



HAL
open science

L'évaluation prédictive en FLE à Grenoble École de Management : conception d'un dispositif hybride de positionnement

Taylor Roholt

► **To cite this version:**

Taylor Roholt. L'évaluation prédictive en FLE à Grenoble École de Management : conception d'un dispositif hybride de positionnement. Sciences de l'Homme et Société. 2015. dumas-01213446

HAL Id: dumas-01213446

<https://dumas.ccsd.cnrs.fr/dumas-01213446>

Submitted on 8 Oct 2015

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.



**L'évaluation prédictive en FLE à Grenoble École de
Management :
conception d'un dispositif hybride de positionnement**

**ROHOLT
Taylor**

Sous la direction de Madame Stéphanie GALLIGANI

UFR LLASIC
Département Sciences du langage

Mémoire de Master 2 professionnel
Spécialité : Français Langue Étrangère
Année universitaire 2014-2015



L'évaluation prédictive en FLE à Grenoble École de Management : conception d'un dispositif hybride de positionnement

**ROHOLT
Taylor**

Sous la direction de Madame Stéphanie GALLIGANI

UFR LLASIC
Département Sciences du langage

Mémoire de Master 2 professionnel
Spécialité : Français Langue Étrangère
Année universitaire 2014-2015

Remerciements

Dans le cadre de ce mémoire, je tiens tout d'abord à remercier Madame Stéphanie GALLIGANI de son aide fondamentale lors de ce travail. Je lui exprime ma reconnaissance profonde pour ses conseils et ses directives.

Mes remerciements vont également à Madame Carole GALLY, coordinatrice du programme FLE à Grenoble École de Management, de m'avoir accueilli et suivi tout au long de mon stage. Son soutien et ses encouragements m'ont été chers durant cette expérience.

Je souhaite également exprimer ma gratitude sincère à Madame Christine BORTOT pour son aide vitale lors de la rédaction de ce mémoire.

Enfin, je remercie mes collègues, mes amis et ma famille pour leur soutien moral durant ce travail.

DECLARATION

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

NOM : Roholt PRENOM : Taylor

DATE : 18/5/2015 SIGNATURE : 

Sommaire

Introduction	6
PARTIE 1. PRÉSENTATION DU CONTEXTE	9
1. L'ORGANISME D'ACCUEIL.....	9
2. LES COURS DE LANGUE, ET SPÉCIFIQUEMENT DE FLE, À GEM	10
3. LES ACTEURS	11
4. CONCLUSIONS SUR LE CONTEXTE ET LE PUBLIC	14
5. LA COMMANDE	15
PARTIE 2 : L'ÉTAT DE L'ART.....	18
1. L'ACTE D'ÉVALUER EN LANGUES : LES DESTINATAIRES ET LES FINALITÉS.....	18
2. LES DIFFÉRENTS TYPES D'ÉVALUATIONS ET LEURS FINALITÉS	20
3. LES QUATRE ÉTAPES DU PROCESSUS ÉVALUATIF ET LES ENJEUX ASSOCIÉS	30
4. LES ITEMS POSSIBLES DANS UN TEST AUTOCORRECTIF.....	41
5. L'HISTOIRE DE L'ÉVALUATION EN LANGUE : DE 1950 À AUJOURD'HUI.....	45
PARTIE 3. LA MISE EN PLACE D'UN DISPOSITIF HYBRIDE DE POSITIONNEMENT À GEM	52
1. L'ÉLABORATION DU NOUVEAU TEST DE POSITIONNEMENT	52
2. LA CONCEPTION DE LA FICHE POUR LE TEST ORAL	85
3. LA CONCEPTION DE L'AUTO-ÉVALUATION	90
Conclusion.....	93

Introduction

De nos jours, l'évaluation en langue occupe une place de plus en plus importante dans la classe de Français Langue Étrangère (désormais, FLE), d'autant plus que les finalités et les formes de ces épreuves se sont multipliées lors de ces trente dernières années. Avec la publication du CECRL en 2001, un éventail étendu d'examens en langue est désormais disponible aux enseignants de langue. Cet ouvrage présente en effet treize pratiques évaluatives qui permettent de sélectionner l'examen le plus pertinent pour la situation d'apprentissage en question.

Cependant, les évaluations décrites dans le CECRL ne sont pas les seules qui sont émergées ces dernières années. La naissance du Diplôme d'Études en Langue Française (DELFL) et du Diplôme Approfondi de Langue Française (DALFL) dans les années 1980 et 1990 montre encore une fois que ce sujet a subi dernièrement des évolutions importantes. Les épreuves certificatives comme celles-ci sont en effet de plus en plus nombreuses, en partie à cause de la nécessité de justifier son niveau de langue pour, par exemple, travailler, effectuer des études supérieures ou bien s'installer dans un pays francophone. Le DELFL et le DALFL ne sont toutefois que deux exemples. La diversité des évaluations en langue est de fait beaucoup plus étendue si nous prenons en compte d'autres épreuves comme le Test de Connaissances du Français (TCF), le Test d'Évaluation de Français (TEF) et le Test de Français International de Naturalisation (TFI). L'existence de ces examens prouve la forte demande d'évaluations en langue aujourd'hui.

Au vu de la diversification des types d'examens linguistiques et de leur nécessité, les enseignants et les directeurs de programmes de FLE doivent désormais prêter une grande attention à la conception et à la mise en place d'épreuves en langue. Un tel soin assure en effet la pertinence, l'efficacité et l'équité de celles-ci dans le contexte donné, quel que soit leur objectif. Il en est de même pour une évaluation dite test de placement ou de positionnement qui sert à déterminer le niveau de langue des apprenants. Le besoin de ce type d'examen est sans doute issu, entre autres, du nombre important d'apprenants de la langue française qui était de 960.000 en 2011 (Site web du journal Le Figaro, s.d.). En considérant ce chiffre, il n'est pas étonnant que l'élaboration de tests de positionnement soit indispensable dans les structures qui proposent des cours de FLE afin d'organiser et de gérer efficacement ces derniers. De surcroît, la qualité d'un tel test en terme de capacité à cerner le niveau réel des apprenants en français est fortement appréciée car elle favorise le bon déroulement des cours et, ainsi, un apprentissage fructueux.

Concernant le terrain de ce stage de master 2 FLE professionnel, il était tout d'abord intéressant de réfléchir sur les types de structures où il était possible de réaliser cette expérience. Il aurait été fort possible qu'il s'agisse d'une institution dont l'unique objectif et l'enseignement du

français, telle que l'Alliance Française, étant donné le nombre important d'établissements de ce type dans le monde qui prennent régulièrement des stagiaires ayant pour vocation à enseigner le FLE.

Je me suis cependant penché sur l'idée d'entreprendre ce projet de master 2 dans une école autre que celle qui ne propose que des cours de français. J'ai ainsi découvert que Grenoble École de Management possède un programme de FLE. En effet, il se révèle intéressant de constater la manière dont une institution comme celle-ci fonctionne en termes de déroulement et de gestion des cours étant donné qu'il ne s'agit pas d'une école de langue. De même, analyser le type de public qui y étudie et comment ce dernier peut différer de celui présent à l'Alliance Française est enrichissant pour un futur enseignant de FLE. Cela permet en effet de prendre conscience des éventuelles différences (les types d'apprenants et les cours, par exemple) auxquelles nous pouvons être confrontés dans ce métier.

Dans le cadre de mes études à l'Université de Grenoble, j'ai eu l'opportunité d'observer des cours de FLE à l'Alliance Française de Grenoble ainsi qu'au Centre Universitaire d'Études Française de Grenoble (désormais, CUEF). De fait, ces expériences m'ont amené à découvrir les caractéristiques de ce type de structure. Leur public est divers quant à ses caractéristiques (âge, origine), sa biographie éducative et ses objectifs d'apprentissage du français. Qui plus est, ces écoles de français proposent des cours très variés quant à la durée et aux objectifs qu'ils visent à atteindre.

Ainsi, pouvoir réaliser une mission pédagogique au sein de Grenoble École de Management m'a permis non seulement de constater le fonctionnement de cette école, mais également de la comparer avec celles évoquées ci-dessus. Qu'est-ce qui différencie cette institution d'une école de langue française concernant l'organisation et le déroulement des cours, ainsi que le public qui y étudie ? De plus, étant donné que la mission du stage se focalisait sur l'évaluation en langue, comment ces épreuves diffèrent-elles selon la nature de l'institution ?

En prenant en compte le terrain de stage, ainsi que l'enjeu que représente l'évaluation en langues actuellement et, plus particulièrement, le test de positionnement, la question suivante se pose : comment élaborer un dispositif de positionnement hybride et en français pour des étudiants étrangers inscrits en école de management ? Cette question est complexe dans le sens où, dans l'optique d'y répondre dans son intégralité, il faut choisir les compétences à évaluer, les types d'items à utiliser et, enfin, les spécificités des trois composantes du dispositif. Ce dernier inclut, d'ailleurs, un test de positionnement autocorrectif, une fiche pour un test oral et une auto-évaluation.

Lors de cette expérience, je chercherai à confirmer ou à infirmer certaines hypothèses en lien avec le montage du projet. Tout d'abord, je voudrais montrer que l'utilisation d'un éventail d'outils évaluatifs variés permet de cerner le vrai niveau des étudiants. Puis, pour que les composantes de ce dispositif soient valides et fiables, il faut qu'elles soient basées sur les objectifs d'apprentissage par niveau d'un seul référentiel de langue. Enfin, je souhaite démontrer que le profil du public influence le

choix des spécificités des composantes du dispositif, à savoir les items, les compétences évaluées, l'organisation et la passation.

Dans ce mémoire, je vais commencer par mettre en évidence l'organisation de Grenoble École de Management pour souligner des caractéristiques de cette structure. Je parlerai ensuite de l'état des théories relatives à l'évaluation en langues, une étape indispensable pour connaître les aspects et les enjeux de ce domaine. Une analyse du processus de positionnement existant de l'école suivra, débouchant finalement sur l'explication de la conception du nouveau dispositif hybride de placement.

Partie 1. Présentation du contexte

1. *L'organisme d'accueil*

1.1. *Histoire et philosophie*

Grenoble École de Management (désormais « GEM ») a vu le jour grâce à la Chambre de Commerce et d'Industrie de Grenoble qui l'a fondée en 1984. Depuis cette date, GEM a continuellement évolué en créant de nouveaux programmes et en se développant à travers la construction des campus antennes. De même, cette institution poursuit sans cesse la création de nouvelles relations avec d'autres écoles de management et universités, en France et à l'étranger, pour favoriser les échanges académiques.

GEM figure parmi les grandes écoles de commerce et de management françaises et internationales et s'insère, de surcroît, parmi les six meilleures écoles de commerce de France, les 20 meilleures d'Europe et les 50 meilleures du monde (Site web de GEM, s.d.¹). De plus, GEM dispose de trois accréditations reconnaissant la qualité de son enseignement : EQUIS (European Quality Improvement System), AACSB (The Association to Advance Collegiate Schools of Business) et AMBA (The Association of MBAs). Il s'agit donc d'une institution académique connue en France comme à l'étranger.

GEM revendique comme valeurs principales l'excellence, l'entrepreneuriat et l'engagement. Elle accorde en outre une grande importance aux technologies, et se dit innovatrice dans ce domaine. De même, elle a pour but de favoriser l'expertise du management de la technologie et de l'innovation et l'entrepreneuriat. Elle privilégie également le respect des pratiques responsables en entreprise. En ce qui concerne ses missions principales, GEM vise le développement de relations de collaborations avec des entreprises afin de leur présenter de jeunes diplômés, formés et compétents pour faire face aux enjeux économiques du présent et de l'avenir. Par là-même, grâce à ces relations, l'école facilite l'insertion des étudiants sur le marché de travail.

1.2. *Organisation*

GEM est divisée en quatre écoles principales parmi lesquelles figurent de nombreux programmes. Il s'agit en effet de Grenoble Graduate School of Business (désormais, GGSB), l'École Supérieure de Commerce (désormais, ESC), l'École de Management de Systèmes d'Informations (désormais, EMSI) et enfin l'École Doctorale. Quant aux différents programmes, GEM en offre plus de quarante, mettant en évidence la grande diversité de cursus mise à disposition des étudiants. Ces derniers sont d'origine internationale et française, et viennent étudier à GEM afin d'obtenir une

¹ <http://www.grenoble-em.com/qui-sommes-nous> (consulté le 20/2/2015)

licence ou un master, d'effectuer un échange universitaire (l'école est en partenariat avec 160 institutions en France et à l'étranger), d'obtenir un mastère spécialisé ou bien de faire une thèse. Il existe également des formations continues au sein de l'école. Quant aux spécialisations, nous pouvons citer, entre autres, le management de technologie et d'innovation, la géopolitique, la finance, l'entrepreneuriat, le marketing, le développement de business, les systèmes d'informations, etc.

L'école est par ailleurs une structure privée. Elle est donc payante et accessible par un concours, par la réussite à un test standardisé centré sur le business (Graduate Management Admission Test (GMAT), par exemple), ou par un programme d'échange avec une autre université.

Enfin, GEM possède deux campus antennes, un à Paris et un à Londres. Concernant le premier, l'école le partage avec l'École de Management Normandie. Inauguré en septembre 2013, ce campus a permis d'accueillir des étudiants internationaux souhaitant effectuer leurs études à la capitale, de se rapprocher des entreprises installées à Paris et aux alentours et, enfin, de proposer aux étudiants en alternance à Paris un espace plus proche de leur lieu de travail (Site de école2commerce, s.d.²). Quant au campus à Londres, il donne la possibilité aux étudiants d'étudier dans un pays anglophone, et d'accueillir plus d'étudiants.

2. Les cours de langue, et spécifiquement de FLE, à GEM

GEM possède de nombreux programmes de langues étrangères. Treize langues y sont en effet enseignées. Bien que la majorité des cours soit en anglais, l'école reconnaît l'importance de la maîtrise d'une langue étrangère dans le milieu professionnel. C'est pour cela que tous les étudiants, y compris les français natifs, sont obligés de suivre des cours de langue afin d'obtenir leur diplôme, mis à part les étudiants des programmes MSc³ et MBA⁴ pour qui le module de langue est facultatif.

Compte tenu du grand nombre de programmes, il n'est pas étonnant que cette diversité se manifeste également dans les cours de FLE. En effet, selon le cursus ou le programme, les cours de français se différencient. Les étudiants à ESC dans le programme *English Track*, par exemple, ont une heure et demie de cours de français par semaine, et doivent valider uniquement l'oral à la fin du cursus. En revanche, les étudiants à GGSB ont trois heures de cours par semaine et doivent valider l'oral et l'écrit à la fin de l'année. Il existe également les programmes *Password* et *Double Diplôme* dont les cours de français ont lieu une fois par semaine pendant trois heures où les étudiants doivent valider uniquement l'oral.

² <http://www.ecoles2commerce.com/1243-un-campus-a-paris-pour-grenoble-ecole-de-management> (consulté le 19/2/2015)

³ Masters of Science

⁴ Master of Business Administration

Bien que cette structure compte plusieurs écoles et programmes, je ne travaille qu'au sein de GGSB et de ESC. En effet, ce sont les étudiants non francophones effectuant leurs études dans ces deux écoles qui sont concernés par les cours de FLE.

3. Les acteurs

3.1. Le personnel

GEM est certes très complexe au niveau du nombre de programmes et des niveaux d'études. Il est toutefois possible de tracer les grandes lignes de sa hiérarchie. Depuis 2011, Monsieur Loïc Roche est directeur de GEM. Par rapport aux langues étrangères, Monsieur Abdelkrim Belkheiri est à la tête du département des langues vivantes et cultures étrangères (désormais, LVCE). Quant à ma tutrice de stage, Madame Carole Gally, elle coordonne l'organisation de toutes les langues étrangères à GEM et gère celle des cours de FLE à GGSB.

Nous pouvons compter une quinzaine de professeurs de FLE à l'école. Il s'agit en général de professeurs expérimentés et ayant enseigné à l'étranger avant de revenir en France pour poursuivre leur carrière. Ils sont de même diplômés d'une maîtrise ou d'un master FLE et, parfois, un Phd⁵. La majorité travaille à temps partiel et la plupart assurent des cours dans d'autres établissements à Grenoble, notamment au CUEF ou à l'Alliance Française de Grenoble. Enfin, hormis deux enseignants, les professeurs sont natifs.

3.2. Le public d'apprenants

Sur une année scolaire, environ 6 500 étudiants étudient à GEM dont 2 500 étrangers. Ces derniers représentent près de 120 nationalités différentes. De même, la diversité linguistique du public d'apprenants est remarquable ; 20 langues étaient, en effet, dénombrées lors de la rentrée de janvier 2015. Nous pouvons citer, entre autres, la plupart des langues romanes (mis à part le roumain), le chinois, le hindi, le néerlandais, l'allemand, le serbe, l'anglais et le kannada.

Étant donné cette diversité des langues et des nationalités, il est important de noter la spécificité du public qui suit les cours de FLE à GEM. En effet, il s'agit d'un public plurilingue car ils parlent tous une langue maternelle et l'anglais, et dont l'apprentissage du français avant d'arriver à Grenoble varie. Par là-même, certains étudiaient le français juste avant d'arriver à Grenoble, certains l'ont étudié auparavant avant d'arrêter pendant un certain temps, et d'autres n'ont jamais fait d'études de français.

Toujours concernant le public d'apprenants, il est important d'évoquer la différence du nombre d'étudiants entre ESC et GGSB. De fait, les étudiants de français à GGSB sont plus

⁵L'équivalent américain d'un doctorat.

nombreux qu'à ESC, ce qui fait que les premiers constituent le groupe principal auquel je m'adresse dans le cadre de mon projet. Cela dit, les étudiants à ESC sont également à prendre en compte, mais leurs caractéristiques risquent de moins influencer la mission du stage. Il s'agit donc de prendre en compte ces faits afin d'être familiarisé avec les caractéristiques du public et de savoir sur quel contexte je dois me concentrer pour que mon travail y soit le mieux adapté.

Enfin, nous remarquons une différence du nombre d'étudiants entre la rentrée de septembre et celle de janvier. En septembre, le département de FLE accueille environ 300 nouveaux étudiants, tandis qu'en janvier, ce chiffre n'atteint qu'environ quatre-vingt-dix. De ce fait, la rentrée de septembre représente un enjeu plus important pour le département quant à l'accueil et la gestion des nouveaux étudiants, notamment leurs placements dans les différents groupes de niveau.

Quant aux motivations, aux besoins, aux objectifs et aux attentes des apprenants, ceux-ci varient selon le programme. Pour les étudiants à ESC, nous pouvons citer l'obtention du Diplôme d'Études Supérieures en Management (DESMA), qui est un diplôme de niveau master, et donc équivalent à un bac + 5. Afin d'obtenir ce diplôme, les candidats doivent être capables de réaliser une présentation de vingt minutes en français sur un sujet d'actualité. Ils ont cinq semaines pour préparer cet exposé. Ils doivent évidemment aussi valider les autres cours du cursus.

Quant aux étudiants à GGSB, ils doivent réussir l'examen final qui évalue les quatre compétences langagières principales : compréhension orale et écrite, et production orale et écrite. La note globale pour le cours est prise en compte lors du calcul de la moyenne. La réussite des cours de français représente ainsi un enjeu important pour les apprenants à ESC et à GGSB.

Quant aux cours de français, bien que ce dernier ne soit pas leur matière principale, les étudiants se montrent généralement motivés, ce que j'ai appris auprès des professeurs. D'ailleurs, même s'ils ne possèdent pas les compétences requises pour changer de niveau, ils souhaitent souvent changer de groupe de niveau afin de progresser plus rapidement en français. Cela met en évidence de nouveau leur motivation pour apprendre cette langue. La raison de cette grande volonté vient sans doute de leur envie de profiter du séjour en France pour découvrir la culture et la langue, et pour pouvoir se débrouiller aisément dans la sphère publique. De plus, étant donné le coût de l'école, qui peut aller de 9.000 et 25.000 euros par an, l'on peut croire que les étudiants souhaitent tirer le plus de profit possible de leurs études à GEM.

Enfin, étant donné que ce sont des étudiants qui ont choisi de venir étudier à Grenoble et donc en France, il n'est pas surprenant que beaucoup choisissent le français comme langue étrangère. Il s'agit en effet d'un choix logique d'un point de vue pratique, car cette langue leur permettra de profiter au mieux de leur séjour à Grenoble.

Afin de connaître de manière plus précise les attentes, les besoins et les motivations des apprenants à GEM vis-à-vis des cours de français, j'ai élaboré un questionnaire sollicitant ces informations. Mon objectif était en effet de savoir si leurs motivations étaient d'ordre professionnel, personnel ou autre, ainsi que ce qu'ils souhaitaient retirer des cours de français. De plus, je souhaitais connaître leurs besoins quant à l'apprentissage du français, c'est-à-dire quelle(s) compétence(s) ils voulaient améliorer et dans quels buts. Quant aux destinataires du questionnaire, je l'ai fait remplir uniquement par des étudiants à GGSB puisque, comme je l'ai déjà évoqué, ces étudiants sont les plus nombreux et constituent le public principal auquel le test est destiné.

Ces données sont importantes dans le cadre de mon projet, car elles me permettent de connaître les caractéristiques du public auquel je m'adresse. Par ailleurs, ces caractéristiques peuvent par la suite influencer la nature du test de positionnement. Savoir quelle importance les étudiants accordent au cours de français, ainsi que leurs motivations, ou non, pour apprendre cette langue, m'ont dévoilé des informations utiles sur leurs caractéristiques. De plus, étant donné que le public est extrêmement varié dans ses origines et son éducation, je tenais à savoir si, malgré ses différences, ils avaient la même attitude vis-à-vis des cours de français.

En regardant les données de ce questionnaire (*cf.* Annexe 1 [p. 4]), il est possible de confirmer ou d'infirmes mes hypothèses sur les motivations, les besoins et les attentes des apprenants par rapport aux cours de français. La première question, par exemple, montre que pour la majorité des étudiants à GGSB, l'apprentissage du français n'est pas une motivation importante pour venir étudier à GEM. Nous pouvons supposer que cela est dû au fait qu'ils viennent surtout pour étudier le management. Cependant, en examinant les résultats de la question 4, nous remarquons que la majorité des apprenants auraient choisi de suivre des cours de français même s'ils n'y avaient pas été obligés. Cela contredit les données de la première question et montre qu'il s'agit effectivement d'apprenants captifs (c'est-à-dire, ils obligés de suivre des cours de FLE) mais toutefois motivés pour apprendre le français.

Quant à la question 2, elle met en évidence que les cours de français sont, pour la plupart des étudiants, moins importants que leurs autres cours, mais un grand nombre (43%) y accorde cependant de l'importance. Enfin, les résultats issus de la troisième question démontrent qu'en général les étudiants à GGSB ne comptent pas utiliser la langue française dans leur profession, et que celle-ci représente donc plutôt un outil de communication. Ce constat est également confirmé en regardant les réponses à la question 6 qui montre que la majorité des étudiants souhaite surtout pouvoir utiliser le français pour communiquer avec des francophones, même si certains veulent utiliser le français à des fins professionnelles (*cf.* Annexe 1 [p. 7], question 6, réponse 8 ; [p. 8] question 7, réponse 7). J'ai appris en outre qu'en terme d'apprentissage du français, les besoins des étudiants sont variés et recouvrent toutes les compétences (*cf.* Annexe 1 [p. 6], question 5).

En résumé, en dépit de l'obligation de suivre des cours de FLE à GEM, les apprenants restent généralement motivés, et un nombre important aurait choisi d'étudier le français de toute façon. Néanmoins, nous remarquons, à travers les questions 5, 6, 7, que beaucoup d'étudiants estiment le nombre d'heures consacrées aux cours de français insuffisant pour acquérir des compétences en français. De même, ils se disent trop occupés pour pouvoir se concentrer sur leurs études françaises à cause des autres cours à GGSB. Certains risquent donc de perdre la motivation d'apprendre le français (cf. Annexe 1 [p. 7], question 6, réponse 6).

4. Conclusions sur le contexte et le public

En analysant le contexte du stage, ainsi que le public avec qui je travaille, je peux tirer quelques conclusions importantes qui influenceront la réalisation de mon projet. Tout d'abord, il est impératif de retenir que GEM n'est pas une école de langue. De ce fait, les apprenants n'ont pas les mêmes besoins ni les mêmes motivations que ceux qui intègrent des structures dont l'unique enjeu est l'enseignement de la langue française telles que l'Alliance Française ou un institut français. Elle se différencie également d'une université française publique qui, quant à elle, recouvre un plus large éventail de matières. Autrement dit, elle propose un choix de filières beaucoup plus divers que GEM qui se focalise sur le management. De même, les différents cours de FLE à GEM ne comptent pas le même nombre d'heures ni d'apprenants, et leurs objectifs peuvent varier selon le programme.

En ce qui concerne le public d'apprenants, celui-ci est particulier et ses caractéristiques sont importantes à prendre en compte avant d'entreprendre ma mission. L'objectif primordial des étudiants n'est en effet pas de suivre des cours de français puisqu'ils sont venus dans l'optique d'étudier le management. Ainsi, ce ne sont pas des spécialistes de langues, et ils ne comptent pas réaliser des études supérieures en langue française.

Étant donné en outre l'obligation de suivre un cours de langue, il s'agit d'apprenants captifs mais néanmoins fortement motivés. Ce constat est compréhensible compte tenu de l'impact positif ou négatif des notes du cours de français sur l'achèvement de leur parcours à GEM. L'enjeu important de la réussite de ces cours pour les étudiants est ainsi démontré.

En résumé, nous remarquons que l'école, et surtout le programme de FLE, comprend presque tous les cas de figures concernant, d'une part le type de public, ses origines, ses motivations, ses besoins et ses objectifs et, d'autre part, les cursus d'études, les différents programmes et diplômes et l'organisation des cours de français. Ce sont des faits qu'il faudra impérativement prendre en compte lors de la réalisation du projet de stage afin que le produit final soit adapté à l'école et au public. Par là même, il sera nécessaire de considérer les attentes et les besoins des enseignants de français vis-à-

vis du test de positionnement, ainsi que les ressources nécessaires en termes de finances et d'outil pédagogiques, lors de la conception et de la mise en place de mon projet⁶.

5. La commande

Ce stage, effectué dans le cadre du master 2 FLE professionnel, est basé sur un projet d'ingénierie, c'est-à-dire une mission de conception d'un produit pédagogique pour l'institution d'accueil. Ainsi, une commande de la part de GEM, et plus particulièrement de Madame Gally et du département du FLE, est le socle de cette expérience. C'est en effet grâce à cette commande que j'ai pu organiser mon travail pour satisfaire un besoin de GEM en élaborant un outil pédagogique dont se servira l'école. En l'occurrence, il s'agit de la conception d'un dispositif hybride de positionnement en français.

Dans cette partie, j'évoquerai, dans un premier temps, l'origine de la commande. J'explicitai ensuite la première formulation de celle-ci. Enfin, l'explication de la commande finale suivra.

5.1. L'origine de la commande

Cette commande vient principalement de Madame Gally qui, au cours des dernières années, a constaté des problèmes avec le test existant au niveau du fond, de la forme, de la passation et de la notation. De fait, le test actuel nécessite une amélioration, d'une part pour le rendre davantage valide et fiable et, d'autre part, pour réduire son coût pour l'école. En effet, la passation et la correction du test actuel représentent un travail pour le département de FLE à la fois fastidieux et onéreux. Il était par ailleurs nécessaire de consacrer à ces deux tâches un temps important étant donné le nombre d'étudiants accueillis dans le département de FLE, surtout à la rentrée de septembre. Faire passer un test oral, ainsi qu'un test écrit, à près de 300 étudiants fait l'objet d'un travail véritablement chronophage et, de ce fait, extrêmement dispendieux pour l'école. En effet, le nombre d'heures nécessaires pour faire passer et pour corriger le test en septembre s'élève jusqu'à 150 heures rémunérées à vingt-cinq euros chacune, et en janvier ce chiffre monte à quarante heures rémunérées. Le test actuel coûte donc environ 4 700 euros par an à l'école, sans compter le coût des photocopies nécessaires. En prenant conscience de ces chiffres, nous ne nous étonnons pas que la motivation principale pour me confier cette mission soit la réduction des heures nécessaires à ce travail, et donc le coût de ce test pour GEM.

Cependant, ce ne sont pas que la passation et la correction qui ont motivé la commande. Madame Gally a également noté que le test ne fournissait pas toujours une image juste du niveau de

⁶ Les attentes et les besoins des enseignants seront évoqués dans la troisième partie de ce mémoire.

l'étudiant, ce qui provoquait des changements de groupe parfois problématiques pour le département⁷. En effectuant un changement de niveau, de fait, il faut trouver un groupe qui n'est pas complet et dans lequel l'étudiant puisse être intégré, ce qui n'est pas toujours évident. Ainsi, la conception d'un nouveau test vise également la réduction des changements de groupes. Il est ici question du fond, c'est-à-dire de sa pertinence au niveau linguistique et thématique, et de la forme du test, à savoir son organisation, les consignes et la rédaction d'items. Les questions qui se posent sont, entre autres, le contenu linguistique et les thématiques correspondent-ils aux niveaux donnés ? Est-ce que le test évalue ce qu'il est censé évaluer en restant fidèle au niveau ciblé ?

L'ensemble des professeurs se dit plutôt satisfaits du test écrit, bien que certains aient pu ponctuellement évoquer quelques lacunes⁸. De même, l'école n'a jamais interpellé le département de français par rapport au coût du test. Aussi pouvons-nous conclure que la commande est venue principalement de Madame Gally.

Pour clore la présentation de la commande, il est important de souligner que celle-ci concerne à la fois une question d'ingénierie politique (réduire le coût du test en diminuant le nombre d'heures nécessaire) et pédagogique (à savoir la conception d'un test de positionnement en assurant l'utilité pour l'institution).

5.2. La première formulation de la commande

L'articulation de la commande a été définie suite à une série de négociations entre Madame Gally et moi-même. Après avoir évoqué les éventuelles missions à entreprendre, nous nous sommes mis d'accord sur le projet suivant : la conception et la mise en ligne d'un test de positionnement autocorrectif dont la passation se fait à distance, c'est-à-dire que les étudiants le passent chez eux, avant d'arriver à Grenoble. Je vise donc premièrement à réaliser une étude et une analyse du test qui existe actuellement afin de repérer ses points positifs et négatifs et, dans un deuxième temps, la conception d'un nouveau test de positionnement autocorrectif.

Madame Gally m'a en outre accordé la liberté d'apporter au test n'importe quel changement jugé nécessaire pour l'améliorer. Elle a d'ailleurs précisé que des parties du test pouvaient être conservées, transformées, remplacées ou bien complètement supprimées.

