (Initiatives d'excellence en formations innovantes), Innovalangues.

1 Introduction

L'intervention que nous proposons s'inscrit dans l'axe de réflexion concernant la place de l'évaluation selon les pratiques d'enseignement-apprentissage en présentiel ou en autonomie guidée.

Nous entendons rendre compte de l'expérimentation du dispositif d'évaluation SELF FLE conçu dans le cadre du projet IDEFI Innovalangues [1]. Cette expérimentation a

1. Innovalangues (`http://innovalangues.fr/`), projet lauréat du programme IDEFI (Initiatives d'excellence en formations innovantes) a permis de soutenir scientifiquement et financièrement la conception et la

eu lieu à l'Université Ca' Foscari de Venise en février 2019 et a impliqué 97 étudiants répartis en deux groupes (38 étudiants inscrits en 2ème année de Licence LICSAAM *Lingue, culture e società dell'Asia e dell'Africa mediterranea* et 59 étudiants inscrits en 2ème année de Licence COMES, *Commercio Estero*). Notre intention est de mettre en exergue, à partir des résultats obtenus, le lien entre l'évaluation à visée diagnostique et formative et la prise en compte, dans les pratiques d'enseignement-apprentissage, de principes clés pouvant guider l'innovation, en tenant compte non seulement du niveau de compétence en langue des candidats, mais aussi de leurs profils langagiers (culture d'appartenance, vécu langagier, biographie).

Comme le souligne Bourguignon (2005), aux traditionnels contrôles notés, s'est substituée une conception holistique de l'évaluation qui intervient avant (évaluation diagnostique), pendant (évaluation formative) et après (évaluation formative) le processus d'apprentissage, revêtant une fonction régulatrice (Bourguignon, 2005 : 460). De plus, les changements de paradigme de la connaissance vers la compétence, de l'activité vers la tâche, la plus grande prise en compte des documents et des tâches authentiques constituent des évolutions qui ont permis d'envisager l'évaluation comme un élément clé de l'apprentissage des langues vivantes.

C'est dans ce contexte que le déploiement de SELF FLE s'est opéré pour la première fois hors de France, précisément à Venise. Contrairement au public FLE en France, qui est très hétérogène, multilingue et pluriculturel, à Venise la langue maternelle des apprenants est majoritairement l'italien. Enfin, le couple de langue italien-français relève des langues apparentées (aire romanophone), ce qui peut constituer un facteur non négligeable dans l'interprétation des données et, le cas échéant, pourrait impliquer des variations dans les résultats des candidats au test.

2 Évaluer les compétences en français par SELF FLE

En décembre 2018, SELF (Système d'Évaluation en Langues à visée Formative), était déployé en six langues : italien (langue pilote), anglais, mandarin, japonais, FLE et espagnol. Le dispositif a désormais atteint un public de plus de 140 000 étudiants en France, accédant ainsi au statut de dispositif d'évaluation à large échelle. Conçu dans le cadre du projet IDEFI Innovalangues (Masperi, 2011), SELF est un test en ligne de positionnement à visée diagnostico-formative, ancré sur un argument de validité (Cervini, Jouannaud, 2015 ; ALTE, 2011). La conception des contenus, notamment des tâches d'évaluation en langue pour trois habiletés (compréhension de l'oral, de l'écrit et expression écrite courte) est basée sur le respect des principes d'authenticité situationnelle et interactionnelle, tandis que la conception d'un système numérique d'édition, d'administration et de *tracking* des résultats, a été soutenue par des processus de validation psychométrique circulaire et dynamique (Cervini, 2016).

La démarche de recherche-action qui est au cœur du projet Innovalangues est fortement ancrée dans les exigences d'enseignement-apprentissage et vise l'innovation sociale, la transformation des pratiques et son accompagnement. Le projet se dis-

réalisation didactique, technopédagogique et informatique du dispositif SELF en six langues. La conception de SELF FLE a été réalisée en collaboration avec l'ADCUEFE (Association des Directeurs de Centres Universitaires d'Études Françaises pour Étrangers).

Published in : Antonella Corvaglia and Françoise Wolf-Mandroux (dir.), *Formation linguistique des apprenants allophones et pédagogies innovantes*, Éditions des archives contemporaines, Coll. «PLIDAM», France, ISBN : 9782813003621,
doi : https ://doi.org/10.17184/eac.9782813003621

tingue également par sa volonté forte de tenir compte du contexte institutionnel et des contraintes de terrain (Masperi, Quintin, 2014).