⁷ Durant les trois premières semaines de cours, les étudiants peuvent demander un changement de groupe s'ils sentent que le cours n'est pas adapté à leur niveau. Ce changement est ensuite effectué avec l'accord de l'enseignant. De même, celui-ci peut également demandé un changement s'il trouve qu'un étudiant a été placé dans le mauvais groupe de niveau.

⁸ J'évoquerai plus en détails les opinions des professeurs de FLE à GEM sur le test original dans la troisième partie de ce mémoire.

5.3. La commande finale

Suite aux premières semaines de stage, Madame Gally et moi nous sommes rendus compte que le problème du placement des étudiants dans le groupe approprié ne venait pas uniquement du test écrit, mais de toute la procédure de positionnement. De fait, ce processus comprend deux étapes : premièrement, un étudiant est reçu par un enseignant lors d'un entretien oral afin d'évaluer sa capacité d'interagir à l'oral, mais aussi pour recueillir des informations sur son expérience de la langue française. Deuxièmement, les connaissances grammaticales et l'expression écrite de l'étudiant sont évaluées par l'intermédiaire d'un test écrit. En examinant la fiche pour l'entretien oral, Madame Gally et moi-même avons conclu que celle-ci n'était pas assez précise pour permettre à l'enseignant de juger correctement du niveau de l'apprenant. Aussi avons-nous conclu que ce test oral était perfectible, et son analyse et l'élaboration d'un nouveau test oral ont été ajoutées à la commande.

De plus, Madame Gally et moi-même avons également décidé qu'il serait avantageux d'intégrer une auto-évaluation dans le processus du placement des étudiants dans les groupes de niveau. Cette idée est venue premièrement de mes lectures sur l'évaluation en langues (Huver et Springer, 2011 ; Lussier et Turner, 1995 ; Conseil de l'Europe, 2001) et, deuxièmement, de l'exemple du programme d'espagnol langue étrangère qui se sert d'une auto-évaluation pour placer les étudiants dans les groupes de niveaux. La responsable de ce programme m'en a confirmé l'utilité. C'est pour ces deux raisons que nous avons pris la décision d'élaborer cette auto-évaluation et de l'intégrer dans le processus de positionnement⁹.

Ainsi constatons-nous que la commande a évolué au cours du premier mois de stage. C'était en examinant non seulement le test écrit mais aussi le test oral que nous nous sommes aperçus de l'intérêt de revoir l'ensemble de la procédure de positionnement, en y ajoutant notamment la conception d'une auto-évaluation, afin de rendre plus complet et efficace le placement des étudiants dans les différents groupes de niveau. C'est donc pour toutes ces raisons que la commande est désormais intitulée « un dispositif hybride de positionnement en français ».

⁹ L'usage de l'auto-évaluation sera expliqué dans la troisième partie de ce mémoire.

Partie 2 : L'État de l'Art

L'acte d'évaluer fait partie intégrante de notre société. Nous subissons constamment différents types d'évaluations, parfois sans en être conscients. À partir de l'entrée dans le système scolaire, par exemple, nous devons faire preuve de différentes compétences et capacités afin de poursuivre notre éducation, d'obtenir une certification ou d'être diplômé. Ces évaluations peuvent avoir un impact plus ou moins important sur notre vie, comme c'est le cas avec le baccalauréat en France. La vie d'un étudiant ne réussissant pas cet examen risque d'être bouleversée, l'obligeant à repenser son parcours académique ou professionnel.

L'évaluation a également sa place dans d'autres contextes. Un bon exemple est celui du permis de conduire où l'on doit réussir un examen afin d'avoir le droit de conduire. Par ailleurs, nous retrouvons des évaluations dans le monde du travail. Un employeur effectue parfois des contrôles de ses employés afin de vérifier leur efficacité au sein de l'établissement. Même au niveau social, nous remarquons l'utilisation d'évaluations, ce qui est le cas avec l'obtention de la citoyenneté dans certains pays où un processus rigoureux est parfois imposé aux étrangers souhaitant l'obtenir. Ils doivent d'ailleurs faire preuve d'un certain niveau de langue, des connaissances du pays et son fonctionnement, etc. Dans le domaine du sport, l'évaluation est également présente, surtout si nous considérons les Jeux Olympiques pendant lesquels les athlètes reçoivent, ou non, une médaille en fonction de leur performance.

En somme, les procédures d'évaluations sont omniprésentes dans nos vies, et constituent un obstacle parfois inévitable pour accéder à de nombreux droits et opportunités non seulement académiques, mais également professionnels et sociaux.

1. L'acte d'évaluer en langues : les destinataires et les finalités

Qu'est-ce qu'évaluer ? Pourquoi évalue-t-on et dans quels buts ? Ce sont des questions qu'il faut se poser lorsque nous menons une réflexion sur la notion de l'évaluation en langue. D'une part, les réponses nous facilitent la compréhension de cet acte et, d'autre part, elles nous aident à mieux évaluer les compétences des apprenants.

1.1. « Évaluer » : une définition

« Estimer », « mesurer », « juger », « noter », ce sont des termes que nous entendons dans le domaine de l'évaluation en langue. Mais concrètement, que veut-dire « évaluer » dans un contexte d'apprentissage de langue ? Tout d'abord, pour définir de manière générale cette notion, nous pouvons nous appuyer sur les recherches de Huver & Springer (2011), qui expliquent que le mot « évaluer » provient de « avaluer » de l'ancien français, qui veut dire « faire une évaluation » ou « fixer la valeur de ». Plus précisément dans le domaine des langues, il s'agit de chercher des indices

ou des indicateurs sur les caractéristiques des compétences de la personne que nous évaluons. Ensuite, nous nous appuyons sur ces indices afin de porter un jugement sur le degré de maîtrise de la langue (Hadji, 1997).

Dans une perspective d'enseignement et d'apprentissage d'une langue étrangère, Lussier (1992 : 13) nous donne une définition plus détaillée de l'acte d'évaluer :

« L'évaluation des apprentissages est une démarche qui consiste d'abord à recueillir des informations (résultats de mesure, appréciations, observations...) et à les rendre significatives ; à porter des jugements sur les données recueillies ; et à prendre des décisions quant à l'action à entreprendre compte tenu de l'intention d'évaluation de départ ».

Cette définition montre la complexité de cet acte. En effet, il s'agit d'une procédure qui se déroule en quatre étapes : (1) le recueil de données sur les compétences de l'évalué, (2) l'analyse de ces données, (3) l'expression d'un jugement qui s'appuie sur ces données et, enfin, (4) la prise de décision en fonction de ce jugement et de l'objectif de l'épreuve.

D'autres chercheurs, comme Huver & Springer (2011), vont plus loin en expliquant que ceci est la mise en relation de différentes procédures évaluatives, ayant des finalités différentes, en fonction du contexte, de l'institution et du public. En d'autres termes, l'évaluation doit s'adapter aux circonstances dans lesquelles elle se déroule, en prenant en compte à la fois le public, le type d'institution et l'objectif. Ce constat sera un guide important pour mon travail étant donné les particularités du contexte de stage et du public évoquées précédemment¹⁰.

1.2. Qui évalue qui et pourquoi

Cette question induit généralement la réponse, « l'enseignant évalue celui qui apprend la langue pour savoir si cette personne a appris telle ou telle chose ». Cela est une réponse parmi d'autres. Dans l'une des situations d'évaluations possibles, c'est effectivement le formateur qui juge de la qualité des compétences des apprenants. En revanche, ce n'est pas uniquement l'étudiant qui est évalué. La passation d'un examen permet également à l'enseignant d'évaluer son enseignement. Ce dernier a-t-il été efficace ? Les apprenants ont-ils atteint les objectifs d'apprentissage visés par le cursus ? Quelles difficultés restent à surmonter ? L'enseignant a-t-il besoin de revenir sur certains éléments qui n'ont pas été systématisés ? L'enseignant peut-il avancer dans son enseignement parce que les objectifs ont été atteints ? Ces questions aident le professeur non seulement à évaluer l'efficacité de son enseignement, mais également à prendre des décisions en fonction des résultats dans le but d'améliorer ses cours.

¹⁰ cf. Partie 1, (3.2. et 4.)

L'évaluation peut aussi servir à mesurer la qualité du programme d'enseignement dans sa totalité, voire celle de l'institution. L'évaluation dépasse ainsi la classe et le formateur pour englober l'ensemble des facteurs et acteurs impliqués dans l'enseignement de la langue. Les résultats donnés par un examen peuvent indiquer, par exemple, si le cursus mis en place par l'institution facilite l'atteinte des objectifs d'apprentissage, ou bien s'il faut le revoir pour qu'il réponde mieux aux besoins des apprenants.

En résumé, il faut retenir que l'évaluation dépasse celle de l'apprenant car elle mesure également la qualité du programme d'études, de l'enseignant ou bien de l'institution elle-même.

2. Les différents types d'évaluations et leurs finalités

Les types d'évaluations en langue sont très variés dans leurs finalités, d'une part dans la manière dont elles se déroulent et à quel moment et, d'autre part, dans leurs impacts, leurs formes et ce sur quoi elles portent. Afin de cerner l'utilité des différentes évaluations, il est important de prendre conscience de la diversité des pratiques évaluatives pour pouvoir choisir la plus pertinente dans le contexte et pour l'objectif. Ainsi, je souhaite maintenant mettre en évidence les différentes évaluations dans l'enseignement des langues. À cette fin, tout comme Veltcheff & Hilton (2003) et Lussier (1992), je me base sur les moments où l'on fait passer une épreuve, à savoir après, pendant ou avant l'apprentissage. Je privilégie cette façon de décrire la nature des types d'évaluations puisqu'elle permet de mieux comprendre l'objectif de chaque épreuve en les mettant en relation avec l'apprentissage.

2.1. Après l'apprentissage : l'évaluation sommative

En aval d'une unité d'apprentissage, l'évaluation qui se met en place est de type sommatif étant donné qu'elle s'effectue « au sommet », ou bien à la fin, du cursus. De même, il s'agit d'un recueil d'informations à la fin de l'apprentissage, permettant à l'enseignant de faire le bilan des apprentissages et, de ce fait, de déterminer l'atteinte ou non des objectifs du programme d'enseignement dans sa globalité (Tagliante, 2005). Le CECRL la définit ainsi : « L'évaluation sommative contrôle les acquis à la fin du cours et leur attribue une note ou un rang. Il ne s'agit pas forcément d'une évaluation de la compétence. En fait, l'évaluation sommative est souvent normative, ponctuelle et teste le savoir. » (Conseil de l'Europe, 2001 : 141). Grâce à cette acception, je peux souligner les particularités d'une évaluation sommative. De fait, elle s'effectue à un moment précis, porte plutôt sur les savoirs et les savoir-faire et fournit, en général, une appréciation à l'apprenant sous forme d'une note ou de l'attribution de crédits (Tagliante, 2005). De plus, elle peut accorder le droit de passer à un niveau d'études ou une classe supérieurs, ou bien de certifier ou diplômer l'apprenant. L'évaluation sommative porte en outre sur la globalité de l'apprentissage, et regroupe donc les objectifs de chaque séquence et en évalue la maîtrise (Mothe, 1975). Il faut également

préciser qu'elle est destinée à des apprenants ayant suivi la même formation, contrairement à d'autres types d'évaluations telle que l'évaluation prédictive¹¹ (ibid.).

2.2. Pendant l'apprentissage : l'évaluation formative

Lorsque nous évoquons les évaluations se déroulant au cours de l'acquisition de la langue, les termes sont de nouveau variés. Mothe (1975) et Tagliante (2005) emploient le mot « diagnostic » pour désigner ce type d'épreuve puisqu'il s'agit d'une évaluation ayant pour but d'effectuer un bilan des apprentissages en fonction des objectifs d'une séquence spécifique.

En revanche, le terme le plus répandu, et donc celui que j'emploie, est « évaluation formative ». D'ailleurs, le CECRL utilise ce même terme et nous en fournit la définition suivante :

«...un processus continu qui permet de recueillir des informations sur les points forts et les points faibles. L'enseignant peut alors les utiliser pour l'organisation de son cours et les renvoyer aussi aux apprenants. On utilise souvent l'évaluation formative au sens large afin d'y inclure l'information non quantifiable fournie par des interrogations et des entretiens » (2001 : 141).

En effet, nous cherchons à évaluer la progression des apprenants en fonction des objectifs d'apprentissage d'une séquence donnée, et non pas de l'unité dans sa globalité comme c'est le cas avec une évaluation sommative (Mothe, 1975). De ce fait, nous voyons que ce type d'épreuve ne s'utilise pas pour certifier ni pour valider les acquis de l'utilisateur. Il est plutôt question de faire un état des lieux de l'apprentissage au fur et à mesure de l'unité d'enseignement (Lussier & Turner, 1995).

Les données d'une épreuve formative permettent de répondre aux questions suivantes : les apprenants ont-ils atteint les objectifs nécessaires pour poursuivre l'apprentissage ? Faut-il revenir sur certains points ? La difficulté du cursus est-elle adaptée au niveau du groupe ? Faut-il modifier les objectifs ? En somme, ce type d'évaluation sert à suivre la progression des étudiants, à réguler et guider l'apprentissage en relevant leurs difficultés et leurs points forts (Lussier & Turner, 1995 ; Tagliante, 2005). Dans un deuxième temps, l'enseignant effectue des changements en fonction des résultats de l'évaluation pour donner aux étudiants les moyens de s'améliorer ou de remédier à leurs erreurs. Les résultats peuvent aussi inciter l'enseignant à modifier le programme ou l'enseignement pour l'adapter davantage aux besoins des apprenants. Ainsi, nous pouvons constater que l'évaluation formative sert d'une part à aider les apprenants dans l'acquisition de la langue et, d'autre part, à montrer à l'enseignant si son enseignement a été efficace ou non.

Enfin, en examinant la définition donnée par le CECLR ci-dessus, nous remarquons que l'évaluation formative peut s'effectuer grâce à un éventail d'outils évaluatifs. De fait, l'enseignant

¹¹ L'évaluation prédictive sera expliquée dans la partie 2 (3.3.) de ce mémoire.

peut aussi effectuer des évaluations formatives à travers des observations, des discussions avec les étudiants, des auto-évaluations ou des questionnaires (Veltcheff & Hilton, 2003 ; Conseil de l'Europe, 2001).

2.3. Avant l'apprentissage : l'évaluation prédictive

Les termes pour désigner l'évaluation prédictive, qui s'effectue en amont des apprentissages, sont assez divers. Nous trouvons, par exemple, « pronostique », « positionnement », « orientation », « aptitude », « niveau », « placement », « d'entrée »... Au vu du nombre de façons de décrire ce type d'évaluation, il n'est pas surprenant que cette pluralité se manifeste aussi dans les usages de cette épreuve. Elle peut servir à orienter la personne vers une certaine formation, à accorder ou à refuser le droit d'entrer dans une institution académique, à déterminer les potentialités de quelqu'un ou bien à créer des groupes de niveau.

Néanmoins, dans le cadre de ce mémoire il est pertinent d'en donner la définition suivante : il s'agit d'une évaluation pour déterminer le profil langagier des apprenants en cernant leurs points forts et faibles par rapport à la langue française (Veltcheff & Hilton, 2003). Cette définition de l'évaluation prédictive est confirmée par l'acception du mot « prédire » qui, d'après le *Trésor de la langue française*, signifie « Annoncer à l'avance un événement par connaissance inductive, rationnelle des causes et des effets ». Ainsi, l'évaluation prédictive sert effectivement à récolter des informations sur les compétences des apprenants, et ensuite de prendre une décision en fonction de ces informations.

Plus spécifiquement, j'emploie les termes « test de positionnement » ou « test de placement » car ils renvoient directement à l'usage de cette évaluation dans le cadre de mon projet de stage. L'objet d'un test de positionnement est en effet de placer ou de positionner les étudiants, en considérant leur profil linguistique fourni par les résultats du test, dans le groupe dont le niveau leur correspond (Hadji, 1997). Ou encore, c'est en vérifiant que les apprenants possèdent les requis nécessaires pour suivre un cours de français de tel ou tel niveau que nous arrivons à les mettre dans le bon groupe (Lussier, 1992). Enfin, il est possible de décrire un test de positionnement en évaluant les connaissances acquises auparavant par les étudiants avec l'intention explicite d'utiliser les résultats pour effectuer un classement des apprenants dans des groupes de niveaux (Mothe, 1975). Ainsi, l'objectif global d'un test de positionnement est de mettre en place des groupes de niveau homogène en regroupant les apprenants ayant des compétences similaires, voire identiques, en français. Tagliante (2005 : 14) nous donne une définition complète d'un test de positionnement :

« Ces tests servent à orienter l'élève vers un groupe de niveau homogène et à l'informer de ses capacités langagières, afin qu'il puisse prendre, en connaissance de cause, les décisions qui concernent son apprentissage. Le test de niveau permet de faire une sorte de bilan des atouts et des points faibles de l'élève et constitue une source d'information essentielle pour l'enseignant. »

Ici, nous remarquons qu'un test de placement n'est pas uniquement utile à l'institution et aux enseignants, il l'est aussi pour l'apprenant. En prenant conscience de son niveau, celui-ci peut davantage prendre en charge son apprentissage. Afin d'être un minimum autonome, l'utilisateur de la langue doit tout d'abord savoir où il se situe parmi les différents niveaux. De même, en prenant conscience de ce qu'il maîtrise et de ce qu'il ne maîtrise pas sur le plan linguistique, il sait ce sur quoi il a besoin de travailler (Tagliante, 2005).

Par ailleurs, en étudiant le concept de test de positionnement, il est important de se rendre compte de l'importance de la bonne qualité de ce type d'évaluation non seulement pour que le cours se déroule bien, mais aussi pour garantir un fonctionnement efficace de l'institution. En ce qui concerne cette dernière, des changements de groupe exige un travail qui coûte forcément à l'établissement en termes de temps et donc d'argent. Les personnes chargées d'effectuer ces changements de groupe doivent de fait consacrer du temps à cette tâche.

Un test de positionnement de qualité favorise également le bon déroulement des cours. Par exemple, un étudiant ayant été placé dans le mauvais niveau va s'efforcer de suivre l'enseignement, ce qui peut avoir de nombreux effets négatifs sur son apprentissage de la langue, notamment la démotivation ou la perte de confiance. De plus, il risque de ralentir l'acquisition de la langue des autres apprenants dans le cours. En effet, les étudiants ayant été placés dans le bon groupe de niveau sont généralement capables de progresser plus vite que l'étudiant mal placé. Cependant, ils ne peuvent pas forcément poursuivre l'apprentissage à la même vitesse étant donné que cet étudiant peut avoir besoin de plus d'aide et de temps pour acquérir les mêmes connaissances. À contrario, un étudiant dont le niveau est plus élevé que celui de son groupe peut s'ennuyer et, de ce fait, son apprentissage ne progresse pas à la même vitesse que s'il était dans le groupe approprié. Ainsi pouvons-nous voir qu'un test de positionnement qui cerne bien le niveau des apprenants favorise une réalisation fluide et sans perturbation importante de la gestion et du déroulement des cours.

Il se révèle en outre important d'examiner les particularités d'un test de placement, car celui-ci diffère considérablement des évaluations formatives et sommatives. Il s'effectue en amont de la séquence d'apprentissage, et ne nécessite aucune préparation de la part des apprenants. Il est même préférable qu'ils ne se préparent pas pour que les résultats témoignent de leurs vraies compétences (Veltcheff & Hilton, 2003). Quant au public auquel est destiné un test de positionnement, il s'agit d'apprenants n'ayant pas forcément étudié le français dans le même contexte (pays, institution, durée, objectifs, etc.), tandis que le public passant une épreuve sommative ou formative a suivi la même formation.

De plus, il faut que le test recouvre une grande quantité de contenus linguistiques, qu'il évalue le plus de compétences possibles, que les types d'items soient variés et, enfin, qu'il comporte

entre 80 et 120 questions (Veltcheff & Hilton, 2003). Quant à ces trois dernières caractéristiques, elles veillent à ce que le test fournisse une image globale des compétences de l'évalué. Autrement dit, en intégrant un grand nombre d'items (QCM, réponse courte) qui évaluent les différentes compétences (grammaire, syntaxe, compréhension écrite, lexique), nous arrivons à un test dont les résultats témoignent véritablement des compétences de l'utilisateur (état, connaissances). Un apprenant appréhende la langue française de manière personnelle, et possède sans doute des compétences partielles ou inégales. Aussi est-il important que le test évalue l'étudiant de manière globale pour prendre en compte ses différences, et donc cerner son vrai niveau.

En ce qui concerne le contenu d'un test de positionnement, nous venons de voir que celui-ci doit être le plus varié possible. Il existe toutefois une autre particularité. L'étendue linguistique de ce type de test est basée sur les objectifs d'apprentissage de l'unité à venir, et non pas ceux d'une unité finie, ce qui est le cas avec une évaluation formative ou sommative. Il reste donc à se décider sur quels outils ou références s'appuyer pour déterminer quels objectifs seront ciblés par l'enseignement, et donc quel contenu devra être intégré dans le test de positionnement.

Enfin, il est indispensable qu'un test de placement soit évolutif au niveau de la difficulté, c'est-à-dire que la difficulté progresse au fur et à mesure que l'étudiant avance dans le test (Veltcheff & Hilton, 2003). Cela est de fait important pour pouvoir situer l'apprenant sur une échelle de niveaux du CECRL allant de début (A1) à avancé (C2).

2.4. L'importance des objectifs d'apprentissage en évaluation

Le terme « objectif d'apprentissage » revient constamment dans les recherches sur l'évaluation en FLE (Lussier & Turner, 1995 ; Huver & Springer, 2011 ; Conseil de l'Europe, 2001). Elles revendiquent en effet qu'une évaluation est basée exclusivement sur les objectifs d'apprentissage de la séquence ou de l'unité d'apprentissage précédente, ce que nous avons vu avec les trois types d'évaluations expliqués ci-dessus. En d'autres termes, l'évaluation ne demande que la mise en œuvre des compétences travaillées en cours et, de surcroît, n'exige pas de l'apprenant qu'il les produise de manière considérablement différente que pendant l'apprentissage. De cette manière, la relation qui lie l'évaluation à l'enseignement est conservée.

2.5. D'autres types d'évaluations

Bien qu'il soit avantageux de se baser sur les moments où l'on évalue pour mettre en évidence les différents types d'évaluation, il en existe d'autres qui peuvent se réaliser à différents moments de l'apprentissage et qui méritent d'être évoqués. Pour ce faire, je m'appuie sur le CECRL (chapitre 9), car cet ouvrage présente treize types d'évaluations, mettant en évidence la complexité des pratiques évaluatives en langue.

2.5.1. L'évaluation critériée normative

C'est la façon dont nous interprétons les résultats qui fait qu'un examen soit normatif ou critérié. En effet, une épreuve normative situe les apprenants les uns par rapport aux autres en fonction des résultats. Comme son nom l'indique, elle renvoie à une norme qui est déterminée par les résultats des personnes ayant passé le même examen. Ainsi, les apprenants sont comparés, et cette comparaison permet à l'évaluateur de leur fournir un résultat concret (« Vous êtes vingtième », par exemple).

Quant à une évaluation critériée, au contraire, elle compare la performance de la personne à des critères basés sur une norme langagière. Dans le cadre d'une évaluation basée sur le CECRL, par exemple, ce dernier fournit des descripteurs sur lesquels il est possible de baser son jugement. Il est donc question de comparer la performance de l'apprenant à ces normes pour, par la suite, déterminer la qualité de cette production langagière. Comme exemple d'une évaluation critériée, nous pouvons citer les épreuves DELF / DALF qui sont à la fois des évaluations critériée et certificatives.

2.5.2. L'évaluation holistique et analytique

L'évaluation holistique renvoie à un jugement globale des compétences de l'apprenant en l'amenant à produire du langage oral ou écrit, tandis qu'un examen analytique porte sur des aspects précis, souvent à travers des tests standardisés, portant sur les compétences grammaticales ou lexicales. Comme le précise le CECRL, nous pouvons caractériser une évaluation comme holistique ou analytique en examinant soit ce que nous évaluons (la performance globale ou des compétences spécifiques), soit la manière dont nous interprétons les résultats (une échelle globale ou bien une échelle permettant de noter des éléments séparés).

2.5.3. L'évaluation de la performance et des connaissances

Lorsque nous évaluons la performance d'un apprenant, notre objectif est de savoir s'il peut produire « un échantillon de discours oral ou écrit » (Conseil de l'Europe, 2011 : 142), tandis qu'une évaluation des connaissances exige qu'il « réponde à des questions de types différents afin de porter la preuve de l'étendue de sa connaissance de la langue et du contrôle qu'il en a » (*ibid.*, 2011 : 142). En d'autres termes, une évaluation de la performance est basée principalement sur la capacité de l'apprenant à produire un langage écrit ou oral en mobilisant correctement différents actes de parole. La qualité de ses performances est déterminée à l'aide de critères précis, comme nous l'avons vu avec l'évaluation critériée. À contrario, celle portant sur les connaissances vérifie si l'apprenant maîtrise des éléments spécifiques et isolés de la langue.

L'explication de ces types d'évaluations mène à une discussion indispensable si l'on s'intéresse à l'évaluation en langues : celle de la performance et de la compétence. En effet, il est

impossible d'évaluer directement une compétence (Lussier & Turner 1995). Il faut s'efforcer de mesurer la qualité d'un ensemble de performances pour ensuite pouvoir mesurer les compétences (Conseil de l'Europe, 2011). De même, par « performance » nous entendons la production langagière orale ou écrite à travers la mise en œuvre d'actes de parole, tandis qu'une compétence est la manifestation de différentes capacités (*ibid.*).

2.5.4. L'évaluation du savoir et de la capacité

Le CECRL explique que la mesure du savoir porte sur ce qui a été enseigné lors des séances en amont, et se rapporte à l'évaluation de l'atteinte ou non des objectifs d'apprentissage spécifiques. Ainsi, l'ouvrage souligne qu'il s'agit d'un point de vue intérieur par rapport à l'apprentissage, c'est-à-dire que l'évaluation est centrée sur le contenu du cours.

On oppose à ce type d'évaluation celle de la capacité qui est basée sur ce que la personne sait ou peut faire dans un contexte de la vie réelle. La notion de capacité se traduit par la mise en œuvre de la compétence. Elle est également synonyme de « performance ».

2.5.5. Maîtrise ou continuum

Ici, il est question de la maîtrise complète des compétences qui ne prend pas en compte le niveau de qualité de la compétence. À contrario, l'évaluation d'une approche continuum considère la qualité partielle de la compétence. De fait, les résultats de ce type d'épreuve montrent à l'enseignant si l'apprenant est capable ou non de faire preuve d'une compétence linguistique donnée. Ces types d'examen sont tous deux basés tous les deux sur des critères spécifiques, démontrant leur relation proche avec une évaluation critériée.

2.5.6. L'évaluation continue ou ponctuelle

L'évaluation continue s'effectue tout au long de la durée de l'apprentissage que celui-ci soit long ou court, et donne, au final, une note globale à l'évalué. Elle mesure la capacité de l'apprenant à réaliser des projets et à faire preuve de performances. Enfin, ce type d'épreuve est généralement pris en main par l'enseignant mais, dans certains cas, c'est l'apprenant qui s'auto-évalue.

Au contraire, l'évaluation ponctuelle se réalise à des moments précis et définis lors de l'apprentissage. La note qui en découle n'est pas influencée par d'autres épreuves du même type effectuées auparavant. De ce fait, une évaluation ponctuelle porte uniquement sur ce que la personne peut faire au moment de l'examen.

2.5.7. L'évaluation formative ou sommative

Comme nous l'avons vu dans la partie 3 (3.1. et 3.2.), l'évaluation formative est utilisée afin de récolter des données qui reflètent les points forts et faibles de l'apprenant. Ces données servent

par la suite à apporter des aides et des remédiations à ce dernier, mais également à modifier l'enseignement en fonction des besoins des apprenants et des enseignants. L'évaluation formative inclut également des évaluations dites informelles telles que des questionnaires et des entretiens auprès des apprenants.

Quant à l'évaluation sommative, il s'agit d'un contrôle de la compétence qui s'effectue à la fin du cursus afin d'accorder une note à l'apprenant ou de le positionner parmi les autres de la classe. Ainsi, ce type d'examen est souvent normatif, ponctuel et mesure les savoirs des apprenants.

2.5.8. L'évaluation directe et indirecte

L'évaluation directe se rapporte à une observation de l'enseignant sur ce que l'apprenant est en train de faire à un moment précis. Il peut s'agir par exemple d'un travail de groupe ou bien d'une production écrite ou orale en cours. Un entretien est un exemple typique d'une évaluation directe. Afin de mesurer la qualité de l'action, le professeur se réfère à des critères précis. Par ailleurs, il est important de souligner que ce type d'épreuve ne peut évaluer que la production orale et écrite étant donné que les activités de compréhension ne sont pas mesurables directement.

À contrario, l'évaluation indirecte prend, en générale, la forme d'un test écrit qui évalue souvent les potentialités de la personne. Nous pouvons citer comme exemple de ce type d'examen les QCM ou bien un texte lacunaire.

2.5.9. L'évaluation objective et subjective

Par évaluation subjective nous entendons un jugement porté par l'examineur. Elle est généralement utilisée pour mesurer la qualité d'une performance. À contrario, une épreuve objective, considérée aussi comme indirecte, est celle qui pour chaque question ne permet qu'une seule bonne réponse. Ainsi, les QCM sont très souvent les composantes d'un test objectif. Le CECLR souligne la complexité de ces enjeux, d'autant plus que les deux types d'évaluations apportent à la fois des avantages et des inconvénients à la pratique évaluative. Cela dit, les auteurs nous rappellent qu'un test objectif prime sur celui qui est subjectif. Afin d'examiner plus en détail ces deux notions, et d'apporter les points de vue d'auteurs chercheurs, l'objectivité et la subjectivité seront abordées plus profondément dans la partie 2 (4.3.1.).

2.5.10. L'évaluation sur échelle ou liste de contrôle

Comme son nom l'indique, une évaluation sur échelle permet à l'examineur de positionner l'apprenant sur une échelle de niveaux verticale tandis que celle sur liste de contrôle amène l'enseignant à porter un jugement en fonction des points supposés être pertinents selon un niveau de langue ou bien un modèle.

2.5.11. L'évaluation sur une impression ou un jugement guidé

Une évaluation à partir d'une impression porte uniquement sur une performance. L'enseignant observe cette dernière et se prononce ensuite sur sa qualité sans se référer à des critères. Ainsi est-elle complètement subjective. Une évaluation selon un jugement guidé s'oppose à cette dernière étant donné que l'enseignant se base sur des critères spécifiques lorsqu'il observe et porte un jugement sur la performance de l'apprenant.

2.5.12. L'évaluation par série ou par catégorie

L'évaluation par série consiste à s'intéresser à la capacité de l'apprenant à effectuer une série de tâches diverses. Nous notons en général de façon globale en fonction d'une échelle chiffrée de 1 à 3, par exemple). Le jeu de rôle est un exemple typique de ce type d'évaluation. À contrario, une évaluation par catégorie est basée sur une seule tâche, et le jugement est porté en fonction des catégories spécifiques d'une grille d'évaluation.

2.5.13. L'auto-évaluation

Développée dans les années 1970, l'auto-évaluation occupe une place de plus en plus importante de nos jours (Huver & Springer, 2011). D'ailleurs, certains tests standardisés courants, tel que *Dialangue*, contiennent une auto-évaluation. De plus, le CECRL l'évoque de manière détaillée et en souligne les principaux intérêts. De fait, cet ouvrage la définit comme un jugement que l'apprenant porte sur ses propres connaissances de la langue (Conseil de l'Europe, 2001). Elle s'effectue en fonction de critères précis définis à partir de descriptions des niveaux de langue. Quant à ces niveaux, le CECRL en propose six : A1, A2, B1, B2, C1, C2.

Le CECRL souligne deux principaux usages d'une auto-évaluation. Premièrement, elle permet à l'apprenant d'autodéterminer son niveau en amont des apprentissages (Tagliante, 2005). Par là même, la personne est amenée à prendre conscience de ses points forts et faibles, ce qui l'implique directement dans son apprentissage. La tâche d'évaluer est ainsi partagée entre l'enseignant et les apprenants (Lussier & Turner, 1995). Puis, grâce à la prise de conscience de ses forces et faiblesses, l'apprenant peut mieux gérer son apprentissage, le rendant ainsi plus fructueux (Conseil de l'Europe, 2001 ; Lussier & Turner, 1995).

Non seulement l'auto-évaluation provoque chez l'apprenant une prise de conscience de ses compétences, mais elle s'avère également efficace pour évaluer de manière plus juste ses compétences. Par exemple, les résultats de l'auto-évaluation peuvent être comparés à ceux de l'évaluation effectuée par l'enseignant, ce qui lui fournit plus d'informations sur les compétences de l'étudiant (Conseil de l'Europe, 2001 ; Lussier & Turner, 1995). En s'appuyant sur ces données, venant à la fois de l'apprenant et de l'enseignant, ce dernier arrive à confirmer ou infirmer les

résultats donnés par sa propre évaluation. Il s'agit ainsi d'équilibrer les types de résultats en variant la typologie d'évaluation pour arriver à une image plus juste du niveau ou des compétences de l'évalué.

En résumé, le premier usage de l'auto-évaluation s'applique à des fins pédagogiques, c'est-à-dire qu'il est en lien avec l'apprenant et son acquisition de la langue. Quant au second, il s'agit d'un usage évaluatif, car l'enseignant juxtapose les résultats avec ceux issus de son évaluation pour porter un jugement plus juste sur les compétences de l'apprenant

2.6. Une conclusion sur les différentes évaluations

Comme nous venons de le voir, l'évaluation est une pratique complexe, surtout lorsque nous examinons le chapitre 9 du CECRL. En étudiant les techniques décrites par cet ouvrage, nous pouvons faire quelques constats importants. Premièrement, ces treize types d'évaluations se superposent très souvent, un fait également souligné par Tagliante (2005). Autrement dit, ces différentes épreuves partagent fréquemment les mêmes caractéristiques et sont donc similaires quant à leurs déroulements et leurs finalités. Nous pouvons noter par exemple l'utilisation de certains termes tels que « subjectif », « objectif », « indirecte », « directe » et « critère » dans les descriptions de plusieurs types d'évaluations. Ce constat provoque la question suivante : pourquoi le CECRL ne présente-t-il pas une liste d'évaluations davantage concise ? Il est vrai que la façon dont cet ouvrage détaille les différentes techniques évaluatives s'approche d'une liste plutôt exhaustive, fournissant à l'enseignant un bagages d'outils adaptables à presque tout contexte. Cependant, il semble qu'une description simplifiée serait éventuellement préférable pour un enseignant qui décide de quelle évaluation à mettre en place.

Deuxièmement, le CECRL, contrairement à d'autres ouvrages, ne présente pas les types d'évaluation en fonction du moment où ils se déroulent. Cet ouvrage les évoque par paires en sorte que l'on puisse juxtaposer l'une à son opposée, mettant ainsi en évidence les grandes différences entre elles.

Enfin, nous remarquons que cette référence ne cite pas le test de positionnement. Il est vrai que quelques évaluations parmi les treize partagent certaines caractéristiques d'une épreuve qui détermine le niveau des apprenants (évaluation analytique, ponctuelle, objective, etc.), mais le CECRL ne l'explique pas de manière explicite¹².

En définitive, au vu de ces remarques, il se révèle nécessaire, d'une part de bien connaître les différents usages de ces épreuves et, d'autre part, de mener une réflexion approfondie avant de mettre

¹² Il est vrai qu'en général le test de positionnement n'est pas toujours évoqué en détails, contrairement aux autres types d'évaluations. De fait, les auteurs qui évoquent le font de manière moins détaillée, ce qui nous incite à nous demander pourquoi.

en place l'examen pour que celui-ci soit le plus approprié possible. Ainsi devons-nous penser à l'objectif et à l'objet de l'évaluation, au moment où elle se déroule, à son fond et à sa forme pour qu'elle soit utile, et renforce l'apprentissage. En connaissant les différentes façons de présenter les évaluations on langue (i.e. selon le moment où elles se déroulent vs. par pairs), nous arrivons à sélectionner l'épreuve la plus pertinente.

3. Les quatre étapes du processus évaluatif et les enjeux associés

Nous venons de définir l'acte d'évaluer, les destinataires et les types d'évaluations. Je voudrais maintenant détailler la procédure par laquelle nous passons pour évaluer. À cette fin, je m'appuie sur l'ouvrage de Lussier (1992), car la manière dont elle explique le déroulement en quatre étapes d'une évaluation, de sa conception à la prise de décision, est, d'après moi, la plus claire et concise. De surcroît, nous pouvons appliquer toutes les étapes à un test de positionnement, ce qui n'est pas le cas avec certains autres types d'épreuve. Cette procédure ne s'applique pas dans son intégralité à une évaluation formative, par exemple, car la dernière étape ne se réalise pas avec ce type d'examen. Ainsi ai-je choisi de me servir de cette démarche pour mettre en évidence la procédure évaluative.

De plus, la façon dont Lussier explique la procédure évaluative me permettra de mettre en évidence les différents enjeux liés à l'évaluation qu'il faut impérativement prendre en compte lorsque nous concevons un examen en langue.

3.1. Étape 1 : l'intention de l'évaluation

Tout d'abord, il est nécessaire de décider de l'intention de l'évaluation qui sera mise en place. Autrement dit, qu'est-ce que nous voulons tirer de l'évaluation ? À quoi sert-elle ? Les objectifs de l'évaluation varient selon plusieurs facteurs dont le moment où elle se déroule, ainsi que l'impact qu'elle produit sur l'étudiant, sur l'enseignant ou bien sur le cursus. Comme nous l'avons vu dans la partie 2 (2.2.) de ce mémoire, l'évaluation peut se réaliser pour évaluer le professeur et son enseignement, les acquis ou non des apprenants, le programme d'études ou l'institution. De même, l'évaluation peut se mettre en place pour attribuer une note aux étudiants ou pour déterminer le passage ou non à un niveau d'études supérieur.

Quelle que soit l'intention, ce qui est primordial à cette étape est que le formateur décide clairement de la finalité de l'évaluation avant de commencer à l'élaborer. Une telle décision est indispensable pour que l'examen soit adapté à la situation évaluative, et qu'il fournisse à l'enseignant des informations utiles pour lui et pour ses étudiants.

3.1.1. La praticabilité

Un facteur lié à l'évaluation est sa praticabilité. Appelé aussi « faisabilité », il s'agit de la facilité avec laquelle l'on fait passer et corrige une évaluation. En d'autres termes, pour que l'évaluation soit économique, il est important que le correcteur ne doive pas consacrer trop d'énergie ni de temps à la tâche, y compris au niveau de la conception de l'évaluation (Noël-Jothy & Sampsonis, 2006). En effet, la question de son coût entre en jeu puisqu'un examen exigeant trop de temps devient trop dispendieux pour l'institution. De même, Lussier & Turner (1995) expliquent que la faisabilité de l'examen relève de l'efficacité des instruments, c'est-à-dire du type de questions, de leurs formes et du dispositif à travers lequel on fait passer l'évaluation. Pour s'assurer de l'efficacité des instruments, il est important de penser au contexte et à l'objectif de l'épreuve.

Il faut par ailleurs que l'instrument soit facile à noter, ce qui est un autre élément de la praticabilité. L'exemple d'un exercice d'appariement met en évidence l'importance de cette caractéristique pour l'évaluateur. De fait, ce type d'item demande à l'apprenant de relier un élément de la colonne de gauche avec un élément de la colonne de droite. S'il se trompe, il doit barrer le trait qui les relie et en redessiner un nouveau, ce qui veut dire que l'évaluateur peut avoir des difficultés à repérer la vraie réponse et donc à corriger l'exercice. Ainsi, un test difficile à noter est une épreuve d'une mauvaise praticabilité.

Cependant, il n'est pas seulement question de la facilité pour l'évaluateur, mais également pour les apprenants. En effet, pour assurer la praticabilité d'une évaluation, il faut que les consignes soient claires, que la durée et la longueur soient cohérentes avec le contexte dans lequel s'insère le test et qu'il soit adapté au niveau prédéterminé. De plus, lorsque c'est possible, l'ajout d'exemples contribue également à une bonne praticabilité de l'évaluation (Noël-Jothy & Sampsonis, 2006).

3.1.2. L'équité

Comme la praticabilité, l'équité est l'une des caractéristiques de l'évaluation qui être prise en compte lorsque nous décidons de l'intention de l'épreuve. Cet aspect sous-entend une implication de la part du concepteur et de l'établissement dans les impacts de l'évaluation (McNamara, 2000). L'institution, le créateur du test et les correcteurs sont en effet tous concernés par l'équité de l'évaluation.

En préservant l'équité de l'examen, il faut se demander si les candidats ont tous les mêmes chances de le réussir, si l'on a mis en place des conditions physiques et psychologiques favorables à tous, et si les conditions de l'examen sont identiques pour tous (Noël-Johny & Sampsonis, 2006). De plus, il est nécessaire que les apprenants soient informés des conditions de l'examen y compris du temps alloué, du nombre d'items et de l'impact de l'évaluation (*ibid.*). Autrement dit, il faut donner aux évalués toutes les informations relatives à l'examen.

3.2. Étape 2 : la mesure

Lors de cette étape, l'objectif est que l'examineur recueille des données qui l'aident à porter un jugement lors de l'étape suivante. Il s'agit en effet de collecter des données, de les organiser en vue de faciliter les deux prochaines étapes de ce processus (Lussier, 1992). Cette étape nous renvoie aux théories de Hadji (1997), qui évoque la notion d'indicateurs ou d'indices.

Plus spécifiquement, ces données sont liées à ce que nous voulons mesurer. Par exemple, est-ce que nous nous intéressons à une performance ou à une compétence spécifique ? Pour obtenir ces données, il est possible de se servir de plusieurs outils évaluatifs dont l'observation, l'auto-évaluation, l'évaluation formelle ou informelle, ou bien le questionnement. En d'autres termes, il est possible d'évaluer de plusieurs points de vue afin d'avoir un échantillon de données représentatif des compétences de l'apprenant (Bonboir, 1972).

De plus, l'évaluateur peut se servir d'une grille d'évaluation afin de guider la mesure. D'après Lussier (1992), ainsi que d'autres chercheurs dans ce domaine (Huver & Springer, 2011 ; Tagliante, 2005 ; McNamara 2000), il est impératif que la mesure réponde à l'objectif de l'évaluation de départ et qu'elle soit en lien avec la prise de décision qui suivra. Autrement dit, il faut mesurer ce qu'on veut évaluer en fonction des objectifs d'apprentissage visés.

Une fois que nous avons ces informations, il faut les organiser en vue de leur interprétation. De fait, il s'agit de regrouper ou de codifier les données en fonction des objectifs de l'examen, c'est-à-dire ce que l'on veut évaluer, afin d'aboutir à une appréciation « brute » de l'évalué. Ensuite, l'évaluateur analyse les données pour en tirer d'éventuelles significations. D'après Lussier (1992), nous pouvons interpréter ces informations soit de manière normative en comparant les résultats des étudiants, soit de manière critériée en comparant les données à une grille d'évaluation conçue en amont de l'évaluation.

3.2.1. La validité

Une explication de l'évaluation en langue ne serait pas complète sans évoquer la notion de la validité, car cette notion est centrale dans le domaine d'évaluation en langues. D'ailleurs, certains chercheurs estiment que cet enjeu est primordial dans la conception d'un examen quel que soit son objectif (Lussier & Turner, 1995 ; Bonboir, 1972).

Mothe (1975, p. 23) nous fournit une définition complète de la validité en expliquant que celle-ci est « ...la qualité qui fait que l'épreuve mesure effectivement ce qu'elle est censée mesurer et non autre chose. C'est donc d'abord une qualité qui n'est pas absolue ni intrinsèque au test, mais relative à son objectif. » Ainsi pouvons-nous constater que la validité va au-delà de l'évaluation elle-même. Elle renvoie à la cohérence entre l'intention de l'examen et le contenu que l'on trouve dans l'évaluation. En somme, afin que l'instrument évaluatif soit valide, il est essentiel qu'il réponde à

l'usage pour lequel il a été conçu, et qu'il y ait une cohérence entre ses finalités et les contenus développés (Lussier & Denis, 1995 ; Tagliante, 2005).

Il est également possible de distinguer différents types de validité. Il y a d'abord la validité faciale ou superficielle qui fait référence à la validité globale de l'évaluation. Est-ce que sa forme et son organisation sont pertinentes par rapport au contexte ? Est-ce que les items sont adaptés au type d'évaluation et à son objectif ? McNamara (2000) explique la validité faciale comme l'acceptabilité faciale de l'examen par ceux qui sont impliqués dans sa conception, sa passation et sa correction.

Ensuite, nous pouvons évoquer la validité du contenu qui s'applique aux détails de l'évaluation. Il s'agit de se demander si le contenu linguistique est adapté au niveau des apprenants, à la situation, aux objectifs et au contexte (Lussier & Turner, 1995 ; Bonboir, 1972). De même, « On dit d'un test qu'il a une validité de contenu si les items ou les tâches dont il est composé constituent un échantillon représentatif des items ou des tâches pour une capacité ou un domaine de connaissance précis » (Noël-Johny & Sampsonis, 2006, p. 43). Pour s'assurer de la validité du contenu, il faut donc un éventail divers d'items, et le contenu doit être cohérent avec l'objectif et la finalité de l'évaluation.

Par ailleurs, au cours de mes lectures, j'ai constaté une différence dans l'explication de la notion de validité entre les publications datant d'avant le 21^{ème} siècle et celles suivant cette date. De fait, la définition est plus complexe de nos jours. Par exemple, si nous nous appuyons sur le *Manuel pour l'Élaboration et la Passation de Tests et d'Examens de Langue* conçu par ALTE (Association of Language Testers in Europe), nous remarquons une définition davantage approfondie. En effet, les auteurs incluent également l'interprétation des résultats et l'impact social de ces derniers comme des facteurs de validité. C'est d'ailleurs l'analyse des résultats par le candidat qui fait aussi preuve, ou non, de la validité du test. Cependant, ils tiennent à préciser qu'en partant de ce principe, aucun test n'est valide de manière absolue (ALTE, 2011). C'est pour cela qu'il est important de se rapporter également à la définition simple de la validité.

Enfin, je souhaite évoquer la validité telle qu'elle est décrite dans le CECRL étant donné l'importance de cet ouvrage dans l'enseignement des langues à l'heure actuelle. D'après cette référence, la validité d'un examen est en lien direct avec les niveaux décrits. Ainsi, un test de niveau A2 est valide lorsque ses contenus reflètent ceux du descripteur de ce niveau. Autrement dit, quand les résultats d'un test de positionnement donnent une représentation exacte du niveau de l'évalué, l'épreuve est valide (Conseil de l'Europe, 2001).

3.3. Étape 3 : le jugement

Une fois que les données sont recueillies en fonction de l'intention de l'évaluation et de sa finalité, il s'agit de porter un jugement sur l'atteinte ou non des objectifs d'apprentissage des

apprenants en examinant les résultats de mesure. En d'autres termes, c'est à cette étape que le formateur donne du sens aux données collectées lors de l'étape précédente. Comme Lussier (1992 : 29) le précise, une grande importance doit être accordée à cette étape. « Le jugement doit s'appuyer sur des faits, sur des observations recueillis en cours d'apprentissage et sur des résultats de mesure. Le jugement est d'autant plus important qu'il sert d'assise à la prise de décision. » Le jugement est en effet ce qui relie la mesure à la prise de décision. Sans lui, l'acte d'évaluer perdrait tout son sens, et c'est pour cette raison que c'est une étape clé du processus.

D'après Lussier (1992), il est important que l'enseignant se fixe des règles afin d'arriver à un jugement juste et précis. Ici, il est question de savoir ce sur quoi nous allons baser notre jugement. Il faut en effet des critères ainsi qu'une échelle d'appréciation pour classer les niveaux de performance possibles, et pour établir le seuil de performance attendue. À cette fin, l'enseignant doit utiliser ces critères pour déterminer le degré de maîtrise des apprenants de tel ou tel objectif d'apprentissage. Puis, en observant la performance de l'étudiant, l'évaluateur se réfère à ces critères pour situer l'apprenant parmi les différents niveaux de maîtrise. S'il s'agit de l'évaluation d'une performance, par exemple une production orale ou écrite, la détermination de critères est d'autant plus importante pour réduire la subjectivité. Cette citation de Lussier (1992, p. 31) met en évidence l'utilité et la nécessité des critères lorsque nous portons un jugement.

« Évaluer à partir des critères de performance est donc essentiel si l'on veut dans un premier temps éliminer autant que possible la part de subjectivité que comporte tout jugement et si l'on veut, dans un deuxième temps, cerner les difficultés des élèves. Il n'est pas nécessaire d'utiliser tous les critères à toutes les occasions d'évaluation. On peut porter son attention sur un ou deux critères d'une fois à l'autre avant des les englober tous au terme des apprentissages visés. »

Lussier explique ici que l'élaboration de critères pour juger de la qualité d'une performance est indispensable pour pouvoir évaluer l'apprenant de manière juste. Puis, grâce à ces critères, l'évaluateur arrive à porter un jugement, et à situer ensuite l'évalué sur une échelle d'appréciation. Cette échelle varie d'ailleurs selon les enseignants, le type d'évaluation et sa finalité. Lorsqu'il s'agit d'une évaluation selon le CECRL, par exemple, les niveaux A1-C2 peuvent constituer l'échelle d'appréciation.

En revenant à la citation ci-dessus, il est possible de relever deux enjeux importants de cette étape du jugement : l'objectivité et la subjectivité. Ces deux notions sont revenues à maintes reprises dans mes différentes lectures (Huver & Springer, 2011 ; Mothe, 1975 ; Horner, 2006 ; Lussier, 1992). C'est pour cela que je voudrais maintenant les évoquer pour essayer de cerner leurs avantages et inconvénients, ainsi que la manière de les contrôler.

3.3.1. L'objectivité et la subjectivité

L'objectivité et la subjectivité sont deux enjeux centraux lors de la phase de jugement, et méritent d'être explicités, d'autant plus qu'un nombre important de chercheurs les évoquent souvent. Mothe (1975), Huver et Springer (2011) et Horner (2012) leur accordent d'ailleurs une grande importance dans leurs écrits.

En premier lieu, je souhaite mettre en évidence deux paradigmes qui facilitent l'appréhension des notions d'objectivité et de subjectivité dans l'évaluation. Premièrement, il s'agit d'un paradigme mécanique selon lequel le monde est univoque, maniable par l'homme, décontextualisé et anhistorique (Huver & Springer, 2011). De plus, l'épistémologie qui la fonde est positiviste et quantitative. Selon une approche mécanique, l'évaluation est objective, quantitative et technique, et permet donc de découper les unités linguistiques, d'y porter un jugement (positif ou négatif), et enfin de regrouper ces unités pour déterminer un jugement globale (*ibid.*).

Deuxièmement, nous pouvons évoquer le paradigme holistique qui, au contraire, défend une vision du monde complexe, polysémique et en évolution perpétuelle (Huver & Springer, 2011). De même, l'épistémologie qui le fonde est constructiviste et qualitative, favorisant ainsi l'interprétation des résultats. L'évaluation est donc subjective (*ibid.*). Ces deux modèles montrent la polarité des épreuves objectives et subjectives, un fait dont il faut prendre conscience afin de cerner les avantages et les inconvénients de ces deux approches évaluatives.

Lorsque nous creusons la question de l'objectivité et de la subjectivité, nous nous demandons si l'une prime sur l'autre, quels en sont les avantages et les inconvénients, et par quels moyens les équilibrer. Comme le précisent Huver et Springer (2011), il existe une certaine tension entre ces deux notions. De fait, les enseignants de langue cherchent d'une part une évaluation quantitative, mécaniste et standardisée pour pouvoir porter un jugement objectif et, d'autre part, une évaluation qualitative, holistique et donc plutôt subjective qui est basée sur la performance. Hadji (1997), par exemple, soutient la nécessité d'atteindre à tout prix l'objectivité d'une évaluation, tandis que l'évaluation subjective, basée sur une performance, est d'ores et déjà fortement installée dans l'enseignement des langues grâce, en partie, à la perspective actionnelle (Horner, 2012). Ainsi remarquons-nous un certain paradoxe, puisque ces deux types d'examens s'opposent à la fois dans leurs formes et dans les résultats qu'ils fournissent (Huver & Springer, 2011).

Une évaluation objective donne des résultats fiables étant donné que la notation n'est pas soumise à des points de vue personnelles (Bonboir, 1975). Elle est ainsi associée à la procédure de notation, car c'est au moment de mesurer la maîtrise, ou non, des objectifs d'apprentissage que l'évaluateur essaie d'être le plus objectif possible (Mothe, 1972). Ainsi, si le but est d'assurer l'objectivité de l'évaluation, il faut qu'il existe clairement une seule bonne réponse, peu importe le correcteur (*ibid.*). C'est le cas avec les questions à choix multiples (désormais, « QCM ») puisqu'il

n'y a qu'une bonne réponse, et celle-ci est déterminée en amont de la passation. Il s'agit ici des tests standardisés comprenant des items fermés et souvent autocorrectifs, c'est-à-dire corrigés par machine (Tagliante, 2005). C'est ce type d'examen qui est d'ailleurs le plus utilisé dans le but d'atteindre l'objectivité, car il n'existe pas de désaccord dans la notion. De plus, il permet d'uniformiser les appréciations, éliminant de cette manière les différences de jugement (Bonboir, 1975). Un test standardisé limite toutefois l'objet de l'évaluation, car il sert principalement à évaluer des compétences isolées telles que les connaissances grammaticales ou lexicales, ou bien la compréhension (Noël-Johny & Sampsonis, 2006). La performance ne peut pas donc être évaluée par un test standardisé.

S'opposant à l'objectivité, la subjectivité entre en jeu lorsque c'est l'évaluateur qui doit décider lui-même de la bonne réponse, car les réponses peuvent varier selon les apprenants (Mothe, 1975). Il s'agit donc d'un jugement, de la part de l'évaluateur, à partir de la qualité d'une performance (Conseil de l'Europe, 2001). De même, avec une telle évaluation, la subjectivité est inévitable, puisque le jugement dépend de la personne qui observe l'apprenant (Horner, 2006). Autrement dit, l'on porte un jugement sur la performance en s'appuyant sur une impression issue de l'observation de la production langagière. Puis, étant donné les différentes façons dont chaque enseignant interprète ces observations, l'impression varie selon celui qui effectue la correction. Une évaluation de la performance est ainsi obligatoirement subjective.

Bien que l'on ne puisse pas éliminer complètement la subjectivité, il est possible de la réduire (Horner, 2006). À cette fin, il est nécessaire que l'évaluateur se fournisse en critères précis sur lesquels il s'appuiera afin de mesurer la qualité d'une performance, ainsi que d'une échelle d'appréciation pour y situer l'apprenant. Les critères font en effet l'objet de descriptions (ou indicateurs) du comportement attendu établies à partir des objectifs d'apprentissage (Bento, 2012) :

« Évaluer à partir des critères de performance est donc essentiel si l'on veut dans un premier temps éliminer autant que possible la part de subjectivité que comporte tout jugement... » (Lussier, 1992 : 31). Ainsi, l'évaluateur observe le comportement de l'apprenant et le compare aux critères afin de déterminer la qualité de la performance tout en étant le plus objectif possible grâce à ces critères. Quant à l'échelle, il s'agit d'un outil permettant d'apprécier l'apprenant en fonction des critères. Elle permet à l'évaluateur de situer l'utilisateur entre « pas de performance » et « une performance satisfaisante selon les objectifs d'apprentissage » (Lussier & Turner, 1995 ; Bonboir, 1972).

Cependant, il est important de préciser que, même avec l'utilisation d'une échelle d'appréciation et des critères, la subjectivité et la variabilité sont inhérentes à l'évaluation de la performance (Horner, 2006). Il est ainsi important de former les évaluateurs pour qu'ils soient tous

familiarisés avec les critères et pour qu'il y ait une homogénéité dans la façon dont ils portent un jugement (Horner, 2006 ; ALTE, 2011).

Il est vrai par ailleurs que la subjectivité est véritablement omniprésente y compris lors d'une évaluation objective. En choisissant les items, les textes ou les documents sonores pour un test standardisé, par exemple, c'est la vision du concepteur qui détermine la pertinence de la nature et de la difficulté des éléments du test. Cela montre que toute évaluation est influencée effectivement par le point de vue de celui qui conçoit l'épreuve. Il est toutefois possible de réduire à nouveau cette subjectivité en se basant sur des référentiels communs tels que le CECRL.

3.3.2. La fiabilité

La fiabilité, ou la fidélité (je considère que ces deux termes ont la même acception), est une autre caractéristique fondamentale d'une bonne évaluation. Comme l'objectivité et la subjectivité, cet enjeu est présent surtout lors de l'étape du jugement. C'est pour cela qu'il est cohérent de la mettre en évidence maintenant.

La fiabilité est de fait en lien avec la régularité des résultats obtenus pour une même évaluation lors de passations différentes. Un test est donc fiable si on obtient les mêmes données lorsque nous le faisons passer auprès de deux groupes de personnes différentes mais ayant les mêmes compétences (Hughes, 1989). De même, quel que soit le jour où l'on note une copie ou quelle que soit la personne qui effectue la notation, le jugement doit être le même. Ainsi, la fiabilité est respectée lorsqu'aucun élément extérieur ne vient influencer la notation (Noël-Johny & Sampsonis, 2006).

Afin de vérifier la fiabilité d'une évaluation, il est impératif de mettre en place un pré-test, c'est-à-dire une phase où l'on effectue une passation auprès d'apprenants de même niveau que ceux qui la passeront dans la situation réelle. En effet, cette étape permet au concepteur de s'assurer que le degré de difficulté des items est adapté au niveau, de confirmer la valeur discriminante des questions (le test permet-il de distinguer les apprenants forts des plus faibles ?) et enfin de certifier la pertinence des distracteurs s'il s'agit des QCM (Noël-Johny & Sampsonis, 2006). Enfin, la fiabilité peut être atteinte en respectant une norme. Dans le cas de l'enseignement des langues en Europe, les niveaux communs de référence du CECRL peuvent être utilisés comme la norme et facilitent la préservation de la fiabilité d'une évaluation.

Par ailleurs, Mothe (1975) nous rappelle que la fidélité peut être garantie au cours des trois grandes étapes de l'évaluation. Lors de la conception, nous pouvons intégrer le plus d'items possible et varier les types de tâches. Ensuite, pendant la passation, il se révèle nécessaire d'établir des conditions stables (pas de nuisance sonore, conditions climatiques correctes, bonne qualité des

enregistrements, par exemple). Enfin, en mettant en place un test autocorrectif et donc à base de QCM, on le rend davantage fidèle lors de l'étape de la correction.

D'après Doucet (2001), lorsque nous nous intéressons à la fiabilité, nous nous interrogeons sur l'origine des erreurs de mesure autres que celles issues de la performance de l'évalué. Autrement dit, pouvons-nous noter des facteurs non liés à la performance de l'élève qui ont fait que ce dernier a donné des réponses fausses ? Ainsi, la fiabilité peut être assurée, par exemple, en prêtant attention aux effets parasites de l'évaluation et en s'efforçant que la correction soit entièrement objective (Tagliante, 2005).

3.3.3. Les effets parasites de la fiabilité

Lors de l'étape du jugement, nous avons vu que les enjeux de l'objectivité, de la subjectivité et de la fidélité sont particulièrement présents. Quant à cette dernière, il existe des facteurs par rapport auxquels il faut être prudent, car ils risquent de nuire à la fiabilité de l'évaluation. De fait, il s'agit des effets parasites qui peuvent entrer en jeu, surtout lors d'évaluations subjectives quand l'enseignant doit mesurer lui-même la qualité de la performance et porter un jugement. Bien que ces facteurs ne puissent pas être entièrement éliminés, il est possible de diminuer leur présence en étant conscients et, de ce fait, de préserver la fidélité de l'examen. En m'appuyant sur les travaux de Taligante (2005), voici les différents effets parasites de la fiabilité :

- L'effet de contraste. Il s'agit de comparer des copies de qualité différente. Si l'on corrige une bonne copie juste avant une copie moyenne, cette dernière risque d'être évaluée plus sévèrement. Le contraire est également possible. Dans le but de réduire cet effet, il s'agit de rester fidèle aux critères et au niveau ciblé.
- L'effet d'ordre. Les recherches dans ce domaine montrent qu'en évaluant un nombre important de copies, l'enseignant a tendance à être plus sévère lorsqu'il note la dernière copie.
- L'effet de stéréotypie. Si l'enseignant connaît un certain apprenant et l'a déjà évalué plusieurs fois, il peut avoir du mal à lui donner une note différente que celles données auparavant. Sommairement, nous avons tendance à établir une image des compétences d'une personne donnée, et cette image nous empêche d'évaluer de manière objective.
- L'effet halo. Dans une situation d'examen, un apprenant se comportant bien, de manière polie et respectueuse, a tendance à être mieux évalué que celui qui se comporte moins bien ou qui montre sa lassitude ou son manque de confiance. Autrement dit, l'effet halo laisse les facteurs non liés à la performance de l'apprenant influencer son évaluation.
- L'effet choc (négatif ou positif). Lorsque nous évaluons la production d'un étudiant, et que nous voyons se produire à plusieurs reprises la même faute, nous sommes enclins à noter plus

sévèrement (choc négatif). À contrario, si nous constatons la bonne utilisation d'un élément linguistique parmi des erreurs, il est facile de se laisser séduire et de donner une bonne note (choc positif) grâce au bon usage de l'élément en question.