Dans le domaine de l'évaluation, les tests en langues orientés vers les dimensions diagnostico-formatives sont très rares (De Ketele, 2010), ce que l'on ne peut que regretter. En effet, une évaluation bien conçue peut avoir un impact significatif sur l'autonomisation des apprenants, sur leur sentiment d'efficacité personnelle (Bandura, 2007) et ce, notamment dans des contextes qui promeuvent la formation hybride ou en autonomie tout au long de la vie. Enfin, l'évaluation diagnostique et formative permet aussi, potentiellement, de mettre en place des parcours de remédiation personnalisés. C'est dans cette perspective que nous avons ciblé l'orientation diagnostique de SELF. En phase de conception, le principe d'authenticité règle à la fois le choix des textes (à comprendre ou à compléter) et le type d'interaction qui se crée entre textes, consignes et opérations cognitives sollicitées sur les étudiants par les tâches/items proposés. En guise d'exemple, en expression écrite courte, la lecture d'une carte postale, entre autres, est proposée. La compréhension de la carte postale ne sollicite pas uniquement la dimension informative et factuelle mais également la dimension socio-pragmatique : quel est le rapport entre écrivain et lecteur de la carte postale ? quel est le registre utilisé ? quelle est la visée illocutoire de cette carte ? La dimension socio-pragmatique de la communication est testée dans ce cas par une "question à réponse ouverte courte" où l'on demande au candidat de trouver la formule de clôture la plus efficace et correcte en fonction de la situation de communication. La dimension diagnostique du test est mise en évidence par la description détaillée des facteurs langagiers et textuels impliqués dans la tâche et les items : la focalisation langagière de chaque question (morphosyntaxique, sémantique, socio-pragmatique), la/les opérations cognitives sollicitées, le type d'interaction demandé au candidat.

Du point de vue de la structure, le test SELF est construit sur base d'un algorithme semi-adaptatif, qui comporte deux étapes : une première étape commune à tous les candidats, et une deuxième étape qui varie selon les résultats obtenus dans la première partie. Techniquement, c'est ce qu'on appelle un test multi-stade. Dans sa substance (calibrage des tâches et des items, définition de l'algorithme de progression), SELF FLE est tributaire des résultats issus des pilotages menés sur un public allophone ayant une gamme très étendue de langues sources : romanes, asiatiques, germaniques, sémitiques. Ainsi, les tâches d'évaluation proposées dans le premier stade du test ne portent-elles que sur la compréhension de l'oral et l'expression écrite courte, à savoir les habiletés qui apparaissent comme les plus discriminantes au regard du public allophone visé par le test. Ce choix apparaît d'autant plus pertinent pour le public ciblé par cette expérimentation, que la langue source (l'italien) est une langue typolinguistiquement proche de la langue cible (le français), ce qui réduit le potentiel de discrimination des tâches en compréhension de l'écrit.

L'affichage des résultats est immédiat, juste après la validation des réponses à la fin du test. L'option retenue pour communiquer aux étudiants le niveau de compétence évalué dans la langue testée s'inscrit dans une logique résolument formative, dans la perspective d'un « cap à atteindre », en rupture avec l'idée d'un « score » sommatif (i.e. résultant d'une « somme d'acquis ») souvent privilégié dans les tests de position-

Published in : Antonella Corvaglia and Françoise Wolf-Mandroux (dir.), *Formation linguistique des apprenants allophones et pédagogies innovantes*, Éditions des archives contemporaines, Coll. «PLIDAM», France, ISBN : 9782813003621, doi : https ://doi.org/10.17184/eac.9782813003621

nement. En l'occurrence, la formulation adoptée est « *vous êtes en route vers...* » déclinée selon les différents niveaux de sortie, qui vont de A1.1 (score le plus bas de l'étape 1) à B2.2/C1 (score plus élevé de l'étape 3).

3 Expériences avec SELF FLE à l'Université Ca' Foscari

3.1 Description de l'expérience et des groupes participants

Les données ont été recueillies à l'université Ca' Foscari de Venise en deux temps, auprès de deux groupes différents : un groupe d'étudiants inscrits dans un cursus d'études de langues orientales (LICSAAM, 2ème année de licence), et un groupe d'étudiants qui suivent un cursus économique (COMES, 2ème année de licence) [2]. Outre les données obtenues par la passation du test, d'autres données ont été recueillies au moyen d'un questionnaire soumis aux deux groupes pour évaluer la perception du test par les étudiants et par le biais d'un autre questionnaire adressé aux enseignants en charge des deux cours de français. Nous nous attarderons ultérieurement sur ces deux questionnaires.