- L'effet « goutte d'eau » ou de débordement. Si l'évaluateur tolère la même faute produite à plusieurs reprises dans une production, la production de cette même faute une dernière fois à la fin de la production risque de « faire déborder la vase ». Ainsi l'enseignant finit-il par accorder une note basse.
- L'écart type. Chaque enseignant a une idée de ce qui est une bonne et une mauvaise note, et généralement les notes qu'il accorde s'inscrivent dans un certain espace. Par exemple, si l'on note uniquement entre 8 et 12, l'écart type est de 4 points. De même, certains enseignants ne donnent jamais une évaluation parfaite (20/20), alors que d'autres estiment qu'il est possible de donner cette note.
- L'effet de favoritisme. L'évaluateur peut avoir un élève « préféré », c'est-à-dire que cette personne possède une bonne image aux yeux de l'enseignant. Celui-ci a du mal donc à donner une mauvaise note à cet apprenant même quand c'est mérité.
- L'effet de contamination. Si l'enseignant sait qu'un apprenant donné à un bon dossier scolaire (il possède des diplômes, il a fait des études en lien avec le sujet de l'examen), l'évaluation peut être influencée de manière positive.
- L'effet de fatigue. Ce facteur est en lien avec la tâche évaluative elle-même, et l'effet qu'elle produit sur l'enseignant lorsqu'il doit corriger un nombre important de productions. La fatigue que suscite cette procédure nuit à la capacité de l'enseignant d'évaluer de manière fiable. De plus, il a été montré que l'on évalue mieux le matin, dans un endroit calme et propice au travail. Cela peut par ailleurs provoquer d'autres conséquences négatives, telles que l'effet de débordement, de contraste ou d'ordre.
- L'effet d'évaluation externe. Dans le but de préserver son image de bon enseignant vis-à-vis de l'institution ou des parents, l'évaluateur ne donne pas de très mauvaises notes.
- L'effet pygmalion. Il est question ici de stéréotypes ou d'images que l'enseignant possède des étudiants ayant certaines caractéristiques (origine ou milieu social, par exemple), et l'effet de ces représentations sur la réussite, ou non, de ceux-ci. La psychologie sociale démontre que si l'élève s'aperçoit que l'évaluateur possède un préjugé par rapport à sa capacité, il finira par confirmer cette image.

En examinant ces effets parasites de la fiabilité, nous remarquons deux choses importantes. Premièrement, il s'agit de facteurs présents lors des évaluations de la production orale ou écrite,

c'est-à-dire de la performance. Deuxièmement, ils s'appliquent à des examens dont la correction s'effectue par des humains et non par des machines. Nous pouvons ainsi faire le lien entre ces effets, la fiabilité, la subjectivité et l'objectivité. De fait, c'est dans le cadre d'une évaluation subjective que ces facteurs peuvent avoir un impact sur le jugement, étant donné que l'évaluateur est chargé de se prononcer sur la qualité de la production. S'il s'agit d'un test standardisé et autocorrectif, par exemple, ces effets n'entrent presque pas en jeu. La conclusion que nous pouvons en tirer est que la fiabilité est donc menacée surtout lors d'évaluations subjectives, et c'est avec ce type de situation qu'il faut être prudent quant aux effets parasites de la fiabilité.

3.4. Étape 4 : la décision

La quatrième et dernière étape de la procédure évaluative est la prise de décision. Celle-ci est menée d'ailleurs en fonction des jugements portés dans l'étape précédente. Ainsi pouvons-nous encore une fois constater le lien entre les différentes étapes de cette procédure.

Comme l'évoque Lussier (1992), la décision est en lien avec l'apprentissage de la langue. D'après cet auteur, il existe deux types de décisions : celle qui a pour but d'assurer la progression des apprentissages par rapport à l'enseignant et sa démarche d'enseignement. Ce type de décision incite, par exemple, à des exercices de renforcement, d'entraînement, d'enrichissement, ou à des activités complémentaires en fonction des résultats des apprenants et de leurs points forts et faibles. Ainsi la décision porte-elle sur des éléments précis de l'apprentissage afin de répondre aux besoins de la classe. De même, elle peut entraîner une transition vers une nouvelle séquence d'apprentissage à l'intérieur du programme global.

Quant au deuxième type de décision, il se rapporte davantage à l'ensemble des apprentissages, et indique le degré de maîtrise ou non des objectifs d'apprentissage du programme. Ainsi entraîne-t-il des changements par rapport à l'institution, au passage, ou non, à un niveau supérieur, à un positionnement dans des groupes de niveau ou à l'obtention d'un diplôme ou d'un certificat. De cette manière, il est possible de constater que le type de décision à prendre est déterminé en fonction de l'intention, c'est-à-dire de la finalité de l'évaluation.

Enfin, lors de l'étape de la prise de décision, l'enseignant doit prendre en compte l'impact de l'évaluation. Autrement dit, quelles conséquences peut-elle entraîner, et est-il possible de remédier à ces conséquences ? L'impact d'un test de positionnement est ainsi généralement faible puisque la décision peut être modifiée en cas de placement dans un groupe de niveau incorrect. En revanche, les évaluations sommative et certificative (DELFI / DALFI) font l'objet d'évaluation ayant un impact important étant donné que la décision prise suite à l'évaluation est définitive.

4. *Les items possibles dans un test autocorrectif*

Il existe plusieurs types d'items possibles à mettre en place dans le cadre d'un test autocorrectif en ligne. En fonction de l'objectif de la question, ces items permettent d'évaluer les différentes connaissances des étudiants. Opposée aux questions ouvertes qui demandent une production écrite ou orale élaborée à partir d'une consigne, les items d'un test autocorrectif sont de type fermé puisque la personne doit simplement répondre en choisissant ou en écrivant l'unique bonne réponse. Il est important de connaître ces outils afin de pouvoir varier les items d'un test de positionnement. Afin d'expliquer les différents item, je m'appuie sur l'ouvrage de Christine Tagliante intitulé *L'évaluation* (1991). Voici donc les types d'items fermés d'après cet auteur :

- Le texte lacunaire. Appelé également « texte à trous », cet exercice est souvent employé pour évaluer les savoirs linguistiques grâce à sa capacité à tester uniquement les connaissances visées. Ainsi représente-il une question dont nous pouvons être sûrs de la validité. De fait, il demande à l'apprenant de compléter un texte à partir d'un choix de mots.
- Le test de closure. Cet outil est élaboré à partir d'un texte dont nous enlevons un mot sur cinq pour évaluer la compréhension écrite. De plus, comme Tagliante le précise, l'avantage de cet item est qu'il peut montrer les erreurs des apprenants non-liées à la compréhension.
- Le test d'appariement. De manière générale, ce test demande à l'apprenant de relier les objets de deux listes juxtaposées. Tagliante spécifie que ce type de question peut être utilisé pour tester la capacité de l'apprenant à reconnaître la cohérence d'une phrase complexe en reliant deux parties d'un énoncé scindé, par exemple.
- Le test de classement. Cette forme de question permet de tester la compétence de lecture sélective. Il sert en effet à savoir si l'étudiant est capable de repérer les parties indicatrices de la cohérence d'un discours.
- Le QROC (question à réponses ouvertes courtes). Il s'agit d'un test de production minimale de la part de l'étudiant. Par exemple, « Que dites-vous à quelqu'un qui vient de perdre son travail ? ».
- Le texte induit. Ce type de question incite une production courte de la part de l'apprenant à partir d'une consigne contraignante et d'un document. L'exemple typique d'un test à partir d'un texte induit est, « Racontez la semaine de monsieur X en observant son emploi du temps ».
- Les QCM. Les « questions (ou questionnaires) à choix multiples » peuvent varier considérablement en terme de forme (nombre de choix, vrai / faux, pour / contre, par exemple), et oblige l'évalué à choisir une seule réponse parmi plusieurs. Réservé à la vérification des connaissances, des transferts, et à la compréhension, ce type d'item ne sert jamais à évaluer l'expression orale ou écrite. Économiques en terme de la correction, les QCM exigent toutefois

un temps important de rédaction et plusieurs révisions afin de vérifier la pertinence des distracteurs. C'est ce que nous allons voir maintenant.

4.1. Les QCM en détails et les enjeux liés

Après avoir vu les types de questions utilisées dans le cadre d'un test autocorrectif, je voudrais évoquer de manière détaillée les QCM car c'est ce type de question qui est le plus utilisé dans le cadre d'une épreuve autocorrective. De plus, connaissant les enjeux de l'objectivité, de la fiabilité et de la validité, ainsi que l'importance qui doit leur être accordée, il est pertinent de faire le lien entre ces différentes caractéristiques d'une « bonne » évaluation et les QCM. Ici, l'objectif est en effet d'analyser dans quelles mesures ce type de question peut être utilisé si l'on souhaite préserver la qualité de l'épreuve ?

Les recherches sur l'évaluation sont fortement concentrées sur les QCM dans l'évaluation objective en langue. Les enseignants s'interrogent d'ailleurs souvent sur l'utilité, ainsi que les limites, de cet outil. En effet, bien que cette forme d'item soit souvent critiquée, il est généralement accepté que les QCM aient leur place dans l'évaluation en langue. Il est tout simplement question de prendre des précautions lors de l'élaboration de ces items pour ne pas nuire à la validité, ni à l'objectivité ou la fiabilité.

Concernant leurs usages, les QCM sont généralement utilisées pour évaluer la compréhension orale et écrite, ainsi que pour tester des connaissances précises sur la langue telles que la grammaire, le lexique et la syntaxe dont nous avons des difficultés à déterminer la maîtrise autrement (Mothe, 1975). Cela montre l'intérêt des QCM dans le cas où nous voulons interroger l'apprenant sur un élément spécifique sans nous intéresser à d'autres connaissances.

Dans son ouvrage, Mothe (1975) évoque en détails la question des QCM, y compris leurs avantages et leurs inconvénients. C'est pour cela que je m'appuie sur ces recherches pour mettre en évidence les enjeux liés à leur utilisation.

4.1.1. Les QCM et la validité

Comme tout type de questions, il est vrai que les QCM peuvent porter atteinte à la validité d'une évaluation. Un concepteur peut avoir tendance, par exemple, à choisir les questions pour leur nature pratique et non pour leur représentativité de l'enseignement, réduisant ainsi la validité. Il faut donc prêter davantage attention à ce que les questions ne soient pas basées seulement sur les objectifs d'apprentissage ciblés par le cursus si l'on veut préserver leur validité. De même, lors de la rédaction des items, il est important de veiller à ce que la question ne pose pas de problème de compréhension. Si tel est le cas, et que l'apprenant se trompe, nous ne pouvons pas identifier la cause de la mauvaise réponse. Par ailleurs, des QCM portant sur la grammaire et contenant du métalangage peu familier

pour l'apprenant, par exemple, peuvent aussi provoquer des problèmes de compréhension, altérant donc la validité. Malgré ces risques, il est généralement accepté qu'une évaluation sous forme de QCM est valide :

« Malgré ces limites dans l'évaluation de la compréhension, le QCM reste néanmoins un instrument de mesure plus valide que d'autres procédés comme la paraphrase, qui pose le délicat problème des invariants (...) ou le résumé de texte (qui)...contraint l'examineur à renoncer à prétendre atteindre l'objectivité de la notation. » (Mothe, 1975 : 65).

4.1.2. Les QCM, l'objectivité et la fiabilité

Lorsque notre objectif principal est de préserver la fiabilité et l'objectivité d'une évaluation, nous avons souvent recours aux QCM, car ce type d'item ne laisse pas intervenir de correction humaine forcément plus subjective. Qui plus est, une correction entièrement objective n'est envisageable qu'avec un test élaboré à partir de QCM (Tagliante, 2005). Ainsi pouvons-vous constater le lien entre une notation de QCM et la fidélité, puisque la notation est stable et constante quelles que soient la situation évaluative et ses conditions (*ibid.*).

De même, à condition que les QCM soient bien rédigées et compréhensibles par tous, les résultats des apprenants seront stables, ce qui favorise la fiabilité de l'évaluation. De fait, une question ouverte laisse la possibilité à l'apprenant d'esquiver la réponse attendue étant donné qu'il peut répondre de n'importe quelle manière (Mothe, 1975). Au contraire, un QCM oblige l'évalué à choisir une réponse parmi un choix limité. Les résultats sont donc plus fiables, car tous ceux qui passent l'évaluation ont les mêmes choix de réponses à leur disposition.

4.1.3. Les QCM, le hasard et la divination

Les QCM sont souvent critiquées comme étant trop faciles à cause du hasard et de la divination. Cela est d'ailleurs confirmé par le fait que les résultats des tests utilisant des QCM sont généralement plus élevés que ceux ayant des questions ouvertes (Mothe, 1975). Afin que les QCM ne soient pas trop faciles, il existe des techniques qui réduisent le rôle du hasard et de la divination. Je vais donc en citer quelques exemples.

Il est possible de rendre plus difficiles des questions type QCM en s'efforçant d'élaborer des distracteurs plausibles et homogènes sur le plan linguistique, ce qui réduit la possibilité de deviner la réponse correcte. Sa place parmi les autres choix doit être complètement aléatoire, les distracteurs doivent être de la même longueur et de la même nature formelle pour éviter que l'étudiant ne la choisisse pour des raisons non-linguistiques.

Pour évoquer plus en détails le rôle du hasard et de la divination avec les QCM, je voudrais exposer le tableau de Mothe ci-après (1974). Les données mettent en évidence la possibilité, en pourcentage, d'obtenir telle ou telle note en répondant au hasard ou en devinant dans un examen de

QCM à deux choix et qui comprend cinq, dix ou vingt items. Ce chercheur souligne que le hasard est en lien direct avec le nombre de réponses possibles (deux distracteurs laissent 50% au hasard ou à la divination, tandis que quatre distracteurs n'en laissent que 25%), ainsi que le nombre d'items du test. Quant à la quantité de questions, le tableau ci-après montre à quel point ce facteur peut réduire la possibilité de réussir un examen constitué de QCM en devinant.

Note au moins égale à : Nombre d'items	5/10 ou 10/20		3/5 ou 6/10 ou 12/20		7/10 ou 14/20		4/5 ou 8/10 ou 16/20		9/10 ou 18/20		5/5 ou 10/10 ou 20/20	
	5	-	50	-	19	-	3					
10	63	38	17	5,5	1	0,01						
20	61	26	6	0,6	0,02	0,0001						

Figure 1 : Le rôle du hasard et de la divination dans une épreuve sous forme de QCM (Mothe, 1975 : 50).

Nous voyons ici que la possibilité de réussir en répondant de façon aléatoire se réduit de manière importante au fur et à mesure que le nombre d'items augmente. Mothe (*ibid.*) précise qu'à droite du trait gras, le rôle du hasard est négligeable et qu'il n'influence plus les résultats que de manière peu significative. Examinons le deuxième tableau qui met également en évidence l'utilité d'augmenter le nombre de choix par QCM pour réduire le rôle du hasard.

Note au moins égale à : Nombre d'options E (x)	1 2 3 4 5 6 7 8 9 10										
	2	5	99,9	99	95	83	63	38	17	5,5	1
3	3,33	98	90	70	44	21	7,6	2	0,4	0,04	0,002
4	2,5	94	76	47	22	8	2	0,3	0,04	0,003	0,0001
5	2	89	62	32	12	3,3	0,6	0,09	0,008	0,0004	0,00001

Figure 2 : Le rôle du hasard dans une épreuve sous forme de QCM (Mothe, 1975 : 51).

Ce deuxième tableau démontre à nouveau la facilité avec laquelle il est possible de réduire de manière importante le rôle du hasard en augmentant le nombre de distracteurs des items. Il s'agit de la possibilité, en pourcentage, d'obtenir telle ou telle note sur dix dans une épreuve de dix QCM ayant deux, trois, quatre ou cinq distracteurs chacune en répondant par hasard. Comme l'auteur le précise, à droite du trait gras le rôle du hasard est négligeable, ce qui justifie le rôle des distracteurs qui réduisent la possibilité que le hasard ait une influence importante sur les résultats.

4.2. Conclusions sur les QCM

En définitive, il est vrai que le hasard et la divination ne peuvent pas être totalement éliminés, mais nous avons vu qu'il est toutefois possible de réduire suffisamment ces facteurs pour rendre fiable et valide un test élaboré à partir de QCM. En revanche, il ne faut pas sacrifier la qualité d'un test en essayant à tout prix de réduire le rôle du hasard et de la divination. Il serait possible, par exemple, de porter atteinte à la validité d'une évaluation en intégrant un quatrième distracteur pour réduire le rôle du hasard. Or, si ce dernier choix ne respecte pas les règles des distracteurs, son intégration dans l'item réduit la qualité de l'épreuve. Il est donc impératif de viser la réduction du hasard et de la divination tout en préservant la validité et la fiabilité des QCM.

5. L'histoire de l'évaluation en langue : de 1950 à aujourd'hui

Puisque nous avons pris conscience du processus d'évaluation en langue, des enjeux liés et des différents types d'évaluations, il est maintenant possible d'examiner l'histoire de l'évaluation en langue. De fait, il est nécessaire de connaître les spécificités de l'évaluation pour pouvoir bien cerner son évolution au fil du temps. De même, il me semble important de connaître cette évolution, d'une part pour mieux comprendre les démarches évaluatives actuelles et, d'autre part, pour cerner l'utilité et les finalités de toutes formes d'évaluations.

À l'origine, il n'existait pas de recherche significative en l'évaluation, ce qui montre à nouveau la place marginale accordée à ce sujet. Avant les années 1970, la notion d'évaluation avait plutôt mauvaise réputation, elle était considérée comme un mal inévitable (Noël-Johny & Sampsonis, 2006). Il a donc fallu attendre les années 1950 pour que l'évaluation occupe une place centrale dans le champ de l'enseignement et de l'apprentissage des langues étrangères, et pour que l'image associée tende vers le positif. Cela dit, c'est principalement dans les pays anglophones que ces recherches ont commencé à émerger. En France, ce n'est qu'à partir de la naissance de l'approche communicative, dans les années 1970, que les didacticiens ont commencé à se pencher de façon significative sur les techniques évaluatives en langue (Huver & Springer, 2011).

Nous pouvons donc constater une certaine négligence en matière de pratiques évaluatives jusqu'aux années 1960. Avant cette date, les didacticiens en langues étrangères s'intéressaient plutôt à l'apprentissage de celles-ci, et ignoraient leur évaluation (Huver & Springer, 2011). De même, il manquait un lien entre l'apprentissage et l'évaluation, un lien dont nous reconnaissons aujourd'hui la forte importance (Noël-Johny & Sampsonis, 2006), et qui est effectivement maintenu en respectant les objectifs d'apprentissage ciblés par l'enseignement.

Bien que les différentes périodes de l'évaluation en langue se chevauchent, et que certains chercheurs en distinguent plus que d'autres, il est possible d'en évoquer trois principales : la période psychométrique structuraliste, la période communicative métacognitive et la période actuelle. Je

considère que ce sont les trois périodes essentielles dans l'évaluation en langues, puisque l'on peut aisément voir les différences entre elles, ainsi que les éléments qui les ont caractérisées. Certains chercheurs évoquent une période nommée psycholinguistique-sociolinguistique qui s'insère entre les deux premières périodes, mais dans le cadre de ce mémoire il n'est pas nécessaire de la distinguer étant donné qu'elle ressemble à la période psychométrique structuraliste.

5.1. 1950 à 1970 : la période psychométrique structuraliste

À partir des années 1950, nous commençons à noter les premiers grands repères historiques concernant l'évaluation en langue (Huver & Springer, 2011). De fait, la scolarisation devenant obligatoire dans un nombre croissant de pays, l'évaluation est devenue nécessaire pour faciliter l'entrée dans des structures académiques, notamment par la création de concours et de tests d'entrée (Lussier & Turner, 1995).

La période psychométrique structuraliste, appelée par certains chercheurs comme Huver & Springer (2011) tout simplement « la période structuraliste », se limite entre les années 50 et 70. En termes d'évaluation, elle est caractérisée, d'une part par une approche structurale de la langue, c'est-à-dire que l'on s'intéressait à l'organisation de la langue et à ses composantes. D'autre part, l'on voulait évaluer les compétences cognitives des utilisateurs, ce qui renvoie au terme « psychométrie » qui signifie effectivement « la science de la mesure des compétence cognitives » (McNamara, 2000).

Influencée de manière importante par l'ouvrage de Robert Lado *Language Testing*, un ouvrage publié en 1961 qui a révolutionné la notion d'évaluation en langue, cette période avait l'avantage d'évaluer de manière isolée les connaissances linguistiques de l'utilisateur. D'ailleurs, et comme l'évoque McNamara (2000), l'évaluation portait sur les connaissances grammaticales, lexicales et phonétiques, et ces éléments étaient évalués indépendamment les uns des autres. Lado employait le terme anglais *discrete point testing* pour désigner cette technique évaluative, car l'on s'intéressait justement à des points spécifiques de la langue.

Par ailleurs, durant cette période les concepteurs d'évaluation s'efforçaient de créer des tests davantage scientifiques, précis et objectifs (Denis & Lussier, 1995). Ainsi, l'on employait souvent des questions fermées telles que les QCM afin d'évaluer objectivement les compétences de l'utilisateur, et par là-même, la correction par machine s'est imposée. Ce type d'épreuve est encore fréquemment utilisé aujourd'hui, même si la diversité des évaluations en langue continue à se développer. Rappelons-nous la mention de l'évaluation des connaissances dans le CECLR, un examen qui se rapproche de l'évaluation utilisée lors de la période psychométrique structuraliste.

Pendant cette période, certaines méthodologies sont nées qui ont considérablement influencé l'évaluation en langue, notamment la méthodologie structuro-globale audiovisuelle. Ces stratégies d'enseignement ont donné lieu à une pédagogie basée sur l'analyse contrastive de la langue cible et

de la langue première, la mémorisation, la répétition et les exercices structuraux, sur les règles grammaticales de la langue étrangère et la traduction (Denis & Lussier, 1995). Le but dans l'enseignement et l'apprentissage des langues étrangères à l'époque était en effet de systématiser des automatismes chez l'apprenant afin qu'il reproduise la langue cible dans d'autres situations (ibid.). Au vu de cette approche, il n'est pas étonnant que l'évaluation de cette période ait été centrée sur la structure de la langue en mesurant le degré de maîtrise grammaticale et lexicale.

5.2. 1970 à 1990 : la période communicative métacognitive

À partir des années 1970, une évolution importante dans les méthodologies de l'enseignement des langues en Europe a bouleversé les pratiques évaluatives utilisées lors de la période psychométrique structuraliste. Ce changement a également donné lieu à une nouvelle période concernant l'enseignement des langues étrangères. Il s'agit de l'émergence de l'approche communicative, une méthodologie née d'une nouvelle vision de l'usage des langues dans la vie courante, ainsi que du rôle de l'apprenant. Selon cette perspective de l'apprentissage de la langue cible, l'apprenant est placé au cœur de son apprentissage et en assume une partie. Il devient donc plus autonome dans son acquisition de la langue cible. De plus, ses besoins, motivations et attentes pour apprendre sont pris en compte. Concernant l'apprentissage de la langue à cette époque, les chercheurs se désintéressaient de la structure grammaticale au profit de l'utilisation de la langue dans des situations de communication. Autrement dit, le développement des compétences globales de communication était désormais privilégié. La langue devenait ainsi un outil d'interaction ayant pour but de réussir la communication entre plusieurs locuteurs.

Comme l'enseignement avait subi des évolutions suite à l'apparition de l'approche communicative, l'évaluation en langue s'est également transformée. Portant jusque-là sur des éléments linguistiques et spécifiques de la langue, l'évaluation s'est alors orientée vers un contexte particulier dans lequel l'apprenant devait faire preuve de capacité à communiquer. De plus, les didacticiens se sont davantage intéressés à l'usage des stratégies métacognitives pour s'approprier des compétences, et la sociolinguistique a joui d'une nouvelle attention puisque la mesure de cette compétence s'intégrait de plus en plus dans les évaluations. Aussi cette période a-t-elle été baptisée « communicative métacognitive » (Denis & Lussier, 1995).

Les processus d'évaluation se sont clairement tournés vers une vérification de la capacité de l'apprenant à effectuer des tâches de communication. Parallèlement, l'enseignement de la langue portait alors sur la réalisation des tâches langagières, tout comme l'évaluation. D'après Huver & Springer (2011), cette nouvelle forme d'évaluation était fondée sur trois éléments : l'analyse des besoins de l'apprenant, les enjeux de l'évaluation et l'auto-évaluation. Étant donné la nouvelle importance accordée à l'autonomie de l'apprenant, il n'est pas surprenant que l'analyse de ses

besoins soit entrée en jeu lors de cette période. De même, afin d'impliquer la personne directement dans son apprentissage, il était jugé avantageux qu'il se rende compte explicitement de ses compétences par le biais d'une auto-évaluation. Enfin, les enjeux de l'évaluation en langues ont été mis en évidence, et les notions de validité et de fiabilité, plutôt ignorées pendant la période psychométrique structuraliste, sont alors devenues centrales dans la conception et l'élaboration de tests de langues (Denis & Lussier, 1995 ; Huver & Springer, 2011).

Toutefois, il est important de noter une certaine contradiction concernant l'histoire de l'évaluation en langues. En effet, comme nous l'avons vu précédemment, en évaluant la performance d'un étudiant, on se rapprochait forcément d'une évaluation subjective et variable. Sachant que cette période accordait de plus en plus d'importance à la validité et à la fidélité, on peut comprendre qu'une tension soit alors apparue puisqu'une évaluation plus subjective réduit obligatoirement la fidélité (Denis & Lussier, 2011). Ainsi, cette période a-t-elle été marquée par une certaine tension entre l'objectivité et la subjectivité, et entre la fiabilité et la validité d'une évaluation en langue. Afin d'assurer la fiabilité d'un examen, il est préférable de se servir d'outils objectifs comme c'est le cas des évaluations standardisées comprenant des items fermés. À contrario, dans le but de préserver la validité d'une épreuve, il faut mesurer à travers des tâches de communication, ce qui donne lieu à une évaluation forcément plus subjective (Huver & Springer, 2011). Pour le concepteur d'évaluation, cette opposition est importante à prendre en compte pour décider quel type d'évaluation est pertinent pour le contexte, le public et, bien sûr, la finalité de l'évaluation à concevoir.

5.3. La période actuelle : 1990 au présent

La dernière période que je souhaite évoquer est caractérisée par deux événements qui ont apporté des changements importants à l'évaluation en langue. Il s'agit de la création de ALTE (Association of Language Testers in Europe) en 1992, et de la publication du CECRL en 2001.

Premièrement, ALTE est née d'une initiative du monde de l'entreprise, ainsi que des travaux du Conseil de l'Europe (Noël-Johny & Sampsonis, 2006). De ce fait, cet organisme est étroitement lié au CECRL et base sa méthodologie de l'élaboration d'examens en langues sur les pratiques et les niveaux de références évoqués dans ce document. Puis, en 2011, ALTE a publié un document détaillant le processus d'élaboration d'évaluation en langues, ainsi que les enjeux associés. L'objectif était de fait de définir les niveaux de compétences pour que les évaluations soient toutes reconnues en Europe, et d'établir des normes évaluatives communes. Il s'agit du *Manuel pour l'Élaboration et la Passation de Tests et d'Examens de Langue*, document sur lequel je me suis appuyé pour concevoir le test de positionnement de GEM, et dont je détaillerai la méthodologie dans la troisième partie de ce mémoire.

Deuxièmement, la publication du CECR, tout comme la création de ALTE, a aussi provoqué des changements importants dans le domaine de l'évaluation. D'ailleurs, son titre évoque explicitement la pratique évaluative (« apprendre, enseigner, évaluer »). Les auteurs de ce document de référence ont détaillé treize pratiques évaluatives¹³, ainsi que les principaux enjeux à prendre en compte en évaluant les compétences langagières des apprenants. Le CECRL aide donc les enseignants à prendre conscience de la notion d'évaluation, et de ce fait à évaluer de manière davantage pertinente, efficace et juste.

De même, non seulement le CECRL explique les différents types d'évaluation ainsi que leurs caractéristiques, mais il introduit également une nouvelle méthodologie : la perspective actionnelle. Selon cette dernière, les tâches de la vie quotidienne, de communication et d'interaction, sont les éléments sur lesquels on se base pour organiser son enseignement. Selon cette méthodologie, une tâche s'applique à divers domaines (personnel, professionnel, éducationnel, ou public), et exige l'engagement stratégique des différentes compétences pour mener un ensemble d'actions. Ces actions aboutissent ensuite à la réalisation d'un produit final (Conseil de l'Europe, 2001).

Avec cette nouvelle méthodologie, les pratiques évaluatives ont évolué. De fait, l'évaluation était intimement liée aux objectifs d'apprentissage, comme nous l'avons vu précédemment¹⁴. Elle porte dorénavant uniquement sur la performance, c'est-à-dire que l'on s'intéresse à la capacité de l'apprenant à réaliser des tâches communicatives et authentiques de la vie réelle, et non plus à des connaissances précises sur la langue. « ...dans cette perspective, il ne s'agit plus de poser des questions de connaissances, de vérifier le savoir mais d'observer les élèves dans l'accomplissement d'une tâche, de porter un jugement sur les compétences mises en œuvre pour comprendre et réussir une tâche complexe » (Noël-Johny & Sampsonis, 2006, p. 109).

Il ne s'agit donc plus de mesurer l'utilisation de la langue de manière morcelée et hors contexte, mais d'évaluer la capacité à communiquer (Bento, 2012). Cependant, il se révèle important de mettre en évidence un autre changement incité par le CECRL et la perspective actionnelle en lien avec le jugement que nous portons sur la performance. En effet, c'est lors de cette période que l'usage d'une grille d'évaluation et de critères précis décrivant le comportement attendu devient impératif afin de guider la mesure et le jugement (Bento, 2012). Il est vrai que cette notion de grille de notation existait avant la publication du CECRL, mais avec ce dernier on a commencé à accorder plus d'importance à cette technique pour évaluer la performance.

¹³ cf. partie 2 (2.5.)

¹⁴ cf. partie 2 (2.4.)

5.4. Remarques sur l'histoire de l'évaluation en langues

En étudiant l'évolution des pratiques évaluatives en langues des années 1950 jusqu'à nos jours, nous pouvons faire quelques constats qui aident à mieux comprendre la situation actuelle de l'évaluation en langue.