Pour ce qui est de l'échantillon testé, le premier groupe est composé de 59 étudiants inscrits en deuxième année de licence de Commerce Extérieur (COMES) qui suivent un cours de français niveau B1 dans le cadre de leur cursus.

Le test a été effectué en salle informatique le 04 février 2019 lors du premier cours du deuxième semestre dans les locaux du campus de Trévise. Lors du premier semestre, les étudiants ont suivi un cours de pratique de la langue assuré par une lectrice de langue maternelle française.

Le deuxième groupe est composé de 38 étudiants inscrits en deuxième année de licence de Langues, Cultures et Sociétés de l'Asie et de l'Afrique Méditerranéenne (LICSAAM). Le test a été effectué en deux temps dans une salle informatique à capacité restreinte dans le campus de San Basilio situé sur la lagune vénitienne. Tout comme le groupe COMES, le groupe LICSAAM a suivi au premier semestre un cours de pratique de la langue. Les enseignements, pour les deux groupes, ont été assurés par la même enseignante. Il est à souligner qu'un faible pourcentage d'étudiants du groupe LICSAAM a commencé l'apprentissage du français en tant que débutants. Au total, 97 étudiants ont effectué le test.

2. Les deux groupes ont été informés du fait que les données obtenues seraient ensuite traitées et analysées à des fins de recherche et ont donné leur accord.

Published in : Antonella Corvaglia and Françoise Wolf-Mandroux (dir.), *Formation linguistique des apprenants allophones et pédagogies innovantes*, Éditions des archives contemporaines, Coll. «PLIDAM», France, ISBN : 9782813003621,
doi : https ://doi.org/10.17184/eac.9782813003621

3.2 Résultats

Sur le plan des résultats globaux (Tableau 1), de prime abord, les résultats issus du test semblent correspondre aux niveaux de compétence en FLE de la classe.

Tableau 1 – Résultats obtenus par les participants à la passation de SELF FLE.

Résultats	Nombre d'étudiants	Temps moyen
En route vers A1.2	1	37
En route vers A2.1	2	25,5
En route vers A2.2	1	26
En route vers B1.1	18	43,38
En route vers B1.2	12	44,33
En route vers B2.1	23	43,13
En route vers B2.2	45	51,48
En route vers B2.2/C1	4	33,25

En effet, comme nous pouvons le constater, les niveaux établis par SELF correspondent au niveau atteint après un semestre de cours puisque le nombre d'étudiants se situant au-delà du niveau B1.1 (84) est supérieur à celui des étudiants en deçà de ce seuil. En outre, le niveau de prérequis des étudiants (A2 minimum) est confirmé par le test puisque la plupart des participants a obtenu un résultat supérieur à A2.2. Cela tend à confirmer la bonne adéquation entre le calibrage du test et le contexte – la situation réelle d'apprentissage – dans lequel le test est administré. De plus, si la gamme des résultats livrés par SELF est conforme aux pré-acquis linguistiques requis par l'Institution et aux niveaux visés, les résultats par compétences montrent que ces derniers sont en adéquation avec ce que nous connaissons des publics italophones. En effet, ces résultats tendent à mettre en évidence la différence entre les habiletés en réception et production (étudiants souvent considérés comme moins forts en production). Si pour les items d'expression écrite courte (EEC), les étudiants COMES testés par SELF obtenant un niveau A2 sont plus nombreux (N = 10) que dans les autres compétences (un seul A2 en compréhension de l'écrit et sept A2 en compréhension de l'oral), cette différence est en fin de compte peu marquée pour le groupe LICSAAM. Par contre, pour les trois compétences, le nombre d'étudiants pour lesquels le test détermine un niveau « en route vers le B2 » est sensiblement le même.

Enfin, par rapport au temps employé en moyenne par les candidats pour terminer le test, nous remarquons que la durée augmente progressivement avec le niveau de difficulté de chaque étape du test. Cependant, nous observons une inversion de tendance pour les quatre étudiants les plus performants qui ont probablement trouvé simples les tâches à résoudre (le résultat obtenu est « en route vers B2.2/C1 », qui correspond au niveau le plus élevé attribué par SELF FLE).

Published in : Antonella Corvaglia and Françoise Wolf-Mandroux (dir.), *Formation linguistique des apprenants allophones et pédagogies innovantes*, Éditions des archives contemporaines, Coll. «PLIDAM», France, ISBN : 9782813003621, doi : https ://doi.org/10.17184/eac.9782813003621

3.3 Perceptions des étudiants

Comme nous l'avons signalé auparavant, un questionnaire a été soumis aux deux groupes qui ont passé le test. Administré sous forme électronique à la fin du semestre, ce questionnaire met la focale sur la valeur formative du test et sur l'adéquation estimée par les deux groupes entre le niveau attribué par le test et le niveau ciblé par le cours de FLE qui leur était dispensé.

De ce point de vue, une correspondance est établie entre le niveau minimum indiqué comme prérequis pour accéder au cours, celui attribué par le test à mi-parcours et le niveau de sortie visé pour les deux groupes.

Une première section du questionnaire s'intéresse aux données personnelles des répondants. Dans la perspective d'une évaluation diagnostique, les données biographiques des candidats présentent un énorme potentiel d'analyse, surtout si l'on dispose d'un très grand nombre de candidats. Cependant, dans cet essai, tous les candidats étaient italophones. C'est pourquoi la variable biographique n'a pas été prise en compte.

La deuxième partie du questionnaire, qui se compose de dix questions, vise à obtenir des données sur le niveau de départ des étudiants pour chacune des compétences. Pour chaque question, il est demandé aux étudiants de dire si une correspondance peut être établie entre le niveau qu'eux-mêmes s'attribuent pour chacune des compétences évaluées et le niveau attribué par le test. Il s'agissait, en d'autres termes, de solliciter une réflexion sur l'autoévaluation, par compétences dissociées, et de comparer cette perception avec les résultats du test.

Enfin, une troisième partie concerne l'appréciation du test lui-même (intérêt des contenus, interaction, interface) à partir de 14 items présentés comme des affirmations à évaluer par une échelle de Likert (de 1-désaccord total à 5-accord total).

Si le nombre de réponses est assez bas (27/59 pour le groupe COMES, 9/38 pour les étudiants du groupe LICSAAM), le questionnaire nous permet tout de même de disposer d'un aperçu de la manière dont le test a été accueilli et de dessiner quelques tendances.

En premier lieu, pour ce qui est de la perception par les étudiants de la justesse des niveaux attribués, nous constatons qu'une grande majorité des sondés (82,3 %) considèrent que le niveau attribué par le test à mi-parcours correspond au niveau requis par l'Institution pour accéder au cours. *A posteriori*, ce niveau a été perçu comme suffisant pour suivre le cours et effectuer les activités proposées au deuxième semestre. Toutefois, si un décalage avait été observé entre les résultats issus de SELF et les exigences institutionnelles (en termes d'inscription aux cours), l'adoption d'un tel test à visée formative permettrait de repenser et de réajuster les niveaux en entrée et en sortie. Nous restituons sous forme de tableau synthétique (Tableau 2), les résultats les plus significatifs. Nous nous efforçons ici de dresser un récapitulatif global sous forme de tableau pour les différents aspects traités en cumulant les réponses positives (plutôt d'accord, accord total) et négatives en signalant que la mention « plutôt d'accord » est toujours largement majoritaire par rapport à la mention « accord total ».

Published in : Antonella Corvaglia and Françoise Wolf-Mandroux (dir.), *Formation linguistique des apprenants allophones et pédagogies innovantes*, Éditions des archives contemporaines, Coll. «PLIDAM», France, ISBN : 9782813003621,
doi : https ://doi.org/10.17184/eac.9782813003621

Tableau 2 – Appréciation du test par les participants.