D'une part, nous observons une tendance vers la subjectivité entre les périodes psychométrique structuraliste communicative car nous nous intéressons davantage à la performance de l'apprenant. D'autre part, en accordant une plus grande importance à l'usage des grilles de notation et des critères, l'évaluation reste certes plutôt subjective, mais tend vers un équilibre entre la subjectivité et l'objectivité grâce à ces outils. Par ailleurs, les enjeux clés de ce domaine, à savoir la validité et la fiabilité, sont de plus en plus valorisés entre la période psychométrique structuraliste et la période actuelle. De nos jours, ces notions restent centrales pour les concepteurs d'examen, montrant ainsi l'importance de ces caractéristiques.

Il est également possible de relever le fil conducteur des changements des examens en langue des années 1950 à aujourd'hui, c'est-à-dire l'élément qui a dirigé l'évolution des types d'examens. En effet, à partir des années 1950, ce sont les méthodologies de l'enseignement des langues qui influençaient majoritairement la nature des évaluations. Par exemple, la méthodologie structurale audiovisuelle a donné lieu à des tests de langues qui mesuraient des compétences précises (grammaire, lexique et phonétique) à travers des questions type QCM, et le *discrete point testing* de Lado. Ensuite, lorsque l'approche communicative est née, l'examen en langue s'est dirigé vers une évaluation de la performance et de la capacité de réussir la communication à travers des tâches. Enfin, l'évaluation a de nouveau évolué avec la publication du CECRL, l'émergence de la perspective actionnelle et une approche holistique de l'évaluation à travers des grilles et des critères précis. Ainsi pouvons-nous noter une forte corrélation entre les méthodologies et l'évaluation en langue, une concordance beaucoup moins présente lors de la période psychométrique structuraliste.

Le maintien de cette relation proche entre la méthodologie d'enseignement et l'évaluation est revendiqué par certains chercheurs comme Lussier (1992), tandis que d'autres, notamment Huver & Springer (2011), se méfient d'un appui uniquement centré sur les méthodologies d'enseignement pour déterminer les caractéristiques de l'évaluation. D'après ces derniers, les enjeux relatifs à l'évaluation relèvent non seulement de la méthodologie d'enseignement, mais également des pratiques sociales, institutionnelles et politiques liées à la société dans laquelle s'insère l'acte évaluatif. « Ainsi, les outils et les pratiques de l'évaluation, au-delà des méthodologies dont ils relèvent, sont à la croisée de dimensions axiologiques individuelles, sociales, politiques, lesquelles sont elles-mêmes traversées par l'ambiguïté originelle du terme évaluation. » (Huver & Springer, 2001 : 9).

Il reste donc aux enseignants à déterminer dans quelle mesure il faut baser l'évaluation sur la méthodologie utilisée en cours en tenant compte des objectifs d'apprentissage, et à quel point il est nécessaire de prendre en compte le contexte, le public, l'impact de l'examen et les caractéristiques de la société lors de sa conception.

Partie 3. La mise en place d'un dispositif hybride de positionnement à GEM

Suite à la présentation du contexte de stage et l'explication des théories pertinentes pour l'évaluation en langue, le projet d'ingénierie que j'ai monté va être maintenant détaillé. Pour rappel, cette mission a été guidée par la question suivante : comment élaborer un dispositif de positionnement hybride et en français pour des étudiants étrangers inscrits en école de management ? J'ai nommé le projet ainsi étant donné que la mission a évolué, de la conception d'un test de positionnement autocorrectif à l'élaboration d'un guide destiné aux évaluateurs pour le test oral, et d'une auto-évaluation. En se servant de ces trois outils évaluatifs pour cerner les compétences en français des étudiants à GEM, l'école dispose d'un dispositif hybride de positionnement pour placer les étudiants dans les groupes de niveau approprié.

Dans le cadre du nouveau processus de positionnement, les étudiants commencent par passer le test écrit en ligne, à distance, pendant les trois semaines qui précèdent leur arrivée à Grenoble. Ce test comprend d'une part la partie autocorrective sous forme de QCM et d'exercices de compréhension écrite et, d'autre part, une production écrite dont le sujet et la difficulté dépendent du niveau donné par le test autocorrectif que l'ordinateur corrige. Ensuite, le jour de l'entretien oral, qui a lieu lors des premiers jours de la rentrée, les étudiants remplissent une auto-évaluation avant de passer le test oral avec un enseignant de GEM. Ainsi, au moment de terminer celui-ci, l'évaluateur s'appuie sur le niveau donné par le test en ligne, la production écrite et le niveau fourni par l'entretien, ainsi que par l'auto-évaluation, pour déterminer le niveau de langue de l'étudiant et le positionner dans le groupe de niveau approprié.

Pour commencer, j'évoquerai les étapes par lesquelles je suis passé pour analyser le test existant, et pour ensuite concevoir le nouveau test de positionnement autocorrectif. Deuxièmement, je mettrai en évidence les étapes de l'élaboration de la fiche pour le test oral. J'expliquerai enfin la conception et l'utilisation de l'auto-évaluation.

1. L'élaboration du nouveau test de positionnement

Le processus de conception du nouveau test de positionnement a comporté plusieurs étapes. Premièrement, j'ai récolté des données à travers des analyses des tests en ligne, des questionnaires et des entretiens. J'ai étudié, en premier lieu, un échantillon de tests de positionnement afin de connaître les caractéristiques de ce type d'épreuve en fonction de l'institution pour laquelle ils avaient été conçus. Ensuite, j'ai mené une analyse minutieuse du test existant pour repérer ses points forts et faibles, ce qui m'a aidé à élaborer le nouveau test. J'ai également interrogé les enseignants et

les étudiants de GEM sur le test original par l'intermédiaire de questionnaires pour savoir comment adapter le nouveau à leurs attentes et besoins.

J'ai fait passer en outre deux entretiens afin de me préparer à entreprendre la mission. Le premier a été réalisé auprès d'une enseignante du CUEF de Grenoble qui a participé à la conception de leur nouveau test de positionnement. J'ai voulu en effet connaître les caractéristiques de ce dernier étant donné qu'il s'effectue en ligne et à distance. Le deuxième entretien a été effectué auprès de la responsable du programme de FLE à GEM. L'objectif était d'obtenir des précisions sur l'origine de la commande, sur le public et sur le test existant. Au fur et à mesure de l'explication de l'élaboration du dispositif de positionnement, je m'appuierai sur ces échanges afin d'expliquer et de justifier les décisions prises pour réaliser ce projet.

Grâce à cette récolte de données, j'ai pu mettre en place un test de placement adapté au public et qui répond à un besoin spécifique du programme de FLE à GEM¹⁵.

1.1. L'analyse d'un échantillon de tests de positionnement

Lorsque la tâche de concevoir un test de positionnement autocorrectif m'a été confiée, je me suis tout d'abord interrogé sur les particularités de ce type d'épreuve. Quels types d'items comprend un test de placement ? Combien de questions contient-il ? Quelles compétences sont évaluées ? Afin d'obtenir les réponses à ces questions, j'ai analysé neuf tests de positionnement de langues¹⁶. La majorité est disponible à tous sur Internet, hormis celui de l'Alliance Française de Grenoble (désormais, AFGrenoble) et celui du CUEF de Grenoble (désormais, CUEF). Pour ces derniers, j'ai pris rendez-vous avec un représentant de l'institution.

En plus de vouloir découvrir les caractéristiques de ces tests, je cherchais également à savoir si leurs particularités variaient en fonction du type d'institution et, de fait, nous constatons que :

- Les test GEMfrançais et GEM pour le programme d'anglais (désormais, GEManglais) sont évidemment utilisés en sein d'une école de management.
- Les trois autres tests (les tests CUEF, AFGrenoble et AFBruxelles [désormais, AFBruxelles]) sont destinés à des écoles de langue française.
- ALTISSIA et *Tell me more* sont faits pour des plateformes d'apprentissage à distance (« e-learning »).
- Enfin *Dialangue* et ELOA sont utilisés par diverses institutions qui achètent les droits de s'en servir.

¹⁵ Pour rappel, ce besoin est la réduction du coût de la passation et de la correction du test pour GEM.

¹⁶ Mis à part le test de placement pour le programme d'anglais à GEM, il s'agit de tests pour le français.

En analysant ces tests à travers le tableau synoptique ci-dessous, j'ai remarqué que leurs caractéristiques variaient considérablement. Tout d'abord, ils n'évaluent pas tous les mêmes compétences. Ainsi, certains ne s'intéressent qu'à la grammaire et au lexique (AFBruxelles, ALTISSIA et GEM pour le français¹⁷ [désormais, GEMfrançais]), tandis que d'autres, en plus de ces compétences, abordent la compréhension orale et écrite (CUEF, AFGrenoble, *Tell me more*, *Dialangue*). Ou encore, certains évaluent le vocabulaire passif et actif (ELAO), alors que d'autres n'évaluent que le lexique passif (AFBruxelles). Enfin, ceux qui se font sur place, c'est-à-dire à l'école (AFGrenoble, CUEF, GEMfrançais), comprennent un entretien oral avec un enseignant, hormis celui de GEManglais.

Ensuite, excepté les tests de AFGrenoble et de GEManglais, il s'agit uniquement de tests dont la passation s'effectue en ligne et à distance, témoignant que cette modalité de passation est de plus en plus répandue. Par là-même, ce sont majoritairement des tests autocorrectifs. Certes, certains tests sont destinés à un public qui souhaite avoir des cours à distance, et ils ne peuvent donc se faire qu'en ligne (ALTISSIA et *Tell me more*). Cependant, d'autres, comme le CUEF et GEManglais, ont choisi de mettre en place un test en ligne à faire passer à distance bien qu'il s'agisse de cours en présentiel.

Nous remarquons en outre que la moyenne du nombre d'items s'élève à 88 par test et, mis à part le test de GEMfrançais et de AFGrenoble, les questions sont majoritairement de type QCM. Par ailleurs, les tests de ELOA, du CUEF, de GEManglais et *Tell me more* imposent une limite de temps qui s'élève à 56 minutes en moyenne.

¹⁷ Il s'agit du test pour le français existant.

		Compétences évaluées	Entreti en oral	Niveaux couverts	Modalité	Limite de temps	Type(s) de questions	Nombre de questions	Modalité de jugement	Modalité de notation	Pour quel contexte	Auto-évaluation
1	<i>GEM français (Test existant)</i>	Grammaire / syntaxe, EE (B1-B2 seulement)	Oui	A1-B2	Sur papier, sur place	Non	Questions ouvertes à réponses courtes + longues	A1-A2 : 102 B1-B2 62 + 3 expressions écrites	Moyenne calculée à partir du test écrit et de l'entretien oral	Correction par enseignant	École de Management	Non
2	<i>GEM anglais</i>	CO, structures syntaxiques / EE, CE	Non	A1-C1+	Sur ordinateur, à distance	Oui (1h)	QCM	71	Note sur 71 (pourcentage)	Autocorrectif	École de Management	Non
3	<i>Alliance Française de Bruxelles</i>	Grammaire, lexique passif	Non	A1-C2	Sur ordinateur, à distance	Non	QCM, 1 production écrite	63 (43 grammaire, 20 lexique)	Niveau donné automatiquement	Autocorrectif et correction par enseignant (PE)	École de langue	Non
4	<i>CUEF</i>	CO, CE, compétences grammaticale et syntaxique	Oui	A1-B2	Sur ordinateur, à distance	Oui (45-60minutes)	QCM (4 choix)	120 (40 par compétence)	Moyenne des trois épreuves	Autocorrectif	École de langue	Non
5	<i>Alliance Française de Grenoble</i>	CO, EO, EC (entretien oral et production écrite)	Oui	A1-B2	Sur papier, sur place	Non	Questions ouvertes	2 EE, 11 questions à l'oral	Estimation subjective	Correction par enseignant	École de langue	Non
6	<i>ALTISSIA</i>	Grammaire, lexique	Non	A1-C2	Sur ordinateur, à distance	Non	QCM, questions ouvertes	35	Moyenne calculée des deux épreuves	Autocorrectif	Plateforme e-learning	Oui
7	<i>Tell me more (Rosetta stone)</i>	Grammaire, lexique, CO, CE	Non	A1-C1	Sur ordinateur, à distance	Oui (30 secondes par question)	QCM	64	Note sur 10	Autocorrectif	Plateforme e-learning	Oui

)							
8	<i>Dialangue</i>	CO, CE, EE, structures grammaticales, lexique actif et passif	Non	A1-C2	Sur ordinateur, à distance / sur place (en fonction de l'institution)	Non	QCM (3-5 choix), questions ouvertes	120 (30 par compétence)	Moyenne calculée à partir du niveau en chaque compétence	Autocorrectif	Divers	Oui
9	<i>ELAO</i>	Structures grammaticales, lexique passif, vocabulaire actif, CO	Non	A1-C2	Sur ordinateur, à distance	Oui (1 minute par question)	QCM, questions courtes ouvertes	76	Note sur 76 (pourcentage)	Autocorrectif	Divers	Non

Tableau 1 : Tableau synoptique de l'analyse de neuf tests de positionnement

En prenant en compte les informations du tableau synoptique, il est possible de tirer quelques conclusions sur la nature générale des tests de positionnement. Premièrement, nous pouvons noter que les spécificités des tests varient de manière plutôt importante. Autrement dit, nous ne trouvons aucun test identique à un autre en ce qui concerne leurs caractéristiques. De surcroît, les tests utilisés dans les mêmes types d'institutions ne partagent pas tous les mêmes spécificités. En comparant les tests des Alliances Françaises et du CUEF, nous observons un écart important concernant les compétences évaluées, le nombre d'items, la modalité de passation et la longueur. Il en est de même pour les tests destinés à une plateforme e-learning, ainsi que pour ceux conçus pour des situations d'apprentissage diverses (*Dialangue* et ELOA).

Ce constat m'a permis de conclure qu'une formule unique pour un test de positionnement n'existe pas. Le fait que certains tests varient n'est pas surprenant étant donné que la plupart sont utilisés par des institutions de langues différentes, et qui ont des besoins variés. En revanche, il était surprenant de constater des différences entre les tests de AFGrenoble, de AFBruelles et du CUEF. Bien que ces structures soient toutes dédiées uniquement à l'enseignement du français, le fait que leurs tests diffèrent montre que les caractéristiques d'un test de placement peuvent varier même au sein d'institutions similaires. En d'autres termes, deux antennes de la même institution, comme Les Alliances Françaises, n'ont pas forcément les mêmes besoins ni le même public. Aussi est-il indispensable de mener une étude sur l'institution, le public, les enseignants et les cours avant de lancer la conception d'un test de placement afin que celui-ci soit adapté aux particularités et aux missions de l'établissement.

1.2. Questionnaire auprès des étudiants sur le test existant

Toujours concernant l'analyse du test existant, j'ai voulu recueillir les opinions des étudiants qui l'ont passé à la rentrée de janvier 2015. La prise en compte de leurs opinions était importante pour l'élaboration du nouveau test car ils représentent le public auquel il se destine. De plus, comme je cherchais à démontrer que le profil des étudiants influence la nature du dispositif, il était important de prendre en compte ces informations dans le travail d'élaboration du nouveau test.

J'ai donc réalisé un questionnaire auprès des étudiants, et ai obtenu 17 réponses (*cf.* Annexe 2 [p. 9]). Bien que ce questionnaire soit court et ne contienne que quatre questions, les réponses se sont révélées suffisantes pour obtenir des informations utiles pour concevoir un nouveau test mieux adapté à GEM.

J'avais plusieurs objectifs lors de cette récolte de données. Je voulais premièrement savoir si les étudiants trouvaient le test cohérent avec leur niveau de langue (*cf.* Annexe 2 [p. 9], question 1 et 2), et deuxièmement, si les items étaient valides (*cf.* Annexe 2 [p. 10]. questions 3 et 4). Les deux premières questions montrent que les étudiants n'ont pas toujours trouvé que les tests débutant et

avancé étaient adaptés à leur niveau. Dans le processus de positionnement existant, les étudiants choisissaient le test débutant ou avancé en fonction de leur ressenti par rapport au français, ainsi que de leur apprentissage ou non de cette langue. Par exemple les questions « Êtes-vous débutant ou avancé ? », et « Avez-vous déjà étudié le français ? » leur ont été posées. Les résultats des deux premières questions du sondage montrent qu'ils n'avaient pas forcément reçu le bon test, ce qui a provoqué de mauvais placements et, de ce fait, des changements de groupes parfois compliqués. J'ai effectivement observé ce problème lors de la rentrée de janvier 2015, avec un changement qui a dû être effectué car un étudiant avait eu le test avancé au lieu de débutant. Compte tenu de ce problème, j'ai conclu que la nouvelle procédure de positionnement devait éliminer la possibilité qu'un étudiant passe le test avancé au lieu du test pour débutant, et vice versa.

Ce questionnaire m'a ensuite permis de savoir si les questions étaient claires pour les étudiants, c'est-à-dire s'ils comprenaient chaque question (cf. Annexe 2 [p. 10], questions 3 et 4). Les résultats de la 3^{ème} question montrent qu'ils avaient trouvé les items plutôt clairs. Cependant, les réponses à la 4^{ème} question montrent que, d'après les étudiants, plusieurs réponses étaient possibles pour certains items. Cela est problématique puisque ces derniers ne savaient pas quelle réponse choisir. Nous pouvons ainsi conclure que le test n'était pas complètement valide puisqu'il n'évaluait pas toujours ce qu'il était censé évaluer.

Pour terminer, j'ai récolté deux commentaires sur le test grâce à la question 5 qui ont permis d'adapter le test aux étudiants de GEM. La première remarque prouve la nécessité de mesurer plusieurs compétences en évaluant un niveau de langue au lieu de se concentrer uniquement sur la grammaire.

« It is difficult because language test doesn't need to include that much grammar questions. There were tons of grammar questions but a little vocabulary question. A language test must include vocab, grammar and daily life questions too¹⁸. » (cf. Annexe 2 [p. 10], question 5, réponse 2).

Ce commentaire m'a poussé à vouloir diversifier le nouveau test en intégrant différents items et en évaluant plus de compétences afin de mieux cerner le vrai niveau des étudiants. De plus, je souhaitais prendre en compte les opinions et les besoins de ceux qui passent le test pour qu'il soit mieux adapté à leur profil.

Quant à la deuxième remarque, elle est particulièrement importante car elle met en évidence une caractéristique importante du public de GEM. En effet, elle montre que leurs études de français

¹⁸ « C'est difficile car (un) test de langue ne doit pas contenir autant de questions d'ordre grammatical. Il y avait beaucoup de questions sur la grammaire et très peu sur le vocabulaire. Un test de langue doit comprendre le vocabulaire, la grammaire et des questions sur la vie quotidienne aussi. » (cf. Annexe 2 [p. 10], question 5, réponse 2, traduction libre)

ont parfois été interrompues, ce qui risque de provoquer des oublis qui peuvent être rapidement comblés par la reprise des études de français.

« I knew what the questions were asking, but I couldn't remember a bunch of the tenses. I took a lot of French in high school and then forgot it all, so if I could take a refresher course I would pick it all up again really quickly!¹⁹ » (cf. Annexe 2 [p. 10], question 5, réponse 2)

Cette remarque a donc suscité la volonté de créer un test adapté à un public qui a éventuellement oublié ses connaissances en français, mais qui s'en souviendra rapidement grâce à quelques cours. Ce phénomène peut être décrit comme la réactivation des connaissances antérieurement acquises. Il était donc question de mettre en place des items appropriés en vue de cette particularité du public.

1.3. Questionnaire de satisfaction auprès des enseignants sur le test existant

Comme l'interrogation des étudiants, le questionnaire auprès de dix enseignants de GEM visait à récolter leurs opinions sur le test existant (cf. Annexe 3 [p. 11]). L'objectif était de savoir si ces professeurs étaient généralement satisfaits de la forme (la longueur, les consignes et l'organisation), des questions (le type et la difficulté des items, et l'adaptation du contenu au niveau donné) et enfin de sa capacité à cerner le niveau réel de l'étudiant. Ce sont les enseignants qui connaissent bien le test et, qui ont donc pu apporter des commentaires utiles.

Les résultats des deux premières questions montrent que seulement 12% des enseignants étaient peu satisfaits du test débutant, tandis que 25% l'étaient avec le test avancé (cf. Annexe 3 [p. 11], question 1). Cela montre que même s'ils étaient globalement satisfaits des deux tests, celui des niveaux B1 et B2 nécessitait sans doute plus d'attention et plus d'améliorations que celui des niveaux A1 et A2. Ensuite, la question 5 souhaitait mettre en évidence l'efficacité avec laquelle les tests déterminaient le niveau des étudiants. Sur une échelle de 1 à 5 (1 étant « très mal positionné » et 5 étant « très bien placé ») les réponses se situaient majoritairement à 3, montrant que les enseignants jugeaient le test moyennement efficace pour cerner le vrai niveau de langue des étudiants. En outre, j'ai appris que les enseignants trouvaient que les tests étaient de la bonne longueur, mis à part 20% à 30% qui ont répondu que les tests étaient soit trop longs soit trop courts (cf. Annexe 3 [p. 12], questions 3 et 4). Ces dernières réponses m'ont incité à examiner le nombre d'items pour savoir s'il fallait le réduire ou l'augmenter.

¹⁹ « Je savais ce que demandaient les questions, mais je ne me souvenais pas de tous les temps verbaux. J'ai suivi beaucoup de cours de français au lycée et puis j'ai tout oublié, donc si je pouvais juste suivre un cours pour me rappeler ces choses, je m'y remettrais rapidement ! » (cf. Annexe 2 [p. 10], question 5, réponse 2, traduction libre)

Je souhaite par ailleurs attirer l'attention sur les résultats des questions 6 et 7 puisqu'ils permettent de savoir si le test existant était valide d'après les enseignants (« Trouvez-vous que le contenu linguistique des tests est adapté aux niveaux ? »). En effet, si le contenu linguistique d'un test ne correspond pas au niveau, il n'évalue pas ce qu'il est censé évaluer. Les résultats montrent que d'après la moitié des enseignants (50%-56%), le contenu des tests était seulement « assez adapté au niveau ». Ainsi ai-je su qu'il fallait que je mène une analyse minutieuse sur la cohérence entre le contenu linguistique et le niveau auquel il était associé dans le test.

Quant à la question 8, les réponses étaient utiles car je voulais m'interroger sur la qualité de la notation des enseignants. L'enjeu de la correction m'a été premièrement dévoilé lors d'un entretien avec ma tutrice de stage. Durant cette conversation, elle m'a en effet dit qu'elle s'était aperçue d'un problème au niveau de la précision avec laquelle les enseignants corrigeaient le test écrit. «...tout dépend aussi de la correction, la correction est hyper importante. » (cf. Annexe 16 [p. 68], entretien 1, tour de parole 89). J'avais moi-même décelé ce problème en analysant le test existant²⁰. Ainsi, les réponses à la question 8 mettent en évidence qu'uniquement 30% des enseignants se référaient souvent au corrigé, alors que 70% s'appuyaient sur ce dernier « parfois » ou « jamais ». Cela a sans doute provoqué des placements inappropriés ou bien des positionnements peu cohérents entre les différents enseignants. Les données issues de cette question montrent que la correction n'était pas assez guidée par manque de critères, et que le jugement était donc trop subjectif. Au vu de ce constat, et de la conception du nouveau test, il était nécessaire que je m'interroge sur comment créer un test de positionnement dont la correction ne soit pas trop subjective.

Enfin, les questions ouvertes à la fin du questionnaire ont dévoilé les changements que les enseignants souhaitaient pour le nouveau test. Ces informations ont aidé dans le travail de conception qui prend en compte leurs attentes. Les enseignants ont évoqué l'usage trop fréquent de métalangage dans le test existant, ce qui risque de provoquer des incompréhensions au niveau des consignes car ce sont des termes linguistique techniques (cf. Annexe 3 [p. 14], question 9, réponses 3, 5 et 6 ; et question 10 [p. 15], réponse 1). Rappelons que j'ai eu la même remarque d'un étudiant dans le questionnaire sur le test existant²¹. De plus, ils ont critiqué la centration du test sur la grammaire, un reproche qui est également ressorti dans le questionnaire auprès des étudiants. En outre, certaines réponses ont révélé la nécessité de compléter le test B1/B2 (cf. Annexe 3 [p. 15], question 12, réponses 6²² et 7²³ ; question 13 [p. 16], réponse 6). Enfin, la volonté d'intégrer des productions écrites dans le

²⁰ J'évoquerai le problème de la notation du test existant dans la partie 3, 1.4.3.

²¹ cf. partie 3, 1.2., 1.3.

²² « Le questionnaire (*les items*) pour le B1.1 me semble un peu haut. Je pensais que c'était au niveau B1.2 que les étudiants apprennent/découvraient le discours indirect par exemple. »

²³ « Le test Bi(I) -B2 n'est pas tout à fait assez complet, on a quelques difficultés à se rendre compte du niveau réel des étudiants. Il faudrait le compléter. »

nouveau test a été évoquée (cf. Annexe 3 [p. 15], question 12, réponse 8 ; et question 13 [p. 16], réponses 2, 3, 4, et 7).

J'ai tenu à prendre en compte tous les commentaires des enseignants lors de l'élaboration du nouveau test de placement, tout comme les remarques issues du questionnaire auprès des étudiants. Cela était important si je voulais adapter avec succès le test aux besoins de ceux qui sont impliqués dans le processus de positionnement des étudiants.

1.4. L'analyse du test existant

Étant donné que ma mission de stage était de créer un nouveau test de positionnement, il était important d'analyser en détails l'existant afin de connaître ses points faibles et forts. Grâce à cette étude, j'ai su par la suite comment l'améliorer sur le plan linguistique, ainsi qu'au niveau de la forme (organisation, progression, consignes et types d'items).

Ainsi, je parlerai premièrement de l'analyse de la forme du test. Ensuite, j'évoquerai les résultats de l'analyse linguistique. Pour terminer, je présenterai l'étude des tests des étudiants qui ont changé de groupe lors des premières semaines de la rentrée de janvier 2015 à cause d'un mauvais placement dans le groupe de niveau.

1.4.1. L'analyse de la forme

Ce test est divisé en deux parties, à savoir le test débutant (A1 / A2), et le test avancé (B1 / B2) (cf. Annexes 4 [p. 17], et 5 [p. 21]). Le premier est ensuite divisé en quatre sections : A1.1, A1.2, A2.1 et A2.2, le deuxième en B1.1, B1.2, B1.3 et B2. Cela correspond aux sous-niveaux qu'utilise le programme de FLE à GEM. Dans le cadre de la procédure en place auparavant, les étudiants passaient le test débutant ou avancé en fonction des questions que l'on posait à l'étudiant sur son apprentissage du français («Êtes-vous plutôt débutant ou avancé ? Avez-vous déjà étudié le français ? Pendant combien de temps ? Où ? Il y a combien de temps ? »). Cette manière de décider quel test donner à l'étudiant ne fonctionnait pas toujours bien. Certains étudiants recevaient ainsi le test avancé au lieu du test pour débutant, ou vice versa, ce qui provoquait parfois des changements de groupe. Le nouveau test devait donc éliminer la possibilité de passer un test non adapté à son niveau.

Quant aux types d'items, il s'agit principalement de questions à réponses fermées étant donné qu'il n'existe qu'une seule bonne réponse (cf. Annexe 4 [p. 17], niveau A1.1, exercices 2 et 3 ; et niveau A1.2 [p. 18], ex. 4 ; niveau A2.1 [p. 18], exercices 1 et 2 ; Annexe 5 [p. 21], niveau B1.1, exercice 1). Cependant, le test contient également des questions ouvertes, c'est-à-dire que l'étudiant peut répondre correctement avec plusieurs réponses. Par exemple, « Aimez-vous les légumes ? » (cf. Annexe 4 [p. 17], niveau A1.2, exercice 2. Possibilité de répondre avec le pronom « je » ou « nous »),

ou encore l'exercice 1 du niveau B1.2 qui demande que l'étudiant écrive un petit texte en utilisant le conditionnel (*cf.* Annexe 5 [p. 22], niveau B1.2, exercice 1).²⁴

L'analyse des types d'items révèle deux conclusions importantes. Premièrement, l'utilisation des questions auxquelles un étudiant peut apporter plusieurs réponses rend la correction fastidieuse car l'évaluateur ne sait pas forcément s'il doit marquer la réponse comme correcte ou pas. Cela peut engendrer des différences de notation, et donc des positionnements incohérents entre les différents enseignants. En d'autres termes, la notation est trop subjective. Deuxièmement, puisque les questions sont principalement à réponses fermées, j'ai conclu que les items étaient transformables en QCM et que le test était donc adaptable à une passation en ligne. Étant donné que la mission était de mettre en place un test autocorrectif, qui est forcément constitué de questions type QCM, j'ai voulu confirmer, à travers cette analyse, que c'était effectivement possible de réaliser un tel examen.

En ce qui concerne les compétences que le test existant évalue, il s'agit principalement de la grammaire, hormis certains exercices dans le test avancé qui demandent d'utiliser correctement des connecteurs logiques ou des expressions de la cause (*cf.* Annexe 5 [p. 24], niveau B1.3 ; exercices 4 et 5), ou bien des productions écrites (Annexe 5 [p. 22], niveau B1.2, exercice 1 ; niveau B2, exercices 1 et 2). À propos de ces dernières, nous remarquons qu'il n'existe pas de grille de notation ni de critères pour guider la mesure de la qualité de ces productions, ce qui fait que ces productions écrites ne donnent pas de résultats fiables pour déterminer le niveau de l'étudiant à cause d'une trop grande subjectivité.

Enfin, l'analyse de la forme du test s'intéresse également aux consignes. À ce propos, j'ai remarqué deux choses importantes. Premièrement, les consignes contiennent souvent du métalangage. Ces termes risquent d'entraver la compréhension, et d'induire ainsi des réponses fausses. Par exemple, l'appellation des temps verbaux (*cf.* Annexe 4 [p. 18], niveau A1.2, exercice 4, « le futur proche » ; Annexe 5, niveau B1.1, exercice 1 « le conditionnel ») n'est pas toujours connue par les étudiants bien qu'ils sachent peut-être l'utiliser. D'ailleurs, lors de la passation du test à la rentrée de janvier 2015, j'ai noté que l'exercice du niveau B1.1 sur le conditionnel a été rarement complété, même par les étudiants ayant un niveau au-dessus de B1.1. Cela m'a montré que soit cet exercice n'était pas adapté au niveau du test, soit l'utilisation du métalangage provoquait une confusion qui empêchait la réalisation de cet exercice. Par ailleurs, je tiens à rappeler que le problème du métalangage avait été souligné par les étudiants et les enseignants à travers les questionnaires évoqués dans la partie 3 (1.2) et (1.2.) Il était donc nécessaire de trouver une solution à ce problème dans le nouveau test.