	Désaccord total, plutôt pas d'accord	**Moyennement d'accord**	**Plutôt/ tout à fait d'accord**
Interface claire et agréable	COMES : 17,6 % LICSAAM : 11,1 %	COMES : 23,5 % LICSAAM : 11,1 %	COMES : 58,8 % LICSAAM : 77,8 %
Durée indicative adaptée	COMES : 11,8 % LICSAAM : 0 %	COMES : 23,5 % LICSAAM : 22,2 %	COMES : 64,7 % LICSAAM : 77,7%
Nature authentique des documents audio	COMES : 0 % LICSAAM : 0 %	COMES : 29,4 % LICSAAM : 44,4 %	COMES : 70,6 % LICSAAM : 55,5 %
Nombre d'écoutes suffisant	COMES : 0 % LICSAAM : 11,1 %	COMES : 23,5 % LICSAAM : 33,3 %	COMES : 76,5 % LICSAAM : 55,5 %
Documents écrits authentiques	COMES : 0 % LICSAAM : 0 %	COMES : 23,5 % LICSAAM : 44,4 %	COMES : 76,4 % LICSAAM : 55,5 %
Longueur raisonnable des documents écrits	COMES : 0 % LICSAAM : 0 %	COMES : 23,5 % LICSAAM : 22,2 %	COMES : 76,5 % LICSAAM : 77,8 %
Ergonomie intuitive pour les items d'EEC	COMES : 11,8 % LICSAAM : 0 %	COMES : 35,3 % LICSAAM : 44,4 %	COMES : 53 % LICSAAM : 55,5 %

L'ensemble des données obtenues concernant l'appréciation du test par les participants tendent à valider les choix de conception qui ont été effectués pour SELF FLE.

En effet, nous remarquons tout d'abord à la lecture des résultats que la correspondance entre les niveaux attribués par le test et la perception préalable du niveau en FLE par les étudiants eux-mêmes est optimale pour les deux groupes (« plutôt d'accord » ou « accord total » pour 90 % des répondants). Qui plus est, à la lecture des résultats extraits par habileté (CO, CE, EEC) cette adéquation entre niveau attribué et représentation du niveau acquis concerne chacune des trois habiletés testées. Par ailleurs, le degré d'appréciation du test est plutôt positif dans les deux groupes.

En ce qui concerne l'interface et les aspects ergonomiques, les résultats sont là aussi globalement positifs pour les deux groupes et cela se voit confirmé par les réponses apportées pour l'ergonomie des items d'EEC.

Au vu des résultats, il semble également que la durée prévue pour les items de compréhension de l'écrit et de compréhension de l'oral soit adaptée.

Enfin, la pertinence et l'intérêt des documents servant de support pour les items de compréhension écrite et orale est très largement soulignée dans les deux groupes.

Toutefois, certains points à améliorer ont été signalés par une faible quantité d'étudiants par rapport à la longueur des questions, le feedback ou la correspondance du résultat avec leur niveau perçu.

Published in : Antonella Corvaglia and Françoise Wolf-Mandroux (dir.), *Formation linguistique des apprenants allophones et pédagogies innovantes*, Éditions des archives contemporaines, Coll. «PLIDAM», France, ISBN : 9782813003621, doi : https ://doi.org/10.17184/eac.9782813003621

3.4 Résultats psychométriques

La psychométrie, discipline statistique appliquée à l'évaluation en sciences humaines, soutient de manière incontournable les démarches des enseignants-concepteurs dans toutes les étapes qui caractérisent le cycle du test (ALTE, 2011), depuis la conception des contenus, jusqu'à l'administration à large échelle. Les procédures à mettre en place sont complexes du point de vue méthodologique et organisationnel. En effet, les didacticiens sont censés d'une part se familiariser avec le domaine généralement peu familier des statistiques ; d'autre part, un bon pilotage requiert l'implication d'un grand nombre de candidats et des passations, dans des conditions de contrainte. Plusieurs modèles de validation existent et SELF se base sur la théorie classique des tests (TCT) et sur la théorie de réponse à l'item. Ces deux moments clés de validation des items sont suivis des sessions de standard setting avec la méthode d'Angoff. Ceci dit, en raison du petit nombre des candidats, les résultats issus des passations de SELF FLE à Ca' Foscari ont pu être analysés uniquement avec la TCT. Ce modèle permet de faire un certain nombre de considérations et, notamment, de vérifier :

— la fidélité du test : l'alpha de Cronbach, coefficient de fidélité, doit être entre 0.6 et 1 (1 représentant une consistance interne parfaite) (ALTE, 2011) ;

— le pouvoir discriminant des items : la valeur de RIR (*biserial correlation Item–Rest total*) doit être > 0.15 ;

— la difficulté des items (p value = pourcentage de réussite à l'item). On considère que toute valeur comprise entre 0.02 (item extrêmement difficile) et 0.98 (item extrêmement facile) est acceptable ;

— l'unidimensionnalité du test par analyse factorielle (AF) [3].

Suite aux déploiements menés à Ca' Foscari, nous avons voulu étayer les données qualitatives recueillies par les questionnaires à l'aide des données statistiques issues des analyses psychométriques et observer, de ce fait, les caractéristiques des contenus et la fiabilité des résultats sur un public différent par rapport à celui auprès duquel le test avait été administré en France et qui avait fourni le contexte de validation du test. Il est nécessaire de préciser que pour l'application du modèle statistique de la TCT, il faut disposer à minima d'un échantillon constitué de 100 candidats. Pour cette raison, les analyses à Ca' Foscari n'ont porté que sur l'étape initiale du test [4], composée de 27 items/questions concernant la compréhension de l'oral et l'expression écrite courte. Les valeurs psychométriques obtenues sont très intéressantes et témoignent de la solidité du test auprès d'un public italophone de non spécialistes. L'alpha de Cronbach est de 0.87, tous les items/questions ont un RIR supérieur à 0.30 (moyenne de 0.49), sauf pour un item à 0.26 (quand même acceptable selon les bornes affichées ci-dessus). Même si les indices de discrimination restent de bonne qualité,

3. « L'AF est une étape essentielle pour la validation du construit. L'analyse factorielle ou les modèles d'équations structurelles permettent de vérifier si les items appliquent le construit prévu. L'analyse factorielle est très utile lors des étapes d'élaboration du test, car elle permet de vérifier que le test ou les spécifications fonctionnent comme prévu. » (ALTE, 2011 : 74)

4. Le premier stade du test est passé par tous les candidats (N = X), qui sont ensuite dirigé vers la deuxième et dernière étape, selon le niveau d'habileté obtenu. Les données issues de la deuxième étape du test, ne concernant que des portions de notre échantillon d'informateurs < 100, ne peuvent être traitées avec le modèle de validation statistique adopté.

Published in : Antonella Corvaglia and Françoise Wolf-Mandroux (dir.), *Formation linguistique des apprenants allophones et pédagogies innovantes*, Éditions des archives contemporaines, Coll. «PLIDAM», France, ISBN : 9782813003621,
doi : https ://doi.org/10.17184/eac.9782813003621

les p values montrent que les items sont traités avec une certaine aisance par un public d'italophones (valeur moyenne de p = 0.81 ; 10 items sur 27 ont un p > 0.90). Si nous observons de plus près le rapport entre indices psychométriques et contenus des items, nous remarquons que la question la plus difficile (p value = 0.42) réfère à l'expression écrite courte et se présente sous la forme d'un QROC (questionnaire à réponse ouverte courte) qui requiert la maîtrise du lexique actif, de la morphosyntaxe et le contrôle de l'orthographe. La question la plus facile (p = 0.98) est une question à choix multiple en compréhension de l'écrit.

Par rapport aux indices de discrimination, l'item le moins discriminant est un QCM (RIR = 26) tandis que le plus discriminant est, encore une fois, un item d'expression écrite courte, cette fois-ci focalisé sur l'usage des connecteurs (RIR = 61).

Lors des premiers pilotages menés en France sur un public d'étudiants allophones, les tâches d'expression écrite courte, constituées pour la plupart de QROC, s'avéraient plus discriminantes et plus difficiles que les tâches de compréhension de l'oral et de l'écrit. Toujours lors des pilotages, les tâches calibrées sur les niveaux B1 et B2 avaient globalement des p values plus bas que les tâches de niveau A1/A2 (elles étaient donc plus difficiles) mais, en même temps, elles avaient des RIR plus faibles, donc elles étaient moins discriminantes. L'explication de cette tendance requiert d'autres expérimentations afin de mener une analyse plus approfondie ; cependant, de manière impressionniste, nous pourrions peut-être faire l'hypothèse que la complexité du construit, pour des tâches de niveau intermédiaire testées sur un public très hétérogène (à la fois en termes d'acquis et de biographie langagière), pourrait impacter la mesure de discrimination des items.

3.5 Perceptions des enseignants

Un questionnaire de dix questions, conçu sur le même format que celui adressé aux étudiants (échelle de Likert de 1 à 5) a été soumis à deux enseignantes de manière à recueillir leurs perceptions sur le test et ses usages.

Les deux enseignantes en question sont des collègues expérimentées [5] qui sont toutes deux amenées à pratiquer l'évaluation sommative dans le cadre des certifications DELF.