²⁴ La multiplicité des réponses a également été confirmée à travers le questionnaire fait auprès des étudiants sur le test existant (*cf.* partie 3, 1.2.)

Par ailleurs, sans logique apparente, certains exercices contiennent des exemples et d'autres non. En effet, si l'exemple ne facilite pas trop la tâche, il est préférable d'en fournir un afin de rendre l'item le plus clair possible pour les étudiants. Cette volonté a été d'ailleurs donc exprimée par les enseignants²⁵. De plus, j'ai appris, par l'intermédiaire d'un entretien avec ma tutrice de stage, que les enseignants constataient parfois que les étudiants ne comprenaient pas toujours bien toutes les questions du test. « Il y a certains enseignants qui disent 'ouais tu vois ils [*les étudiants*] comprennent pas la consigne' » (cf. Annexe 16 [p. 67], entretien 1, tour de parole 77).

Ainsi, en constatant le manque d'exemples, je me suis aperçu que d'en inclure dans le nouveau test remédierait aux deux problèmes des consignes évoqués ci-dessous. Les questions seraient plus claires, c'est-à-dire que les étudiants sauraient ce qu'il faut faire. De plus, l'usage du métalangage n'entraverait pas la compréhension de la consigne car les étudiants verraient l'exemple et comprendraient mieux l'exercice.

1.4.2. L'analyse du contenu linguistique

Dans un deuxième temps, j'ai voulu examiner le contenu linguistique du test afin de savoir si ce dernier était cohérent avec le niveau associé. Cette étude a permis de déterminer la validité du test car s'il demandait des connaissances à un niveau donné qui sont normalement abordées dans un niveau supérieur ou inférieur, alors sa validité était réduite. De plus, j'ai souhaité vérifier la progression de la difficulté, puisque cela est une caractéristique importante d'un test de positionnement²⁶.

Tout d'abord, nous remarquons que le test original évalue principalement les connaissances grammaticales. De fait, il demande que les étudiants fassent preuve de savoirs sur des éléments précis et isolés de la langue. Mises à part les productions écrites du test avancé qui se rapproche davantage d'une évaluation de la performance, il s'agit principalement d'évaluer des connaissances grammaticales telles que les conjugaisons, l'utilisation des compléments et des articles définis et indéfinis, ou la maîtrise de la phrase interrogative. Ainsi, ce test représente plutôt une approche mécaniste similaire à une évaluation de la période psychométrique structuraliste.

Ensuite, dans le but de vérifier la cohérence entre les contenus linguistiques et les niveaux du test, il a été indispensable que je m'appuie sur des outils de référence de langue. Ainsi, j'ai choisi le *Référentiel de programmes pour l'Alliance Française*, les référentiels des niveaux A1, A2 et B1 de Beacco (ouvrage publié par le Conseil de l'Europe), la méthode de français *À Propos A1* et, évidemment, le référentiel du programme de français à GEM (cf. Annexe 18 [p.75]). J'ai choisi ces

²⁵ Cf. partie 3, 1.3.

²⁶ Cf. partie 2, 2.3.

deux premiers car ils sont basés sur le CECRL et détaillent ce que l'on doit aborder sur le plan linguistique (grammaire, lexicale, actes de parole) aux niveaux A1 à C1. Ensuite, j'ai voulu examiner la méthode *À propos A1*, d'une part car elle est basée sur le CECRL et, d'autre part, pour avoir une idée de ce qui est enseigné au niveau A1 du point de vue d'une méthode. Enfin, l'étude du référentiel du programme de français à GEM était nécessaire pour m'assurer de la cohérence entre le test de positionnement et ce qui est enseigné dans les cours de cette institution.

En examinant ces outils, je me suis d'abord aperçu que dans la méthode *À propos A1*, le contenu linguistique était très avancé. En d'autres termes, nous trouvons à la fin de cette méthode des contenus qui figurent au niveau A2 selon les autres référentiels. Par exemple, elle aborde la distinction entre le passé composé et l'imparfait en fin A1, tandis que, d'après les trois autres outils, ce sont des temps verbaux à enseigner en A2. De même, à GEM la différence entre ces temps verbaux n'est pas encore abordée au niveau A1. Aussi ai-je pu conclure que les méthodes de français incluent parfois plus de contenus que ceux qui peuvent être abordés lors d'un semestre de cours à GEM. Les méthodes ne représentaient donc pas une ressource assez fiable pour déterminer ce que le nouveau test de positionnement devait évaluer, en termes de grammaire et de syntaxe, par niveau. Face à ce constat, j'ai décidé de me baser uniquement sur les trois référentiels pour analyser le contenu linguistique du test existant.

Avec les référentiels de l'Alliance Française, de GEM et de Beacco, j'ai pu valider la pertinence du contenu par rapport aux quatre niveaux de langues (A1-B2). De fait, si l'élément linguistique figurait au même niveau dans le test que dans les référentiels, sa place dans le test était validée. Au contraire, lorsque j'ai noté une ou des discordance(s) entre le niveau de l'élément dans le test et son niveau d'après un ou plusieurs référentiels, j'ai estimé qu'il ne figurait pas dans le bon niveau du test.

Ainsi, tous les éléments linguistiques qui n'étaient pas cohérents avec le niveau du test figurent dans l'un des quatre tableaux ci-après (cf. Tableaux 2 à 5). Le niveau suivant les deux points dans chaque case est celui auquel l'élément linguistique en question doit être abordé selon le référentiel. Les tableaux mettent ainsi en évidence les écarts entre le niveau des contenus du test existant et le niveau auquel ceux-ci devraient appartenir selon les trois référentiels. En outre, les contenus en rouge sont ceux dont le niveau auquel ils doivent être enseignés diffère selon un ou plusieurs des référentiels.

<i>Selon le référentiel des niveaux de l'Alliance Française</i>	<i>Selon « Niveau A1 pour le français » (J.C. Beacco, 2007)</i>	<i>Selon le référentiel de GEM</i>
Les verbes pronominaux : A2	Les verbes pronominaux : ? ²⁷	Les verbes pronominaux : ?
La négation « ne...personne » : A2	La négation « ne...personne » : A2	La négation « ne...personne » : A2
Le futur proche : A2	Le futur proche : A2	Le futur proche : A1
La phrase interrogative inversée : A1	La phrase interrogative inversée : B1	La phrase interrogative inversée : ?
Les verbes à deux et trois bases (sauf « comprendre ») : A2	Les verbes à deux et trois bases (sauf « comprendre ») : A2	Les verbes à deux et trois bases (sauf « comprendre ») : A2

Tableau 2 : Test existant niveau A1

<i>Selon le référentiel des niveaux de l'Alliance Française</i>	<i>Selon « Niveau A2 pour le français » (J.C. Beacco, 2008)</i>	<i>Selon le référentiel de GEM</i>
Les articles partitifs / contractés : A1	Les articles partitifs / contractés : A1	Les articles partitifs / contractés : A1
Conditionnel présent (sauf quelques verbes usuels tels que « vouloir, pouvoir, avoir, être et des formes de politesse [je voudrais, vous pourriez, tu aurais]) : à partir de A1, à B1	Conditionnel présent (sauf quelques verbes usuels tels que « vouloir, pouvoir, avoir, être et des formes de politesse [je voudrais, vous pourriez, tu aurais]) : B1	Conditionnel présent (sauf quelques verbes usuels tels que « vouloir, pouvoir, avoir, être et des formes de politesse [je voudrais, vous pourriez, tu aurais]) : B1
La distinction passé composé / imparfait : B1	La distinction passé composé / imparfait : B1	La distinction passé composé / imparfait : A2

Tableau 3 : Test existant niveau A2

²⁷ Un point d'interrogation apparaît lorsque que le niveau de l'élément en question n'est pas précisé dans le référentiel.

<i>Selon le référentiel des niveaux de l'Alliance Française</i>	<i>Selon « Niveau B1 pour le français » (J.C. Beacco, 2011)</i>	<i>Selon le référentiel de GEM</i>
Les pronoms compléments (le, la, les, lui, leur, y, en) : A2	Les pronoms compléments (le, la, les, lui, leur, y, en) : A2	Les pronoms compléments (le, la, les, lui, leur, y, en) : A2
Les pronoms relatifs composés (préposition + qui, lequel, laquelle, etc.) : B2	Les pronoms relatifs composés (préposition + qui, lequel, laquelle, etc.) : B2	Les pronoms relatifs composés (préposition + qui, lequel, laquelle, etc.) : B2
La distinction entre le subjonctif et l'indicatif : B2	La distinction entre le subjonctif et l'indicatif : B2	La distinction entre le subjonctif et l'indicatif : B2
La cause et la conséquence (malgré, avoir beau, à force de, de peur de) : B1-C1	La cause et la conséquence / la cause (avoir beau, à force de, de peur de) : B1-B2	La cause et la conséquence (avoir beau, à force de, de peur de) : B2
Les pronoms relatifs simples (qui, que) : A2	Les pronoms relatifs simples (qui, que) : A2	Les pronoms relatifs simples (qui, que) : A2

Tableau 4 : Test existant niveau B1

Le contenu du test B2 existant est adapté au niveau selon les trois référentiels.

Tableau 5 : Test existant niveau B2

Grâce à l'analyse du contenu linguistique du test en fonction des niveaux, et au tableau synoptique, j'ai pu tirer deux conclusions importantes qui ont aidé considérablement la conception du nouveau test. Premièrement, cette étape a montré que la difficulté du test existant n'était pas toujours adaptée aux niveaux. Par exemple, nous pouvons voir dans les tableaux des tests A1 et A2 que les contenus sont trop difficiles pour ces niveaux, ce qui est également le cas pour le niveau B1. Par là-même, la progression de difficulté n'est pas toujours correcte étant donné que nous trouvons dans le niveau B1 du test des contenus du niveau A2 (i.e. les pronoms relatifs « qui » et « que »). Quant au niveau B2, ce test ne contient que deux exercices qui sont effectivement cohérents avec le niveau B2 selon les référentiels. Aussi ce tableau est-il vide.

Deuxièmement, j'ai remarqué des différences entre les trois référentiels concernant le niveau auquel un élément grammatical donné doit être enseigné (pour rappel, ces discordances sont marquées en rouge). Bien qu'elles ne soient pas nombreuses, il s'agit d'un constat important à prendre en compte car cela montre que les normes du CECRL ne sont pas toujours respectées dans les référentiels. Or, lorsque nous élaborons un test de placement, il faut choisir une ressource sur laquelle s'appuyer pour que le niveau fourni par le test soit non seulement le niveau réel de l'étudiant, mais également le même selon le CECRL. Il est en effet question de s'assurer de la validité des items en s'efforçant de concevoir un test de positionnement qui évalue ce qu'il est censé évaluer selon les niveaux du CECRL.

Au vu des écarts montrés dans les tableaux, je peux conclure que le test existant n'est pas parfaitement en accord avec les niveaux du CECRL, une hypothèse d'ailleurs émise par ma tutrice de stage lors de l'entretien.

« ouais ouais ouais c'est (*le fond du test*) un des enjeux, c'est un des enjeux, après j'ai aussi noté que le test n'est pas toujours représentatif du niveau de l'étudiant, et que, à mon avis, peut-être qu'on peut (*l'*)affiner encore en fonction du niveau (...) en proposant des exercices, ou plus élaborés, ou plus simples en fonction du niveau, au niveau de la consigne. Aussi souvent ils (*les étudiants*) comprennent mal la consigne et ils font mal l'exercice » (*cf. Annexe 16 [p. 66], entretien 1, tour de parole 36*).

Un des objectifs de la mission était donc de créer un test de placement dont le contenu serait cohérent avec les niveaux du test, en m'appuyant sur des ressources basées sur le CECRL. De même, les discordances ponctuelles entre les référentiels que j'ai pu relever lors de cette analyse m'ont poussé à me demander comment choisir le contenu linguistique pour un test de positionnement en fonction des niveaux. Ce sont effectivement ces conclusions qui ont dirigé la création du nouveau test²⁸.

1.4.3. L'analyse de quatre tests faits lors de la rentrée de janvier 2015

²⁸ J'évoquerai de quelle manière elles ont influencé la conception du nouveau test dans la partie 3, 1.5.

Étant donné qu'une des motivations pour refaire le test de positionnement était de réduire les changements de groupe, il a été nécessaire d'étudier les tests faits lors de la rentrée de janvier 2015 par les étudiants qui avaient changé de groupe au cours des premières semaines de cours. Cette étude a permis en effet de savoir pourquoi le niveau fourni par le test n'était pas toujours le vrai niveau de l'étudiant, et donc comment éviter ce problème dans le cadre du nouveau test.

Lors des premières semaines du deuxième semestre de l'année 2014/2015, le programme de FLE a effectué douze changements de groupes dont cinq pour des étudiants qui avaient passé le test en janvier 2015. Quant aux sept autres, il s'agit d'étudiants qui étaient déjà à GEM avant cette rentrée et qui avaient donc passé le test auparavant. Les cinq changements effectués en janvier sont ainsi ceux que j'ai analysés faute de pouvoir disposer des tests des sept autres.

Les informations relatives aux changements figurent dans le tableau 5 ci-dessous. Pour chacun des repositionnements, je me suis intéressé à l'importance du changement en terme de niveau (colonne 1), à sa cause (colonne 2), et à la source de la demande (colonne 3). J'ai procédé ainsi pour déterminer l'efficacité du test existant quant au le niveau d'un étudiant, pour savoir avec quel degré de précision et de quelle manière les enseignants notaient ces tests et, enfin, si c'était l'étudiant ou l'enseignant qui demandait le changement.

	<i>Colonne 1</i>	<i>Colonne 2</i>	<i>Colonne 3</i>
	Remonter ou descendre en niveau	Cause du changement (tirée de l'analyse du test)	Source de la demande (Étudiant ou professeur)
Changement 1	Descendre 1 sous-niveau (A2.2→A2.1)	L'étudiant a eu le test avancé au lieu de débutant.	L'étudiant
Changement 2	Descendre 1 sous-niveau (A2.2→A2.1)	La notation	L'étudiant
Changement 3	Monter 2 sous-niveaux (A1.2→A2.2)	La notation	L'étudiant
Changement 4	Monter 1 sous-niveau (A1.2→A2.1)	La notation	L'étudiant
Changement 5	Monter 1 sous-niveau (A2.1→A2.2+)	La notation	L'étudiant

Tableau 5 : Informations liées aux changements de groupe de niveau effectués suite à la rentrée de janvier 2015.

En examinant les données du tableau, nous remarquons qu'il s'agit de changements d'un seul sous-niveau, hormis le changement 3 qui fait l'objet d'un écart de deux sous-niveaux. Ainsi, le test existant ne fournissait jamais un niveau très éloigné de celui que l'étudiant possédait en réalité, ce qui montre qu'il était plus ou moins précis dans sa détermination du niveau. De même, cette colonne montre que deux étudiants sont descendus, et que les trois autres sont montés. Cela signifie que

lorsque le test n'était pas précis dans sa détermination du niveau, il s'agissait parfois d'une surévaluation, ou parfois d'une sous-évaluation. Nous voyons ensuite qu'il s'agissait à chaque fois d'une demande de changement de la part de l'étudiant, ce qui met en évidence la conscience des étudiants par rapport à la difficulté du cours en relation avec leur niveau de langue.

Un aspect du test original important à noter est que les questions n'étaient pas pondérées, c'est-à-dire qu'elles n'étaient pas accordées de points. Ainsi l'évaluateur ne pouvait-il pas calculer le nombre de points total gagnés par un étudiant pour ensuite lui attribuer un niveau. Ce fait montre à nouveau que les correcteurs du test existant ne pouvaient pas le corriger de manière exacte, et donc que la détermination du niveau de l'étudiant n'était pas assez guidée.

Enfin, cette analyse a dévoilé le point faible le plus important du test existant. De fait, mis à part le premier changement qui est dû au fait que l'étudiant ait reçu le mauvais test, il s'agit d'une notation inexacte. Par exemple, en examinant les tests des changements 2, 4 et 5 (*cf.* Annexe 6 [p. 25-36], tests 1, 2 et 3), nous remarquons l'imprécision et la subjectivité avec lesquelles ceux-ci ont été notés. Si l'étudiant répondait correctement aux questions jusqu'à tel ou tel sous-niveau (l'ancien test était divisé en sous-niveaux), l'évaluateur devait positionner cette personne à ce niveau. Par exemple, un étudiant qui avait pu répondre aux items du niveau A2.2 était placé dans ce niveau. Nous constatons que ceci est une manière subjective de prendre une décision, ce qui est mis en évidence par le test 1, et la raison pour laquelle ces étudiants ont dû changer de groupe. En examinant ce test, nous observons que l'étudiant a répondu correctement à la plupart des questions jusqu'au niveau A2.2. Or, le correcteur l'avait positionné au niveau A2.1. Les tests 2 et 3 mettent en évidence de la même manière la subjectivité de la notation du test original.

En somme, le test écrit existant ne contenait pas de barème de notation, ni de points totaux à compter afin de se prononcer avec davantage de précision sur le niveau de l'étudiant. Cette manière de noter l'ancien test était ainsi trop subjective et manquait d'outils pour guider la détermination du niveau.

1.5. La conception du nouveau test de positionnement

J'ai donc pu élaborer un nouveau test de placement en fonction des données récoltées grâce aux questionnaires et à l'analyse de test existant. D'une part, celui-ci résout les problèmes découverts par le biais de ma collecte de données (la subjectivité de la notation, les changements de groupes, les consignes et les compétences évaluées) et, d'autre part, répond au besoin de réduire le coût du test pour l'école.

Concernant la méthodologie utilisée pour réaliser ce travail de conception, je me suis servi du *Manuel pour l'élaboration et la passation de tests et d'examens de langue*, conçu par ALTE. Cette ressource explique en effet la procédure par laquelle il faut passer pour mettre en place une évaluation

de langue en lien avec le CECRL. Premièrement, j'ai sélectionné cet ouvrage étant donné qu'une partie de la commande de stage était de concevoir un test de positionnement basé sur le CECRL. Deuxièmement, il explicite non seulement les étapes qui constituent cette tâche, mais également les raisons pour lesquelles il est préférable de procéder ainsi.

1.5.1. Étape 1 : la décision de fournir un test ou un examen

La décision de concevoir un examen est la première étape de ce processus. Il s'agit d'examiner les raisons pour lesquelles l'institution, en l'occurrence GEM, souhaite disposer d'une nouvelle évaluation. Cette décision de la part de l'école nous fournit des informations qui, par la suite, influencent l'élaboration du test au niveau de sa forme, de son contenu, de sa passation et de sa correction (ALTE, 2011).

Dans le cas de mon stage, les informations liées à la décision sont issues de la commande (*cf.* partie 1, 5.). Pour rappel, le besoin primordial posé par la commande est la réduction du coût du test de positionnement pour l'école. C'est pour cette raison qu'il était pertinent de créer un test autocorrectif dont la passation s'effectue en ligne puisque celui-ci n'exige pas d'heures de travail, ne pour le faire passer, ni pour le noter. La décision a suscité également la volonté d'intégrer dans le test un contenu linguistique diversifié pour qu'il soit davantage valide. Quant à la plateforme pour mettre en place le test, j'ai choisi *Moodle* car c'est celle que GEM utilise. Qui plus est, elle permet de mettre en place des épreuves adaptables au contexte en mettant à disposition une gamme d'items divers et une notation automatique.

1.5.2. Étape 2 : la planification

Lors de cette étape, il est question de se rendre compte des caractéristiques des destinataires du test, de l'objectif de ce dernier, du contexte éducatif dans lequel s'insère la passation, de la norme sur laquelle sera fondée le test et, enfin, de comment les résultats seront utilisés (ALTE, 2011). Les données recueillies sont nécessaires pour que le test soit non seulement cohérent avec son objectif et adapté au contexte et aux caractéristiques des parties prenantes, mais aussi qu'il réponde aux besoins de ces dernières.

Les destinataires	Il s'agit d'étudiants d'origine diverse, âgés de 20 ans et plus qui réalisent des études supérieures de niveau licence, master et doctorat. Certains sont captifs, c'est-à-dire qu'ils sont obligés de suivre des cours de français, alors que d'autres ne le sont pas.
L'objectif du test	Déterminer le vrai niveau de langue des étudiants qui vont suivre des cours de français à GEM. Il s'agit de niveaux A1-B2
Le contexte éducatif	Une école de management comprenant un grand nombre de programmes.
La norme	Le CECRL et plus particulièrement les descripteurs de niveaux qu'il décrit.

L'usage des résultats	Positionner les étudiants dans les niveaux A1-B2 afin de créer des groupes de niveau homogène.
------------------------------	--

Tableau 6 : La planification du nouveau test de positionnement de GEM

Ce tableau permet de relever les principales informations qu'il faut prendre en compte avant de commencer la conception du nouveau test, à savoir le contexte, les caractéristiques du public, les niveaux que l'épreuve doit couvrir et la norme sur laquelle le test doit être basé. Il est vrai que ce ne sont pas les seules informations à considérer, bien entendu. Elles sont toutefois nécessaires pour élaborer le test.

1.5.3. Étape 3 : la conception du nouveau test

En prenant en compte les informations figurant dans le tableau 5, j'ai pu commencer la conception du nouveau test de positionnement. D'après le manuel de ALTE, il s'agit de prendre des décisions quant aux compétences à évaluer, les types et le nombre (total et par partie) d'items, l'organisation de l'épreuve (la durée et la structure), le nombre de points à donner à chaque question et, enfin, l'échelle de notation (ALTE, 2011).

Étant donné que le nouveau test (*cf.* Annexe 10 [p. 42]) évalue en fonction des niveaux A1-B2, il est composé de quatre parties, chacune correspondant à l'un de ces niveaux. Le nouveau test sera toutefois passé en entier par tous les étudiants afin d'éliminer la possibilité qu'un étudiant passe le test débutant au lieu du test avancé (ou vice versa), comme c'était le cas avec le test existant. Il est composé de 97 items au total (25 items pour les niveaux A1, A2 et B1, et 21 items pour le niveau B2). Je suis arrivé à ce nombre d'items suite à trois constats. Premièrement, l'étude des neuf tests de positionnement démontre un nombre moyen d'items de 88 (*cf.* partie 3, 1.1.). Deuxièmement, les recherches en évaluation nous rappellent que la multiplicité des questions améliore la validité du test car elle fournit une image plus étendue et plus exhaustive du niveau de l'apprenant (Mothe, 1975). Troisièmement, nous avons vu dans la partie 2, point 4.1.3. qu'à partir d'un certain nombre d'items et de distracteurs, le hasard et la divination ne représentent plus de risque important. Étant donné que le nouveau test contient 73 QCM avec 3 à 4 distracteurs, le risque que le score d'un étudiant soit influencé de manière importante par le hasard ou la divination est marginal. J'ai estimé ainsi que ce nombre d'items était approprié pour le nouveau test.

Précisément, il s'agit de 73 questions fermées sous forme de QCM et de textes lacunaires, ainsi que de 24 questions type QROC sous forme de conjugaison de verbe ou de morphologie syntaxique (la transformation d'une phrase au négatif, par exemple). Premièrement, étant donné que le test est autocorrectif, les QCM sont les plus pertinents car ils peuvent être corrigés automatiquement par l'ordinateur. Il s'est agi donc d'une obligation d'utiliser des questions fermées. C'était en effet la nature du test qui a déterminé les types d'items qui le constitue. Par là-même, le problème de la

subjectivité de la notation du test existant est résolu puisqu'il s'agit d'un test entièrement objectif. De plus, certains effets parasites de la fidélité, présents chez les correcteurs dans le processus de positionnement original, sont éliminés tels que l'effet de fatigue, de contraste, d'ordre et de contamination.

Néanmoins, utiliser des QCM n'est pas uniquement une question d'ordre pratique. Ce type de question est également bien adapté au public de GEM puisque, comme nous l'avons vu, ce dernier peut avoir oublié certains aspects du français à cause d'une interruption d'études. Ce sont des connaissances linguistiques qui peuvent revenir vite lors des premiers cours. Un commentaire issu du questionnaire sur le test existant et évoqué dans la partie 3, 1.2. démontre bien ce fait.

« I knew what the questions were asking, but I couldn't remember a bunch of the tenses. I took a lot of French in high school and then forgot it all, so if I could take a refresher course I would pick it all up again really quickly!²⁹ » (cf. Annexe 2 [p. 10], question 5, réponse 2)

Cette citation montre que les QCM sont pertinents et adaptés au public car ils indiquent si la personne est capable ou non de reconnaître la bonne forme. Il est en effet déplorable que, le jour de l'examen, un étudiant soit placé dans un niveau trop bas parce qu'il n'a pas pu conjuguer tel ou tel verbe correctement. Avec l'ancien test, cela était souvent le cas. Puis, après seulement quelques cours, un étudiant se rendait compte qu'il connaissait déjà, par exemple, le passé composé et l'imparfait. Il demandait ainsi un changement de niveau. Les QCM aident à éviter ce type de cas.

Cependant, il faut aussi vérifier que les étudiants sont capables de conjuguer eux-mêmes des verbes par le biais de questions ouvertes. C'est pour cela que le test contient également vingt-quatre questions à réponses courtes qui sollicitent principalement des conjugaisons de verbes, mais aussi des morphologies syntaxiques. Comme les QCM, ce type de question est adapté à un test autocorrectif car la plateforme *Moodle* permet de prédéterminer toutes les réponses possibles. J'ai également intégré dans ces items un outil permettant d'insérer des accents pour éliminer la possibilité que l'étudiant donne une réponse erronée pour des raisons non liées à ses compétences.

Par ailleurs, il s'agit de diversifier le test en intégrant différents types de questions. Quant à ces items qui demandent une conjugaison à tel ou tel temps verbal, le test fournit des exemples pour que le métalangage dans la consigne ne pose pas de problème de compréhension. Nous avons vu que les enseignants et les étudiants trouvaient l'utilisation des termes linguistiques problématique dans le test existant, d'où la nécessité de mettre des exemples dans le nouveau. En outre, pour que les exemples ne facilitent pas la tâche, j'ai fait en sorte que ceux-ci utilisent différents verbes et sujets. Par exemple :

²⁹ «Je savais ce que demandaient les questions, mais je ne me souvenais pas de tous les temps verbaux. J'ai suivi beaucoup de cours de français au lycée et puis j'ai tout oublié, donc si je pouvais juste suivre un cours pour me rappeler ces choses, je m'y remettrais rapidement !» (cf. Annexe 2 [p. 10], question 5, réponse 2, traduction libre)

19. *Conjuguez le verbe à l'impératif. (Type only the conjugated verb.)*

Exemple : « Faites (faire, vous) attention, s'il vous plaît ! »

« _____(travailler, tu) bien ! » (cf. Annexe 10 [p. 44], item 19)

J'ai également introduit dans le nouveau test un item type « Cette phrase contient-elle une erreur ? » (cf. Annexe 10 [p. 44], item 13). J'avais en effet découvert ce type de question en étudiant le test de positionnement *Dialangue*. Ici, l'objectif était à nouveau de varier la nature des items.

Le test est par ailleurs limité à une heure et à une seule passation. De fait, étant donné que les étudiants vont passer ce test chez eux au lieu de venir dans les locaux de l'école sous surveillance des enseignants, il est nécessaire qu'il soit limité dans le temps. Cela les empêchera de chercher les réponses dans une ressource quelconque (internet, dictionnaire, etc.) ou bien de demander à un tiers. De plus, les consignes du nouveau test (cf. Annexe 10 [p. 42]) explicitent l'importance de passer le test en autonomie, dans un endroit calme avec une bonne connexion internet. Une passation à distance élimine également les facteurs souvent présents chez les étudiants nouvellement arrivés en France tels que la fatigue issue du voyage, le décalage horaire ou le choc culturel. Ce sont des effets liés à l'arrivée dans un pays étranger qui risquent d'influencer de façon négative les résultats d'une évaluation. De cette manière, des conditions similaires pour tous et propices au travail sont mises en place, conservant ainsi l'équité du test.

Enfin, l'intégration de l'outil pour insérer des accents, des exemples pour les questions à réponses ouvertes et la présence de consignes claires et détaillées préservent la praticabilité du nouveau test.

1.5.3.1. Les compétences évaluées et le contenu linguistique du nouveau test

Après avoir décidé de la structure de l'épreuve, ainsi que des items qui la constituent, il a fallu choisir les compétences et les éléments linguistiques à évaluer. Ainsi, le nouveau test évalue les connaissances linguistiques (la grammaire et la syntaxe), la compréhension écrite et la production écrite.

Comme nous l'avons vu, le test existant n'évaluait que les connaissances grammaticales des étudiants. J'ai également appris, à travers les questionnaires, que les étudiants et les enseignants à GEM trouvaient que le test original était trop axé sur la grammaire³⁰. De plus, nous avons vu qu'une épreuve ayant pour objectif de déterminer le niveau d'un étudiant doit recouvrir beaucoup de contenu et contenir un nombre important d'items pour pouvoir cerner le niveau réel de la personne. Chaque étudiant n'a pas les mêmes compétences en grammaire, qu'en lexique ou en compréhension écrite, et

³⁰ Par exemple : « Ils (*les tests*) sont trop longs, trop centrés sur la grammaire, utilisent parfois une terminologie peu claire (notamment en niveaux A1 et A2). Il ne faut pas oublier que les étudiants GGSB n'ont pas vocation à écrire des mémoires en français. Ce qui est important pour eux, c'est l'oral. » (cf. annexe 3 [p. 15], question 12, réponse 2), ou bien, « (*Les tests sont*) Trop axé(s) sur le linguistique et pas assez sur le communicatif. » (cf. Annexe 3 [p. 15], question 12, réponse 3).

celles-ci sont parfois partielles. Il ne faut donc pas supposer qu'un test qui n'évalue que la grammaire cerne bien le niveau de l'étudiant.

J'ai donc diversifié les compétences évaluées dans le nouveau test en intégrant, en plus des questions d'ordre grammatical et syntaxique, sept questions de vocabulaire, une compréhension écrite par niveau et une production écrite à effectuer à la fin du test écrit.

Concernant le vocabulaire, j'ai voulu intégrer cette compétence dans le test sans qu'elle y occupe une place trop importante. De fait, dans le cadre d'un test de positionnement le vocabulaire est assez problématique puisque, même si un étudiant ne connaît pas tel ou tel mot en français, il peut très bien avoir le niveau associé à ce mot. De plus, nous ne pouvons pas savoir quels champs lexicaux sont abordés dans tel ou tel pays, contrairement à la grammaire qui est davantage standardisée. Aussi ai-je conçu uniquement sept questions de vocabulaire dans le nouveau test.

Quant à la compréhension écrite, sa prise en compte vient également du questionnaire des enseignants dans lequel ils ont évoqué la nécessité d'inclure l'évaluation de cette compétence dans le nouveau test (*cf.* Annexe 3 [p. 16], question 13).

Enfin, la production écrite fait partie du test dans le but de varier non seulement les compétences évaluées, mais également les types d'évaluations qui composent le processus de positionnement. Une production écrite est en effet une évaluation de la performance, contrairement à l'évaluation des connaissances que représente le test écrit sous forme de QCM. Pour obtenir une vraie image des capacités en français d'un utilisateur de cette langue, il est préférable d'évaluer également la compétence écrite. En revanche, la production écrite ne sera pas notée. La raison de ce choix est qu'une notation de la production écrite augmente le temps nécessaire pour le processus de positionnement, ce qui va contre l'objectif de réduire au maximum le coût de cette procédure pour l'école.

Ainsi, nous avons décidé, avec ma tutrice de stage, qu'au moment de faire passer le test oral, l'enseignant aura la production écrite devant lui pour pouvoir prendre en compte la capacité de l'étudiant à écrire en français lorsqu'il détermine son niveau. Ainsi, après avoir effectué l'entretien oral, l'enseignant détermine le niveau de l'étudiant en fonction des résultats du test oral, du niveau donné par le test écrit et de la qualité de la production écrite. Cette dernière est donc un appui complémentaire qui aide le professeur à placer la personne dans le groupe de niveau approprié.

En ce qui concerne le contenu linguistique du test, je me suis servi du référentiel de l'Alliance Française car cet ouvrage présente, de manière simple, cohérente et compréhensible les éléments de la langue qui correspondent aux niveaux A1-B2. Cette référence est également en lien direct avec le CECRL, ce qui respecte la demande ainsi que la validité du test. Il était préférable de m'appuyer sur

une seule référence étant donné les discordances entre les différentes référentielles relevées dans la partie 3, 1.4.2.

Comme nous l'avons découvert, le test existant n'évaluait pas forcément ce qu'il était censé évaluer en fonction des niveaux, portant ainsi atteinte à sa validité. C'est pour cela que j'ai fait en sorte que le contenu de chaque partie du test soit cohérent avec le niveau de ces parties en m'appuyant sur le référentiel de l'Alliance Française. Ainsi, les éléments évalués de la partie niveau A1 figurent dans le niveau A1 du référentiel, idem pour les niveaux A2, B1 et B2.

En outre, le nouveau test recouvre plus d'éléments linguistiques que l'existant, ce qui est nécessaire pour que le nouveau test de positionnement fournisse le vrai niveau de l'évalué. Le test existant contenait 28 points grammaticaux, tandis que le nouveau en contient 44 (*cf.* Annexe 13 [p.62]). En augmentant le nombre éléments linguistiques de cette manière, la capacité du test à déterminer le niveau est parallèlement accrue puisqu'il évalue ce qui est censé être évalué dans le cadre d'un test de positionnement, c'est-à-dire une grande quantité d'éléments linguistiques qui correspondent aux niveaux A1-B2.

Bien qu'il ait été pertinent et avantageux de me baser sur le référentiel de l'Alliance Française pour choisir le contenu linguistique du test, j'ai estimé qu'il était nécessaire de vérifier l'adéquation de celui-ci avec trois autres sources afin de m'assurer une dernière fois de la validité du contenu linguistique. Il s'est agi premièrement de vérifier que les éléments intégrés dans le test figuraient dans le même niveau que dans le référentiel pour le programme de FLE de GEM³¹. Deuxièmement, trois locuteurs francophones natifs qui ne sont pas enseignants de FLE ont passé le test sur papier pour vérifier que les questions étaient claires et que les réponses n'étaient pas ambiguës. Troisièmement, les enseignants de GEM ont lu le test pour vérifier que le contenu était approprié au niveau.

J'ai voulu m'adresser d'une part aux Français extérieurs à l'école pour vérifier que les réponses étaient claires. J'ai estimé en effet que les questions devaient être tout d'abord claires pour des francophones natifs avant de vérifier leur clarté auprès des étudiants de GEM. Lorsque pour eux la réponse n'était pas claire, je savais que la question n'était pas bonne et l'ai donc modifiée. Quant aux professeurs de FLE de GEM, la volonté de les interroger à propos de l'adéquation du contenu vient non seulement de la nécessité de vérifier la pertinence linguistique du test en fonction des niveaux, mais également d'adapter le test aux attentes de ceux qui enseignent le français à GEM. Après avoir réalisé ces deux étapes, j'ai estimé que le test était valide non seulement selon le CECRL, mais également selon les professeurs de GEM.

³¹ Le contenu linguistique du nouveau test figure dans le même niveau du référentiel de GEM. Cependant, ce dernier est incomplet (*cf.* Annexe 18 [p. 75]), et donc le niveau de certains éléments n'a pas pu être vérifié.

En somme, au vu des compétences évaluées, ainsi que du contenu linguistique plus étendu, le nouveau test est plus complet dans le sens où il mesure la maîtrise de plus de compétences qu'uniquement celle de la grammaire, comprend une évaluation de la performance, et est constitué d'items variés. De plus, grâce à la vérification des éléments linguistiques du test auprès du référentiel de l'Alliance Française et des professeurs de GEM, il cerne avec précision le vrai niveau de langue des étudiants, et est plus fiable.

1.5.3.2.L'échelle de notation

La troisième partie de la conception selon ALTE consiste à établir une échelle de notation. En lien avec les niveaux et les sous-niveaux des cours de FLE à GEM, cette échelle est importante pour que le test fournisse un score qui, par la suite, indique le niveau de l'étudiant. Nous avons vu que le test existant ne contenait pas d'échelle, et que les items n'étaient pas pondérés. Il était par conséquent imprécis dans sa détermination du niveau.

Le nouveau test contient 110 points, un point étant accordé non pas à chaque question mais à chaque réponse que l'évalué doit fournir. Par exemple, les QCM comptent pour un point, ainsi que les questions à réponses ouvertes courtes. En revanche, les exercices lacunaires (items 55 et 91) comptent pour un nombre de points égal au nombre de lacunes dans la question. Par exemple, l'item 55 contient 9 verbes à conjuguer et donc compte pour 9 points :

*55. Complétez le dialogue en conjuguant les verbes au passé composé, à l'imparfait ou au plus-que-parfait.
« Hier soir, ma copine _____(organiser) une fête surprise pour mon anniversaire ! Avant la soirée, elle _____(cuisiner) mon plat préféré, et elle _____(aller) acheter un gâteau délicieux chez le pâtissier du quartier. Quand je _____(rentrer) chez moi du travail, il y _____(avoir) tous mes amis et ma famille ! Nous _____(danser) et nous _____(rigoler). À minuit, nous _____(être) tous fatigués, et donc tout le monde _____(partir). Quel bon anniversaire ! » (cf. Annexe 10 [p. 49], item 55)*

Ensuite, pour que le test fournisse un niveau et un sous-niveau précis, j'ai divisé les points en fonction des sous-niveaux qu'utilise le programme de FLE à GEM pour constituer les groupes, à savoir A1.1, A1.2, A2.1, A2.2, A2.3, B1.1, B1.2, B1.3, B2 et B2+. Plus précisément, j'ai divisé le nombre de points total pour chaque niveau par le nombre de sous-niveaux. Je suis ainsi arrivé à l'échelle de notation suivante :

- Niveau A1 : Items 1 à 25 pour un nombre de points entre 0-25
 - A1.1 : 0-12 points
 - A1.2 : 13- 25 points
- Niveau A2 : Items 26 à 50 pour un nombre de points entre 26-50
 - A2.1 : 26-33 points
 - A2.2 : 34-41 points
 - A2.3 : 42-50 points
- Niveau B1 : Items 51 à 75 pour un nombre de points entre 51-83

- B1.1 : 51-61 points
- B1.2 : 62-72 points
- B1.3 : 73-83 points
- Niveau B2 : Items 76 à 97 pour un nombre de points entre 84-110
 - B2 : 84-97 points
 - B2+ : 98-110

Afin de m'assurer que les sous-parties des quatre parties du test correspondent également aux sous-niveaux, j'ai fait en sorte que le test soit évolutif au niveau de la difficulté. Ainsi, le contenu au début du niveau A1 correspond à ce que l'on aborde au niveau A1.1 à GEM, etc. À cette fin, j'ai à nouveau interrogé les professeurs de GEM dans le but de savoir si les items étaient correctement situés dans les niveaux. En fonction de leurs avis, j'ai remonté ou descendu des items. Le test est ainsi évolutif en difficulté, et le contenu correspond à ce qui est abordé dans les différents niveaux à GEM.

Il est vrai que j'étais dubitatif quant au fonctionnement de cette échelle après l'avoir établie. En effet, un étudiant qui possède un niveau A2.2 peut très bien passer le test entier³², ce qui veut dire qu'évidemment il ne va pas forcément terminer le test entre les items 33 et 41. J'ai donc attendu la phase d'expérimentation pour voir si l'échelle de notation permettait d'interpréter les résultats et ensuite si cette interprétation fournissait le vrai niveau de l'étudiant.

1.5.4. Étape 4 : l'expérimentation et l'analyse des résultats

Lorsque la première version du nouveau test a été achevée et mise en ligne sur la plateforme *Moodle*, la phase de l'expérimentation a commencé. Il était effectivement question de tester le test sur le même terrain, auprès du même public et dans des conditions pareilles que celles de la passation réelle (ALTE, 2011). L'objectif global était donc de vérifier sa capacité à déterminer avec précision le niveau d'un étudiant de français à GEM. Néanmoins, les productions écrites n'ont pas été intégrées dans le test pour la phase de l'expérimentation étant donné que celles-ci serviraient uniquement comme appui lors de la détermination du niveau de l'étudiant, et ne seront pas notées. De plus, il s'agit de ne pas décourager les étudiants de passer le test à cause du temps important qu'exigeraient la passation du test écrit et la production écrite.

J'ai donc demandé aux étudiants de GGSB³³ des niveaux A1.1, A1.2, A2.1, A2.2, B1.1, B1.2, B1.3 et B2 de passer le test. La phase d'expérimentation a duré trois semaines dans le but d'avoir le plus de réponses possibles. De fait, afin de vérifier la fiabilité d'un test ainsi que son bon

³² Dans les consignes du test, j'explique que, si la personne se rend compte que le niveau est trop élevé pour elle, elle doit arrêter le test et ne pas procéder par divination. Or, rien de l'empêche de continuer tout de même. Malheureusement, cela fait partie des choses que l'on ne peut pas contrôler dans le cas d'un test dont la passation s'effectue à distance.

³³ Pour rappel, les étudiants de GGSB représentent le public cible du test car ils sont les plus nombreux.

fonctionnement, il est préférable de le faire tester auprès d'un grand nombre d'étudiants (ALTE, 2011). Au final, j'ai eu 71 passations dont 18 de niveau A1, 32 de niveau A2, 14 de niveau B1 et 7 de niveau B2³⁴. Ces chiffres sont cohérents car, lors du deuxième semestre de l'année 2014/2015, les étudiants de niveau A2 étaient les plus nombreux, tandis que ceux de B2 étaient le moins nombreux.

Les graphiques des données issues de la phase d'expérimentation (*cf.* Annexe 11 [p. 56], figures 1-4) permettent d'analyser les résultats. Les conclusions tirées ont indiqué quels changements il fallait effectuer dans le but d'améliorer le test, tout au niveau de la forme que du fond.

Avant de présenter l'analyse des données de l'expérimentation, il est important de signaler que lorsque les étudiants ont passé le test, ils avaient déjà suivi neuf semaines de cours de français. En prenant cela en compte, nous pouvons supposer qu'ils avaient progressé, et qu'ils ne possédaient donc sans doute pas le même niveau que celui qui leur avait été attribué au début du semestre.

Pour souligner les points importants de l'analyse des résultats, je commencerai par évoquer les résultats de la phase d'expérimentation afin de constater la précision avec laquelle le nouveau test cerne le niveau des étudiants. Je parlerai ensuite de la difficulté du test et de son adéquation dans le cadre du positionnement à GEM, avant de mettre en évidence les données issues d'un questionnaire sur le nouveau test rempli par les étudiants. Enfin, j'expliquerai les changements que j'ai effectués en fonction de l'analyse des résultats.

1.5.4.1. L'analyse de la capacité du nouveau test à déterminer un niveau de langue

Il était tout d'abord nécessaire de constater la précision avec laquelle le test détermine le niveau des étudiants. À cette fin, j'ai comparé leur niveau fourni par le test avec celui du groupe dans lequel ils étaient au moment de la passation.

Pour commencer, le premier graphique (*cf.* Annexe 11 [p. 56], figure 1) montre que la majorité (70%) des étudiants ont reçu un niveau supérieur à celui de leur groupe au moment de la passation, tandis que 16% ont reçu un niveau égal, et 14% un niveau inférieur. Ces données sont cohérentes puisque, comme je l'ai expliqué, les étudiants avaient déjà étudié le français pendant neuf semaines à GEM. Ainsi, il est logique de constater que 70% ont obtenu un niveau plus élevé. En ce qui concerne ceux qui ont obtenu le même niveau, en partant du principe que tous les étudiants ne progressent pas de la même façon, il est normal que certains possèdent encore le même niveau. Enfin, il est vrai que certains ont obtenu un niveau inférieur à celui de leur groupe. Cela peut être expliqué par le fait qu'un groupe de langue n'est jamais complètement homogène par rapport au niveau. Une explication

³⁴ Je me suis efforcé d'obtenir davantage de passations de la part des étudiants en B1 et B2 en leur envoyant des rappels par courriels et en passant dans les cours pour leur demander de passer le test. Toutefois, je n'ai pas eu beaucoup de réponses de leur part. Malheureusement, il s'agit d'un facteur incontrôlable.

possible est que ces étudiants ont été mal positionnés au début de l'année et n'ont jamais été replacés dans le groupe approprié.

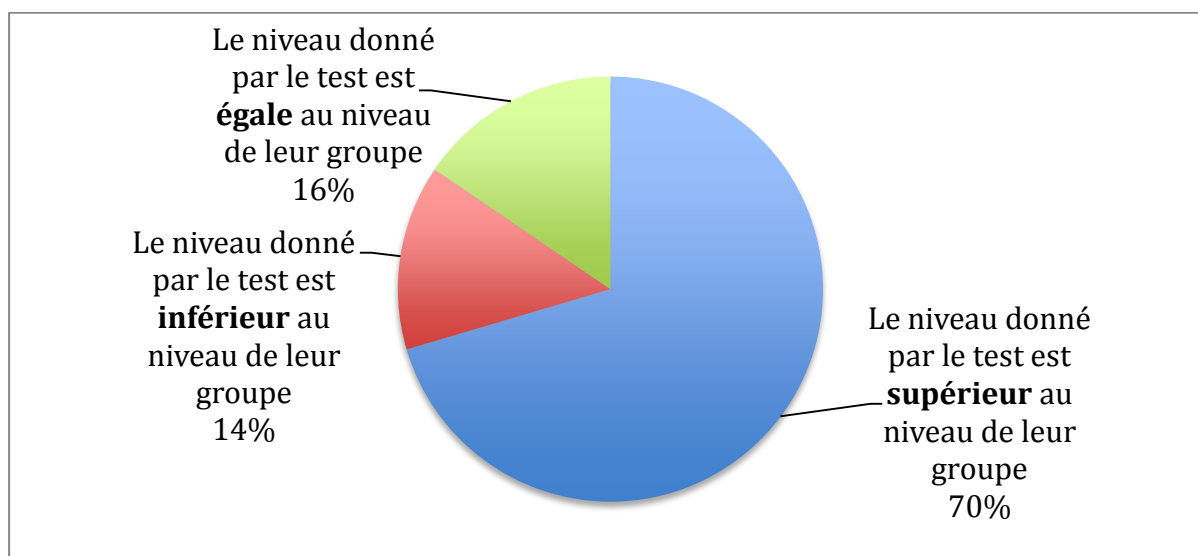


Figure 1 : Pourcentage des étudiants dont le niveau donné par le test était supérieur, égal ou inférieur à leur niveau au moment de la passation.

Il faut toutefois souligner qu'un test de positionnement n'est jamais à 100% parfait dans sa détermination du niveau des apprenants. Les changements de groupes sont d'ailleurs inévitables, ce que j'ai appris grâce à l'entretien avec une enseignante du CUEF (cf. Annexe 17 [p.73], entretien 2, tour de paroles 151-154) :

E : Et donc la première fois que vous l'avez utilisé (le nouveau test de positionnement) (...) est-ce que vous avez dû faire des changements de groupe après ?

L1 : Oui pour certains.

E : C'est (...) toujours le cas j'imagine.

L1 : C'est toujours le cas mais oui.

De plus, cette même école de langue, qui a conçu récemment un test de positionnement autocorrectif, a constaté un nombre de changements plus important qu'avec l'ancien test (cf. Annexe 17 [p. 73], entretien 2, tour de parole 166). Cela montre que les changements de groupe sont non seulement inévitables, mais qu'ils peuvent augmenter, même avec un nouveau test plus facile à faire passer et à corriger, ce qui est le cas avec celui que le CUEF a mis en place récemment. « Mais on a fait un test plus simple que celui que tu as fait (celui utilisé en 2012) » (cf. Annexe 17 [p. 73], entretien 2, tour de parole 168).

J'ai voulu ensuite examiner de manière pointue les différences entre le niveau fourni par le nouveau test et celui du groupe. De fait, lors de la première analyse des résultats, il était facile de constater que les étudiants recevaient rarement le même niveau que leur groupe. J'ai donc voulu noter l'importance de l'écart en termes de sous-niveau afin de mesurer la capacité du nouveau test à cerner

avec précision le niveau d'un étudiant. Un grand nombre d'écarts importants signifierait en effet que le test n'était pas assez précis dans sa détermination du niveau.

Un graphique indique ces décalages (cf. Annexe 11 [p. 56], figure 2). Nous voyons que 46% des étudiants ont obtenu un niveau supérieur à celui de leur groupe, et 10% un sous-niveau inférieur. Quant aux écarts importants, à savoir 3 ou 4 sous-niveaux inférieurs ou supérieurs au niveau réel, ceux-ci sont négligeables par rapport aux autres écarts. Aussi les ai-je estimés anodins et n'ont-ils pas influencé l'analyse des résultats.

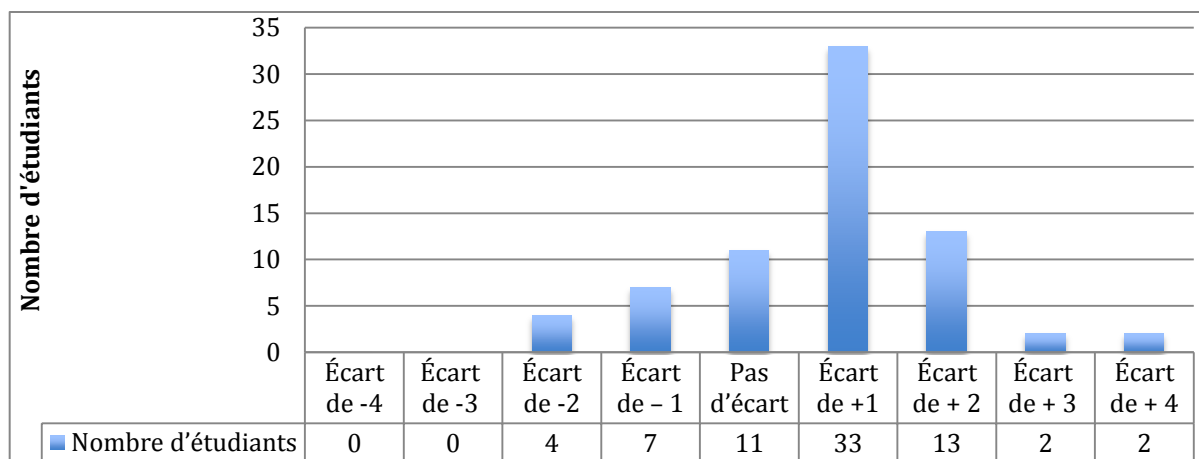


Figure 2 : Écart en sous-niveaux entre le niveau réel des étudiants et le niveau fourni par le nouveau test de positionnement

Enfin, les données les plus intéressantes à examiner sont celles qui montrent que 18% ont obtenu deux sous-niveaux supérieurs à leur niveau réel. Cela a montré que le test était un peu trop facile sur le plan linguistique, et qu'il fallait donc rendre quelques items plus difficiles. Cela dit, en tenant compte du fait que les étudiants avaient suivi neuf semaines de cours, ce graphique montre que le test était précis étant donné que la majorité (70%) ont reçu un niveau soit légèrement plus élevé (i.e. par un seul sous-niveau), soit identique à celui de leur groupe.

En partant du principe que les étudiants avaient progressé depuis le début du semestre, et qu'ils avaient donc un niveau légèrement supérieur à celui de leur groupe, le tableau 8 ci-dessous justifie à nouveau ces écarts.

	A1.1	A1.2	A2.1	A2.2	B1.1	B1.2	B1.3	B2	B2+
Le nombre d'étudiants de chaque niveau au moment de la passation	5	13	16	16	10	1	3	7	0
Le nombre d'étudiants ayant tel ou tel niveau du test	0	5	11	16	12	9	8	4	0

Tableau 8 : La mise en évidence des écarts en sous-niveaux.

De fait, j'ai comparé le nombre d'étudiants de chaque niveau et sous-niveau avec le niveau et le sous-niveau au-dessus, ce qui est indiqué les flèches. Nous remarquons que les deux chiffres ne sont jamais très éloignés (une différence de quatre étudiants ou moins), mis à part le niveau B1.2 au niveau B1.3 et le niveau B2 au B2+ (marqué par une flèche rouge). En ce qui concerne les premiers, cette incohérence peut être expliquée par le manque d'étudiants de niveau B1.2 dans la phrase d'expérimentation. Concernant les niveaux B2 et B2+, cette incohérence a indiqué que le niveau B2 du test était éventuellement trop difficile, et qu'il fallait donc examiner les items de cette partie.

Ainsi, la conclusion tirée de la figure 2 selon laquelle le test cerne bien le niveau a été renforcée. Suite à cette étude, j'ai pu conclure que le test était globalement fiable et valide, et que l'échelle de notation fonctionnait bien.

1.5.4.2. La vérification de la difficulté du nouveau test de positionnement

Après avoir examiné les résultats, dans l'optique de savoir avec quel degré d'efficacité et de précision le test cernait le niveau des étudiants, j'ai effectué une analyse afin de m'assurer que la difficulté du test était appropriée aux niveaux. À cette fin, j'ai premièrement calculé les notes moyennes par niveau, tout étudiant confondu. La graphique de ce calcul (*cf.* Annexe 11 [p. 58], figure 4), démontre une moyenne descendante, en allant du niveau A1 au niveau B2. Étant donné qu'il s'agit d'étudiants de tout niveau, et que le nombre d'étudiants des niveaux B1 et B2 est plus petit par rapport aux autres niveaux, il est normal que nous observions une telle tendance des moyennes. Cependant ces données ont confirmé que le niveau B2 était un peu trop difficile.

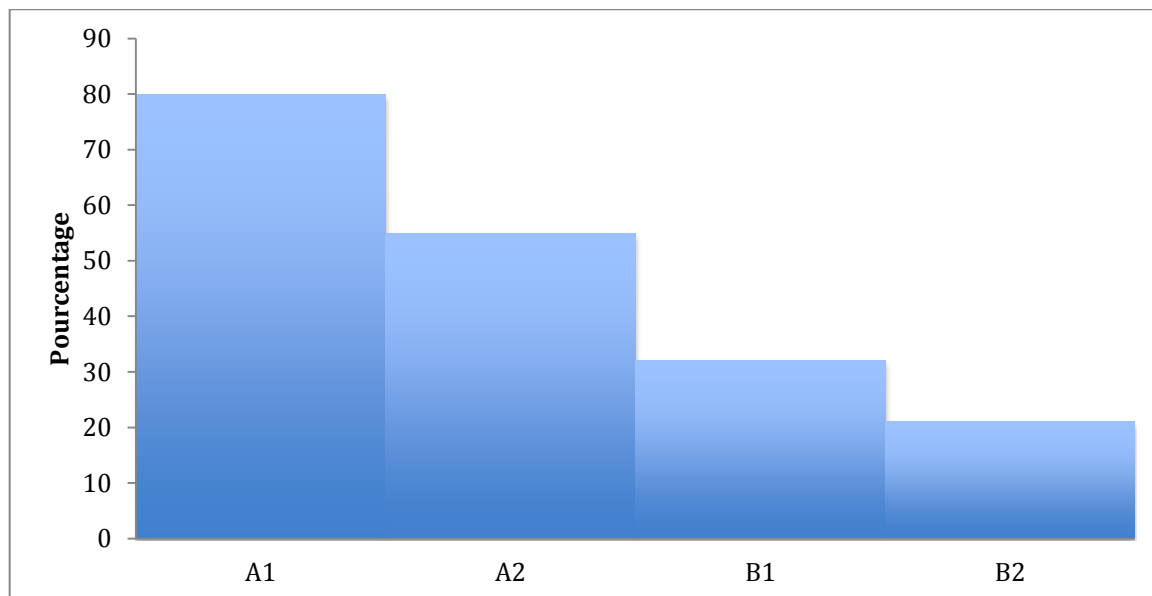


Figure 3 : La note moyenne en pourcentage par niveau, tout étudiant confondu

Deuxièmement, j'ai chiffré le pourcentage de bonnes réponses des étudiants dans le niveau auquel ils appartiennent. Autrement dit, j'ai sélectionné les résultats des étudiants en groupe de niveau A1, et j'ai calculé leur moyenne uniquement dans la partie A1 du test. J'ai ensuite effectué la même démarche pour les étudiants de niveau A2, B1 et B2. Les résultats de cette analyse (*cf.* Annexe 11 [p. 58], figure 5) montrent une moyenne qui se situe entre 43% et 62% et qui, de plus, reflètent les données du calcul des moyennes tout étudiant confondu. Cette deuxième analyse de la difficulté confirme non seulement le niveau de difficulté des quatre sections du test, mais plus spécifiquement que la difficulté du niveau B2 était trop élevée dans la première version.

Enfin, j'ai examiné de manière plus précise le niveau de difficulté de chaque item en analysant tout d'abord la facilité moyenne de tous les items. Celle-ci s'élève à 47%, ce qui est considéré comme une facilité moyenne acceptable (McNamara, 2000). Ensuite, nous avons décidé, avec ma tutrice de stage, que dans le cadre de notre test de positionnement, la facilité la plus basse acceptable était de 20%. De fait, une facilité inférieure à 20% signale un item trop difficile, même pour certains étudiants du niveau dans lequel figure l'item. J'ai donc modifié ces questions pour les rendre plus faciles.

Par ailleurs, il était normal de constater une facilité de 80% à 90% pour certains items aux niveaux A1 et A2 car ce sont des étudiants ayant déjà eu des cours de français à GEM, qui ne sont donc pas des débutants complets. Cela dit, j'ai estimé que l'item ayant eu une facilité de 100% était trop facile. Ainsi était-il nécessaire de le rendre plus difficile³⁵.

1.5.4.3. Questionnaire sur le nouveau test auprès des étudiants

³⁵ Les changements effectués dans le nouveau test seront évoqués dans la partie 3, 1.6.

La dernière étape de l'analyse des résultats du nouveau test a pris la forme d'un questionnaire réalisé auprès de 37 étudiants l'ayant passé lors de l'expérimentation. L'objectif était, d'une part de savoir si les consignes et les questions étaient claires et compréhensibles du point de vue des étudiants et, d'autre part, de recueillir des commentaires par rapport au test afin de l'adapter au public présent à GEM et donc de l'améliorer. En tant que concepteur, il se révèle difficile de considérer le test du point de vue de ses destinataires. Aussi est-il bénéfique d'interroger le public cible sur l'épreuve pour s'assurer de sa clarté.

Tout d'abord, ce questionnaire a démontré que la majorité des étudiants ont trouvé le test mieux que celui qu'ils avaient passé en arrivant à GEM, mettant ainsi en évidence son adaptation aux attentes des destinataires (*cf.* Annexe 12 [p. 60], question 3). J'ai ensuite appris que 97% des étudiants avaient trouvé les consignes générales claires, ce qui est important pour garantir l'équité de l'épreuve (*cf.* Annexe 12 [p. 59], question 2). De même, la troisième question a montré que la plupart des étudiants avaient compris la consigne de chaque question (*cf.* Annexe 12 [p. 59], question 2). Cette question a toutefois récolté quelques réponses « non », m'incitant à examiner avec attention les items afin de déceler les éventuels problèmes.

En outre, les commentaires généraux que j'ai recueillis à travers la question 4 du questionnaire étaient particulièrement utiles car ils ont permis de savoir quels changements spécifiques il fallait effectuer pour que le test soit plus clair et mieux adapté au public présent à GEM. Par exemple, le commentaire 10 (*cf.* Annexe 12 [p. 60], question 4) m'a incité à examiner en détail l'item 55. Je l'ai ensuite modifié afin de vérifier sa clarté pour l'étudiant et de m'assurer qu'il était valide.

“The only thing I didn't like about this new test was the question with passé composé, imparfait, and plus que parfait. I am familiar with all three forms. The question was confusing. If the question had been more obvious, I think I could have scored better on that problem.” (*cf.* Annexe 12 [p. 60], question 4, réponse 10)³⁶.

De même, j'ai appris que certains étudiants trouvaient le nouveau test trop long, ce qui renvoie à la praticabilité de l'examen. Or, il n'était pas possible de répondre à ses remarques en réduisant la longueur du test puisqu'il ne faut jamais porter atteinte à la fidélité ou à la validité d'une épreuve en vue de favoriser d'autres enjeux liés à l'évaluation. Ainsi n'ai-je pas jugé bon de réduire le nombre d'items pour rendre le test plus court.

« Quel que soit l'usage envisagé, une seule règle subsiste : dans l'étendue et dans la forme du mode d'évaluation, jamais on ne sacrifiera la validité à la recherche de l'économie ou de la facilité d'utilisation ; ces deux aspects doivent être conciliés dans la mesure du possible, mais toujours la validité prévaut » (Bonboir, 1972 : 80).

³⁶ « La seule chose que je n'aie pas aimée du nouveau test était la question sur le passé composé, l'imparfait et le plus-que-parfait. Je connais ces trois formes. La question n'était pas claire. Si la question avait été plus évidente, je pense que j'aurais eu plus de points. » (*cf.* Annexe 12 [p. 60], question 4, réponse 10, trad. libre).

1.5.5. Étape 5 : Spécifications du test final

La dernière étape de la conception d'un examen d'après ALTE est la concrétisation des spécificités de l'épreuve. En d'autres termes, ce sont des changements effectués au test en fonction des résultats de l'expérimentation. La version définitive du nouveau test a ainsi été finalisée.