Dans l'ensemble, l'utilité d'un test à visée formative est mise en avant, avec une réserve dans l'un des deux cas. À la question « Pensez-vous, en général, qu'un test de positionnement facilite l'action didactique et si oui à quel niveau ? » l'une des enseignantes répond : *« Oui et non. Oui. Il me permettra de cibler les activités en fonction du niveau, si le groupe classe est homogène. Non. Cela pourrait m'amener à revoir mes attentes ou mes objectifs en fonction d'un test qui ne prend en compte que des facteurs performatifs (savoir/ ne pas avoir) ».*

En ce qui concerne la correspondance entre les niveaux attribués aux étudiants par le test et leur niveau tel qu'il est estimé par les collègues, si l'une des collègues le juge surévalué en compréhension, du fait de la transparence entre les deux langues

5. Les collègues en question ont un statut de « Collaboratore Esperto Linguistico » (contractuel CDI ou CDD).

Published in : Antonella Corvaglia and Françoise Wolf-Mandroux (dir.), *Formation linguistique des apprenants allophones et pédagogies innovantes*, Éditions des archives contemporaines, Coll. «PLIDAM», France, ISBN : 9782813003621, doi : https ://doi.org/10.17184/eac.9782813003621

apparentées (« Il me semble que pour des étudiants italiens, le niveau est surévalué. Notamment en compréhension car les langues sont transparentes »), l'autre collègue pense au contraire que le niveau est conforme à ce qu'elle observe chez ses étudiants.

Il n'y a pas non plus unanimité sur la potentielle utilité du test pour des enseignants novices qui pourraient alors s'inspirer de SELF FLE pour alimenter leur réflexion en matière de construction des compétences langagières à partir de la progression observée dans la succession des tâches et des items proposés par SELF FLE. L'une des réponses mentionne « La progression d'un niveau à l'autre n'apparaît pas avec évidence. Par moments j'ai eu l'impression que l'on revenait en arrière par rapport au niveau de difficulté ». Les avis sont également partagés quant à la possibilité que la passation de SELF se révèle être une situation potentiellement acquisitionnelle.

Si un tutoriel a été rendu disponible avant la passation, l'une des collègues note des hésitations sur l'ordre de l'écoute (question à écouter avant ou après l'écoute du document d'input). Une autre remarque porte sur la dimension grammaticale :

> *« L'aspect grammatical est très marginal, le test se base essentiellement sur les compétences communicatives et il arrive (dans les niveaux hauts B2 et plus) que des apprenants aient de bonnes compétences communicatives (acquises par des séjours ou des contacts avec des natifs) sans pour autant avoir de maîtrise grammaticale. Et inversement, pour les apprenants de certains pays où l'enseignement est très notionnel. Il faudrait valoriser les connaissances notionnelles aussi, au moins en partie ».*

Cette remarque est tout à fait intéressante, car elle tend à conforter le choix qui a été fait de neutraliser la perception du focus de la tâche, très clairement identifié et renseigné en phase de conception (syntaxique, lexical, pragmatique…), au profit de la mise en exergue de la compétence communicative.

De plus, cette informatrice évoque les biais éventuels induits par les réponses par élimination [6]. Hormis ces réserves, l'évaluation est positive concernant l'appréciation globale du test (choix des options 4 plutôt d'accord » ou 5 « tout à fait d'accord »), l'authenticité des items de compréhension, le calibrage difficulté/longueur/temps imparti. La variété des items est soulignée, ainsi que leur capacité à être source d'inspiration pour les pratiques de classe, notamment pour la compréhension orale.

4 Conclusions et perspectives

Les perspectives envisagées pour l'avenir sont principalement de deux ordres : la première consiste à poursuivre les expérimentations de SELF FLE hors de France. Il nous semble en effet utile de répéter les passations sur un public italophone plus vaste, pour être en mesure d'appliquer le modèle TCT sur les autres étapes du test, mais également pour être à même d'appliquer la théorie de réponse à l'item (modèle de Rasch). De nouvelles passations pourraient, d'un côté, compléter le cadre que nous venons d'esquisser par ce premier tour d'analyses et les questionnaires, de l'autre

6. Cette réserve est assez fréquente chez les enseignants qui émettent des doutes quant à l'efficacité des items à choix multiples.

Published in : Antonella Corvaglia and Françoise Wolf-Mandroux (dir.), *Formation linguistique des apprenants allophones et pédagogies innovantes*, Éditions des archives contemporaines, Coll. «PLIDAM», France, ISBN : 9782813003621,
doi : https ://doi.org/10.17184/eac.9782813003621

côté, nous permettre d'estimer, le cas échéant, de possibles adaptations des contenus pour un public cible d'italophones. Le deuxième ordre concerne une nouvelle phase de design, dans le cadre d'un nouveau projet [7] qui se donne comme objectif la conception d'un SELF FLE international. SELF FLE international prendra clairement en compte les variétés de français du monde francophone. Il sera de ce fait conçu en utilisant des sources authentiques de la francophonie en Europe (Luxembourg, Belgique, Suisse, Val d'Aoste), voire du continent africain et du Québec. A ce sujet, nous soulignons l'importance de passer par des validations menées, tout au long du processus de création du test, par des institutions à caractère multilingue, traditionnellement tournée vers la valorisation de la diversité linguistique. Ce qui devrait permettre in fine de se confronter non seulement à un panel de ressources plus riche et varié mais également à une manière sans doute différente de concevoir l'approche à la langue, sensiblement enrichie par les apports exogènes. Cette nouvelle version du test relance nous permettra également d'approfondir la dimension diagnostique des compétences en langues, pour que le dispositif SELF puisse devenir un véritable levier vers des parcours d'apprentissage plus adaptés au profil et aux exigences de chaque étudiant.

Bibliographie

ALTE, *Manuel pour l'élaboration et la passation de tests et d'examens de langue. Division des Politiques linguistiques*, Conseil de l'Europe, DG II – Service de l'éducation, Strasbourg, 2011.

Bandura, Albert, *Auto-efficacité. Le sentiment d'efficacité personnelle,* 2e édition, De Boeck Université, 2007.

Bourguignon, Claire ; Delahaye, Philippe ; Vicher, Anne (2005) « L'évaluation de la compétence en langue : un objectif commun pour des publics différents », *Éla. Études de linguistique appliquée*, vol. 140, No. 4, p. 459-473. doi : 10.3917/ela.140.0459.

Cervini, Cristiana (2016) « Approcci integrati nel testing linguistico : esperienze di progettazione e validazione in prospettiva interlinguistica », *in* Cervini, Cristiana (dir.), *Interdisciplinarità e apprendimento linguistico nei nuovi contesti formativi. L'apprendente di lingue tra tradizione e innovazione*, Bologna, Quaderni del CESLIC. Dip. LILEC. doi : http://doi.org/10.6092/unibo/amsacta/5069.

Cervini, Cristiana ; Jouannaud, Marie-Pierre (2015) « Ouvertures et tensions liées à la conception d'un système d'évaluation numérique multilingue en ligne dans une perspective communicative et actionnelle », *ALSIC – Apprentissage des langues et systèmes d'information et de communication.* Numéro spécial 'Des machines et des langues', vol. 18, No. 2. doi : https://doi.org/10.4000.

Cervini, Cristiana ; Masperi, Monica (2021) « Designing a Multilingual Large-scale Placement Test with a Formative Perspective : A Case Study at the University of Grenoble Alpes » *in* Betty Lanteigne, Christine Coombe, James Dean Brown (dir.), *Challenges in Language Testing Around the World – Insights for Language Test Users*, Springer Publishing.

De Ketele, Jean-Marie (2010) « Ne pas se tromper d'évaluation », *Revue française de linguistique appliquée*, vol. 15, No. 1. p. 25-37. doi : 10.3917/rfla.151.002515

Masperi, Monica ; Quintin Jean-Jacques (2014) « L'innovation selon Innovalangues », *LEND – Lingua e Nuova Didattica*, vol. 1, p. 6-14, Milano.

Masperi, Monica (2011) Innovalangues : Innovation et transformation des pratiques de l'enseignement-apprentissage des langues dans l'enseignement supérieur. hal-02004250.

7. Erasmus+, Partenariat Stratégique, qui rassemble cinq différentes universités européennes.

Published in : Antonella Corvaglia and Françoise Wolf-Mandroux (dir.), *Formation linguistique des apprenants allophones et pédagogies innovantes*, Éditions des archives contemporaines, Coll. «PLIDAM», France, ISBN : 9782813003621, doi : https ://doi.org/10.17184/eac.9782813003621