Bien que la conclusion générale tirée de l'analyse des résultats ait été que le nouveau test de positionnement cerne avec assez de précision le vrai niveau des apprenants, j'ai tout de même choisi d'apporter des changements pour l'améliorer, en m'appuyant sur les données issues de la phase d'expérimentation. Il s'agit de modifications effectuées en fonction, d'une part de l'analyse de la difficulté de l'examen et, d'autre part, des commentaires du questionnaire réalisé auprès des étudiants.

Premièrement, j'ai modifié les items que je jugeais trop difficiles ou trop faciles. Comme je l'ai évoqué, quelques items étaient trop faciles, tandis que d'autres étaient considérés trop difficiles. Concernant les derniers, j'ai soit modifié ces questions pour les rendre plus faciles, soit les ai descendues dans le test pour qu'elles soient davantage cohérentes avec la difficulté évolutive de l'épreuve.

En examinant ces changements (*cf.* Annexe 14 [p. 63]), nous notons que la plupart figurent dans le niveau B2 du test. Cela est logique étant donné que l'analyse de la difficulté du test a montré une facilité trop basse dans cette partie du test (*cf.* Annexe 11 [p. 58], figures 4 et 5). D'ailleurs, c'est dans le niveau B2 du test que nous trouvons le plus de questions avec une facilité inférieure à 20%, ce qui a entraîné une modification de ces items. De plus, il est possible que le niveau B2 de GEM soit généralement moins avancé en comparé au niveau des référentiels. Cela expliquerait pourquoi la moyenne s'est montrée assez basse après l'expérimentation. Cela dit, il était nécessaire de modifier cette partie pour qu'elle soit légèrement plus facile.

Dans un deuxième temps, j'ai effectué des changements en fonction des commentaires récoltés dans le questionnaire rempli par les étudiants (*cf.* Annexe 15 [p. 64]). L'objectif était à nouveau de considérer le point de vue ainsi que les avis du public cible, pour que le test leur soit adapté. Cependant, je n'ai pas pu prendre en compte tous les commentaires. Par exemple, les remarques sur la limite de temps (*cf.* Annexe 12 [p. 60-61], question 4, réponses 7 et 15) n'ont pas été prises en compte car cette limite de temps est nécessaire pour l'équité de l'épreuve. De plus, seulement huit étudiants ont consacré une heure entière à compléter le test, montrant que, pour la grande majorité, la limite de temps était suffisante. En outre, pour que les étudiants gèrent correctement le temps et puissent compléter le test en une heure, les consignes générales du test soulignent qu'il faut consacrer environ 30 secondes à chaque item.

Suite à ces changements, j'ai pu concevoir un nouveau test de positionnement qui répond au besoin qui a motivé la commande, à savoir la réduction du coût du positionnement des étudiants.

Ainsi, la passation s'effectue en ligne et à distance, éliminant la nécessité d'une correction par les enseignants. De plus, ce nouveau test est adapté au public et au programme de français. Il est par ailleurs en accord avec les niveaux de langue selon le CECRL et le référentiel de l'école. Son contenu et l'évolution de sa difficulté ont été vérifiés et confirmés par les enseignants de GEM. Pour toutes ces raisons, le nouveau test de positionnement est valide.

2. La conception de la fiche pour le test oral

Lorsque nous évaluons les compétences d'un apprenant en français dans le but de déterminer son niveau, il est nécessaire d'évaluer toutes les compétences, y compris la compréhension et la production orales, autrement dit celle de l'interaction orale. Ces dernières sont en effet mobilisées lors d'une interaction entre deux personnes qui jouent alternativement le rôle d'interlocuteur et de locuteur (Conseil de l'Europe, 2001). Un entretien est ainsi un exemple typique de l'évaluation de cette interaction orale.

Une mesure complète de ses compétences permet d'obtenir une image globale et juste de son niveau, d'où la conception d'un guide d'entretien pour évaluer la capacité de l'étudiant à interagir à l'oral à travers une conversation avec un enseignant de GEM. Plus précisément, l'enseignant pose des questions à l'étudiant et mesure le niveau de sa compréhension et de sa réponse, pour ensuite lui donner un niveau à l'oral situé entre A1 à B2.

Concernant le contexte à GEM et les compétences essentielles, j'ai appris, grâce à l'entretien avec ma tutrice de stage, que la compréhension et la production orales priment dans les cours de français.

« nous ici ça va être très difficile (de constituer les groupes avant que les étudiants n'arrivent à Grenoble) parce que ce qui va primer c'est l'oral et qu'on n'aura pas autant d'affinités (et) que de toute façon on fera l'entretien oral » (cf. Annexe 16 [p. 66], entretien 1, tour de parole 24).

Il est donc facile de comprendre que l'entretien oral est une étape clé dans le processus de positionnement des étudiants. Ainsi, la capacité du test oral à cerner le niveau des étudiants en compréhension et en production orales est extrêmement importante. Cette fiche doit donc être la plus valide et fiable possible, et fournir une image la plus exacte possible des compétences de l'évalué à l'oral.

En outre, ma tutrice de stage a souligné que les professeurs de français à GEM étaient généralement satisfaits avec le guide, et qu'elle souhaitait donc que je conserve sa forme. Ainsi, l'objectif de cette tâche était d'élaborer un nouveau guide pour le test oral organisé de la même manière que l'original, mais plus valide.

2.1. L'analyse de la fiche existante pour le test oral

À l'instar de la conception du test écrit, j'ai commencé l'élaboration de la fiche pour le test oral par une analyse de l'existante (cf. Annexe 7 [p. 37]). Premièrement, comme il s'agit de l'évaluation de la performance (en l'occurrence, l'interaction orale), et donc d'une évaluation subjective, l'intégration dans le guide de critères précis et d'une grille d'évaluation est essentielle. Or, en examinant la fiche existante pour le test oral, nous constatons qu'elle contient une grille, mais qu'il manque des critères qui décrivent les performances attendues (cf. Annexe 7 [p. 37], « Niveau des réponses »). De ce fait, le guide existant ne permet pas de situer la qualité de la production dans la grille. Voici un extrait de cette dernière :

<i>B/ CONVERSATION</i>	<i>Niveau des réponses</i>			
<i>1/ Avez-vous déjà étudié le français ?</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
<i>2/ Pendant combien de temps ?</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>

Figure 4 : Extrait de la grille de notation de la fiche existante pour le test oral.

En étudiant les tests oraux administrés à la rentrée de janvier 2015 (cf. Annexe 6 [p. 25, 30, 35]), j'ai pu relever certains problèmes avec la fiche existante. Par exemple, les enseignants entouraient parfois des chiffres pour indiquer la qualité de la réponse, mais ils ne prenaient pas forcément en compte ces informations pour déterminer ensuite le niveau de l'étudiant. Par ailleurs, les évaluateurs marquaient parfois entre deux chiffres ou bien n'entouraient pas de chiffres (cf. Annexe 6 [p. 25 et 35], tests 1 et 3). Cela est d'ailleurs plutôt compréhensible étant donné le manque de critères et d'explication correspondant à ces chiffres. En somme, ils n'avaient pas d'appui concret lors du jugement de la compétence orale, ce qui réduisait ainsi la justesse de l'évaluation.

Non seulement la fiche manquait de critères pour aider l'enseignant à porter un jugement juste sur le niveau de réponse de l'étudiant, mais l'organisation de la fiche manquait de cohérence. De fait, les premières questions (cf. Annexe 7, « A/ Présentation »), qui sollicitent une présentation personnelle de la part de l'étudiant, et donc des actes de parole de niveau A1, ne sont pas liées à la grille. L'enseignant ne peut pas donc juger de la qualité des réponses à ces questions. Ensuite, les interrogations suivantes portent sur l'expérience que la personne a eue avec la langue française (cf. Annexe 7 « B/ Conversation » questions 1 à 3). Ce sont des questions nécessaires, mais elles ne sollicitent pas d'acte de parole correspondant à un niveau spécifique. Ainsi n'aident-elles pas à situer la personne dans l'échelle des niveaux A1-B2 à l'oral.

Enfin, dans la création d'une procédure de positionnement, je visais également à mettre en place des outils cohérents entre eux. Le test oral et le test écrit devaient ainsi fournir les mêmes niveaux et sous-niveaux pour que l'évaluateur puisse comparer les deux résultats, pour prendre ensuite une décision juste et réfléchie par rapport au niveau de l'étudiant. Or, la fiche existante ne permettait pas de déterminer le sous-niveau de l'étudiant, ce qui est le cas avec le nouveau test écrit. J'ai donc voulu créer une nouvelle fiche pour le test oral qui aille dans le même sens.

2.2. L'élaboration de la nouvelle fiche pour le test oral

Suite à l'analyse présentée ci-dessus, j'avais deux objectifs à atteindre. Je voulais premièrement créer une nouvelle fiche pour le test oral (*cf.* Annexe 8 [p. 38]). Pour ce faire, j'ai ajouté, modifié et réorganisé les questions pour que les réponses de l'étudiant donnent à l'enseignant une meilleure idée de son niveau. En effet, en l'amenant à mobiliser différents actes de parole d'un niveau donné à travers l'entretien, l'évaluateur se rend compte du niveau des compétences de la personne en interaction orale³⁷. Je voulais par ailleurs refaire la grille et rédiger des critères qui correspondent à cette dernière pour guider davantage le positionnement des étudiants dans les niveaux.

Deuxièmement, je souhaitais concevoir un guide, destiné aux enseignants qui leur explique comment se servir de la fiche pour le test oral (*cf.* Annexe 9 [p. 40]). Cette volonté vient de la nécessité de former les enseignants, lorsqu'il s'agit de l'évaluation de la performance, pour qu'ils exploitent au mieux la fiche et pour que la subjectivité ne soit pas trop présente.

Au vu de ces objectifs, j'ai commencé par élaborer une fiche comprenant des questions dont la difficulté est progressive. Ainsi l'enseignant sait si l'étudiant dispose des compétences orales nécessaires pour comprendre et pour répondre à des questions de telle ou telle difficulté, et peut ainsi cerner le niveau qu'il possède à l'oral. Il était donc important de faire en sorte que la première case de questions corresponde au niveau A1, la deuxième au niveau A2, la troisième au B1 et la quatrième au B2. Ensuite, en fonction du niveau de compréhension des questions, ainsi que des réponses obtenues, l'évaluateur détermine le niveau de l'étudiant en interaction orale, en cochant une des trois cases de la grille d'évaluation, à droite des questions.

Cette grille, qui figure sur la fiche, est divisée en trois cases : « non-acquis », « en cours d'acquisition » et « acquis ». C'est dans ces cases que l'enseignant marque avec quel degré de maîtrise l'étudiant a compris et a répondu aux questions. Ces termes sont appropriés pour plusieurs raisons. Premièrement, j'ai constaté l'usage récurrent de ses termes pour décrire le niveau dans les tests de positionnements. De plus, ils sont pertinents pour positionner la personne dans les niveaux et les sous-niveaux. En effet, il ne s'agit pas de donner une note sous forme de chiffre, mais de savoir où se situe

³⁷ Avant de commencer l'entretien, les enseignants expliqueront à l'étudiant qu'il doit répondre aux questions avec des phrases complètes pour éviter qu'il ne réponde par "oui" ou "non". Cette information figure dans le guide pour la fiche.

l'apprenant en terme d'acquisition, ou non, des actes de langage de tel ou tel niveau. Ces termes permettent donc à l'enseignant de donner le sous-niveau, ce qui est important, d'une part pour lier le test oral au test écrit et, d'autre part, pour déterminer le niveau de l'étudiant aussi précisément que possible. Les sous-niveaux ont été établis ainsi : « non-acquis » = les sous-niveaux .1 et .2, et « en cours d'acquisition » = les sous-niveaux .3 et .5³⁸. La catégorie « acquis » ne contient pas de sous-niveau car si l'enseignant y positionne la personne, celle-ci a atteint le niveau et se situe donc à un niveau supérieur.

Afin de rédiger des questions d'une difficulté cohérente avec les niveaux A1 à B2, je me suis basé sur deux normes : premièrement, sur les actes de paroles évoqués dans *Le Référentiel de Programmes de l'Alliance Française* et, deuxièmement, sur l'échelle globale des niveaux communs de compétences du CECRL (Conseil de l'Europe, 2001 : 25).

Le référentiel de l'Alliance Française donne effectivement les actes de paroles qui correspondent à chaque niveau. En appuyant sur eux, j'ai pu formuler des questions pertinentes pour l'évaluation de la compétence à interagir de chaque niveau (dans la fiche, les niveaux sont divisés en quatre cases qui correspondent chacune à un niveau : A1, A2, B1 et B2). De plus, ma tutrice de stage et moi avons convenu que ces questions correspondent aux niveaux dans le cadre des cours à GEM.

Par exemple, en regardant la première case, qui correspond au niveau A1, nous constatons que les questions qui s'y trouvent demandent que l'étudiant se présente, ce qui est un acte de parole de niveau A1 (L'Alliance Française de Paris Ile-de-France, 2008). En mobilisant cet acte de langage, l'évalué doit également se servir des éléments linguistiques du niveau A1, à savoir du vocabulaire des nationalités et des chiffres, ainsi que du présent de l'indicatif (*idem.*).

De même, en regardant la case du niveau A2 (*cf.* Annexe 8 [p. 38]), nous voyons que la mise en œuvre de l'acte de parole « parler du passé » est induite par la question 8. Ainsi, si la personne possède un niveau A2, elle pourra utiliser le passé composé et l'imparfait, deux temps verbaux qui doivent être maîtrisés à ce niveau d'après le référentiel. Quant au niveau B1, les questions dans la case 3 exigent que l'étudiant parle du futur (*cf.* Annexe 8 [p. 38], question 10 et 11) en se servant du conditionnel présent et du futur simple. Enfin, les questions du niveau B2 cherchent à savoir si l'étudiant peut exprimer la cause (question 12) et émettre des hypothèses (question 13) en mobilisant le conditionnel passé et le plus-que-parfait ensemble, des actes de paroles qui sont, encore une fois, associés au niveau B2.

³⁸ Il est vrai que l'enseignant doit choisir un sous-niveau, ce qui est une décision forcément subjective. Comme nous l'avons appris, il est impossible d'éliminer entièrement la subjectivité lorsqu'il s'agit de l'évaluation de la performance. Le choix entre les sous-niveaux met effectivement en évidence ce fait. Les critères aident toutefois à ce que la subjectivité n'influence pas trop la prise de décision.

Quant à l'échelle globale des niveaux du CECRL, elle a permis de vérifier que les questions de chaque niveau du guide pour le test oral étaient cohérentes avec les descriptions des niveaux A1-B2 qui figurent dans le tableau du CECRL. Par exemple, nous voyons dans la description globale du niveau B1 qu'un apprenant possédant ce niveau peut parler de ses espoirs et de ses projets. Dans la fiche pour le test oral figurent donc les questions « Qu'aimeriez-vous faire après vos études ? », ainsi que « Quels sont vos projets professionnels ? ». Chaque partie du test oral est donc reliée aux descripteurs des niveaux dans le CECRL, assurant ainsi la validité de l'épreuve.

Pour terminer, la question 14, (« Habitez-vous seul, en colocation ou en famille d'accueil (nationalité(s) de(s) colocataire(s) ou de la famille d'accueil ? »), est importante puisque si l'étudiant vit dans un logement où le français est la langue dominante, il est fort possible qu'il progresse plus vite que quelqu'un qui n'est pas exposé régulièrement à la langue française. C'est par ailleurs à l'enseignant de décider si ces informations doivent influencer le positionnement de la personne dans les niveaux. Si oui, il doit écrire un commentaire à ce propos en bas de la fiche où se trouve un espace à cet effet. Cependant, cette question ne correspond pas à un acte de parole particulier d'après le référentiel de l'Alliance Française. Aussi se situe-t-elle en dessous des cases, et l'enseignant ne coche-t-il pas de case à droite en fonction de la réponse de l'étudiant.

2.3. L'élaboration du guide pour les enseignants

Pour que les enseignants sachent comment utiliser au mieux la nouvelle fiche pour le test oral, et pour que le jugement ne soit pas trop subjectif, j'ai rédigé un guide explicatif. C'est en effet ici que se trouvent les critères qui décrivent les performances attendues pour les trois catégories de la grille. Des explications synthétiques des critères figurent également au dos de la fiche pour que les professeurs puissent s'y référer lors de la passation (*cf.* Annexe [p. 39], « Explication synthétique des critères »). De fait, ces critères détaillent les performances attendues, de sorte que l'évaluateur puisse repérer des indices dans la compréhension et dans l'expression orales de l'étudiant. Ces compétences indiquent sa maîtrise des actes de parole et, de ce fait, son niveau de langue.

Comme j'ai estimé que le test oral existant était trop subjectif à cause d'un manque de critères et de grille, les premiers visent à réduire la subjectivité. Il est toutefois important de préciser que l'objectif de cette nouvelle fiche pour le test oral n'est pas d'éliminer entièrement la subjectivité. Au contraire, une interaction entre l'enseignant et l'étudiant nous un ressenti sur les capacités de la personne que nous n'obtenons pas avec le test écrit car ce dernier mesure uniquement les connaissances linguistiques, et non pas la performance. À force d'observer les différents niveaux de langue en évaluant la compétence orale des étudiants à GEM, les professeurs expérimentés arrivent à déterminer le niveau tout simplement en les écoutant parler. Il s'agit d'un ressenti qui vient avec l'expérience.

Toujours concernant le guide pour les enseignants, en plus des critères, s'y trouvent des descriptions des éléments linguistiques à chercher dans les réponses des étudiants (*cf.* Annexe 9 [p. 40], « L'organisation de la fiche »). Ce sont des informations essentielles pour l'évaluateur car elles lui permettent de constater avec quelle maîtrise linguistique l'étudiant s'exprime à l'oral. Ces descriptions mettent en évidence quels temps verbaux et quel vocabulaire devraient être employés pour mobiliser correctement l'acte de parole. De plus, elles décrivent l'aisance avec laquelle l'étudiant doit comprendre et répondre aux questions en fonction des trois catégories de la grille (i.e. possibilité de répétition et degré de réussite de la communication). Ces explications, couplées avec les critères, mettent à disposition de l'enseignant une base fiable pour évaluer le niveau à l'oral de l'étudiant correctement et avec précision.

2.4. Conclusion sur la conception de la nouvelle fiche pour le test oral

En comparant la nouvelle fiche pour le test oral avec l'ancienne, il est facile de constater que la première est davantage valide étant donné que les questions sont organisées de manière cohérente dans les niveaux A1-B2. De plus, grâce à la grille d'évaluation et aux critères décrivant les performances attendues, le programme de FLE à GEM peut désormais évaluer de manière plus pointue l'interaction orale. La constitution de groupes de niveau homogène sera ainsi facilitée.

Je n'ai pas pu tester la nouvelle fiche pour l'entretien oral à cause de la courte durée du stage. De plus, j'ai jugé l'expérimentation du test écrit en ligne plus importante étant donné que les étudiants devront le passer en autonomie et à distance. Il était ainsi primordial de m'assurer de sa praticabilité et de sa validité afin d'éviter d'éventuelles péripéties.

En définitive, ce test représente une évaluation subjective critériée qui mesure la capacité d'interagir à l'oral (donc la compréhension et l'expression orales). Elle est équilibrée par le test écrit en ligne, forcément plus objectif, qui évalue les connaissances en termes de grammaire, syntaxe, vocabulaire, production et compréhension écrites. Par ailleurs, le test oral de GEM se rapporte à l'évaluation selon la perspective actionnelle.

Ainsi, les outils évaluatifs qui constituent la procédure du placement des étudiants dans les groupes de niveaux appropriés sont-ils diversifiés dans leur nature que dans leur finalité, permettant de fait de cerner le vrai niveau des étudiants.

3. La conception de l'auto-évaluation

Comme je l'ai déjà évoqué, l'intérêt d'une auto-évaluation pour compléter la procédure de positionnement des étudiants dans les groupes de niveau a été suscité lors d'une discussion informelle avec l'une des professeurs d'espagnol de GEM. Cette personne a en effet expliqué que le programme d'espagnol langue étrangère se sert d'une auto-évaluation lorsqu'un doute est relevé par rapport au

niveau d'un étudiant. Ainsi, si les résultats du test oral et du test écrit ne correspondent pas, les enseignants d'espagnol consultent le niveau donné par l'auto-évaluation. Ceci nous a ensuite mené à nous interroger sur l'éventuelle utilité d'intégrer une auto-évaluation dans le dispositif de positionnement.

Je me suis par ailleurs aperçu de l'intérêt d'une auto-évaluation grâce à mes lectures. Au vu donc de l'attention accordée à l'auto-évaluation, et surtout de l'intérêt qu'elle représente pour diversifier le processus de positionnement en variant le type d'outils de mesure, la mission de créer une auto-évaluation m'a été confiée.

Pour concevoir l'auto-évaluation (*cf.* Annexe 19 [p. 78]), je me suis basé sur les actes de paroles de chaque niveau tels qu'ils sont présentés dans la même référence de langue utilisée pour élaborer le test en ligne et la fiche pour le test oral. De fait, l'auto-évaluation contient pour chaque niveau un paragraphe qui décrit ce qu'un locuteur ayant le niveau en question est capable de dire et de comprendre, à l'oral et à l'écrit, en français. Dans le but de confirmer l'adéquation de ces actes de langage par rapport aux niveaux, je me suis également appuyé sur *Tests de niveau*, un ouvrage réalisé par Tagliante (1998). Cette référence contient pour chaque niveau une auto-évaluation, et a permis de rédiger des paragraphes qui représentent les compétences écrites et orales de chaque niveau.

Par ailleurs, pour que les descriptions soient compréhensibles par tous les étudiants, elles ont été rédigées en anglais et ne contiennent pas de métalangage (*i.e.* appellation des temps verbaux). L'objectif était en effet de décrire les niveaux de manière simple et claire en évoquant les thèmes associés (la famille, le travail, les études, etc.), ce dont on doit pouvoir parler (le passé, le présent, le futur), et dans quels buts (décrire, argumenter, expliquer). De plus, les textes évoquent ce qui est encore difficile à faire au niveau donné, ce qui permet de distinguer de manière plus claire les différences entre les niveaux (« Je commence à comprendre certains films, mais cela reste encore plutôt difficile pour moi », ou bien « J'ai toujours des difficultés à comprendre un locuteur natif s'il parle vite ou avec un autre natif », par exemple). Cela aide aussi les apprenants à se rendre compte, non seulement de ce qu'ils peuvent faire, mais également de ce qu'ils ne peuvent pas encore faire, favorisant ainsi une meilleure appréhension de leur niveau en français.

Comme je l'ai évoqué, un des défauts du processus de positionnement existant était que l'étudiant recevait parfois le test débutant au lieu du test avancé (ou vice versa). J'ai donc voulu, premièrement, créer un outil qui puisse résoudre ce problème. C'est pour cela qu'au départ, la finalité de l'auto-évaluation était de mener l'étudiant vers le test écrit débutant (A1/A2) ou avancé (B1/B2). Ainsi, la personne aurait fait l'auto-évaluation en ligne juste avant de passer le test écrit autocorrectif. Si elle s'était auto-évaluée à un niveau A1 ou A2, elle aurait été menée au test débutant. Au contraire, un niveau B1 ou B2 issu de l'auto-évaluation l'aurait dirigée vers le test avancé. Cependant, suite à

l'expérimentation de deux versions différentes de l'auto-évaluation³⁹, je me suis rendu compte qu'aucune des deux n'était assez fiable pour servir à cette fin.

C'est pour ces raisons que l'auto-évaluation finale (*cf.* Annexe 19 [p. 78]) servira à confirmer ou à infirmer la décision de l'examineur. Ce dernier pourra le consulter en cas de doute ou par simple curiosité. Ce n'est toutefois pas une obligation de prendre en compte le niveau issu de l'auto-évaluation. Pourtant, même si celui-ci n'influence pas le positionnement, il se révèle intéressant de constater l'écart, ou non, entre le niveau qu'un étudiant s'auto-accorde et le niveau que détermine l'enseignant. Ces deux perspectives convergent-elles ? Pourquoi ou pourquoi pas ? Qu'est-ce que cela indique sur le public à GEM ? Une telle comparaison confirmerait éventuellement la volonté des étudiants de progresser plus vite en français. En d'autres termes, étant donné qu'ils ont tendance à vouloir être placés dans un niveau supérieur à leurs compétences réelles, il se peut qu'ils se donnent souvent un niveau plus élevé par rapport à celui que déterminent le test écrit en ligne et le test oral.

Cela a d'ailleurs été le cas lors des phases d'expérimentation où entre 30% et 50% des étudiants se sont surévalués. Ce constat m'a à nouveau incité à émettre des hypothèses à ce propos. Est-il dû effectivement à leur volonté de suivre un cours de niveau plus avancé, ou bien la cause vient-elle du fait que ces étudiants ne possèdent pas forcément une certaine réflexivité linguistique qui permet de se rendre compte de ses vraies compétences ? Il est important de retenir que, pour ce public, la langue française représente en général un outil de communication. Ils l'apprennent pour pouvoir communiquer, et non pour connaître ses particularités linguistiques. Ainsi, lorsqu'ils appréhendent cette langue, il est possible qu'ils ne soient pas toujours attentifs à ce qu'ils sont capables de faire. Cela expliquerait l'écart entre le niveau de l'auto-évaluation et celui qu'ils possèdent en réalité.

Malheureusement, je n'ai pas pu me pencher davantage sur ces questions faute de temps pour revoir l'auto-évaluation et son utilisation. Cela a toutefois éveillé une envie en moi de réexaminer le concept de l'auto-évaluation lors de ma carrière d'enseignant de FLE.

³⁹ Dans ces deux phases d'expérimentation, un quart des étudiants auraient reçu le mauvais test.

Conclusion

En étudiant les évaluations en langues, nous constatons continuellement des évolutions. Des tests autocorrectifs, la volonté de réduire la subjectivité du jugement, et la diversité croissante des types d'épreuves en langue, sont juste quelques exemples. Ces constats nous amènent à nous interroger sur les éventuels changements dans ce domaine. Par exemple, l'évaluation de l'expression orale à distance par le biais de l'ordinateur apparaît comme l'une des évolutions possibles. Cette pratique est d'ailleurs déjà utilisée aux Etats-Unis par le ACTFL (American Counsel on the Teaching of Foreign Languages), l'équivalent américain du CECRL. Leurs tests pour déterminer le niveau de langue comprennent effectivement une partie d'expression orale qui est ensuite évaluée à distance par un examinateur du ACTFL. Cette pratique permet ainsi de mesurer toutes les capacités des utilisateurs de la langue qui ne passent pas l'examen sur place, s'adaptant à un monde transactif basé de plus en plus sur la technologie. Cette évolution plutôt récente provoque une réflexion sur le rôle de la technologie dans, non seulement l'évaluation en langues, mais également dans l'enseignement de ces dernières. Allons-nous continuer à nous éloigner d'une situation d'apprentissage et d'enseignement traditionnelle, c'est-à-dire en présentiel ?

Quant à la mission du stage, le nouveau dispositif hybride de positionnement comprend trois outils évaluatifs variés. Premièrement, le test autocorrectif en ligne fait l'objet d'un examen analytique, indirect et entièrement objectif qui évalue les connaissances des étudiants à travers des items fermés, ou à réponse courte. Ceci est une évaluation typique de la période psychométrique structuraliste. Deuxièmement, la fiche pour le test oral permet de mesurer la compétence d'interaction orale des étudiants à travers un outil évaluatif holistique et critérié. Ainsi s'agit-il d'un jugement guidé de ces compétences mais tout de même subjectif, typique des évaluations de la période actuelle. Enfin, l'auto-évaluation permet de prendre en compte la perspective de l'étudiant vis-à-vis de son niveau, ce qui fournit à l'enseignant une information utile ou tout simplement intéressante pour déterminer le niveau des étudiants de GEM.

Il est vrai que ces trois composantes du dispositif ont été finalisées. Cependant, il est important de souligner les limites de cette mission pédagogique car il ne s'agit pas d'un travail parfait. Certaines restrictions ont malheureusement empêché de mener d'autres actions afin de s'assurer davantage de la qualité du dispositif. De fait, le temps alloué pour réaliser ce projet, à savoir quatre mois et demi, représente une durée plutôt courte pour une telle conception pédagogique.

Par exemple, il n'a pas été possible de mettre en place une phase d'expérimentation de la fiche pour le test oral à cause des contraintes temporelles. De même, il aurait été préférable d'effectuer une deuxième phase d'expérimentation du test autocorrectif pour vérifier à nouveau sa fiabilité. Par ailleurs, une étude davantage détaillée du rôle de l'auto-évaluation dans le processus de

positionnement aurait été également intéressante et enrichissante. Ainsi, il est dorénavant évident que le temps alloué à une telle mission est un enjeu à considérer soigneusement avant d'entamer le projet.

Il était en outre regrettable de ne pas avoir eu plus de participation des étudiants des niveaux B1 et B2 dans la phase d'expérimentation du test autocorrectif. Il s'agit malheureusement d'un facteur difficilement contrôlable lors d'un recueil de données qui nécessite un volontariat de la part des participants. Un nombre plus élevé de ces étudiants aurait permis d'analyser davantage la pertinence des items dans les parties B1 et B2 du test, garantissant ainsi sa validité de manière plus forte.

Ce projet d'ingénierie politique et pédagogique répond tout d'abord au besoin de l'école de réduire le coût du positionnement des étudiants. De plus, étant donné la diversité et la validité des outils qui constituent le dispositif hybride, le programme de FLE à GEM peut désormais cerner avec précision le niveau des étudiants étrangers venant étudier le français dans cette école. Les contenus linguistiques des trois parties étant tous établis à partir d'une seule référence de langue basée sur le CECRL, la validité du dispositif est assurée. De plus, grâce à la récolte de données par le biais des questionnaires, il a été possible de constater les particularités du profil du public de GEM, et d'adapter ensuite les trois composantes du processus de positionnement à ses caractéristiques.

En définitive, nous constatons qu'un processus de positionnement va au-delà de la simple détermination du niveau des apprenants. L'enjeu est en effet beaucoup plus grand. Il s'agit non seulement de créer des groupes de niveau homogène, et un environnement propice à l'apprentissage de la langue française, mais aussi de s'assurer de l'efficacité du fonctionnement de l'institution. Un dispositif de positionnement valide, fiable, et adapté au contexte, aux apprenants et aux enseignants, est effectivement l'outil pédagogique qui peut répondre à cet enjeu.

Bibliographie

- ALTE (2011). *Manuel pour l'élaboration et la passation de tests et d'examens de langue*. Strasbourg : Conseil de l'Europe.
- l'Alliance Française de Paris-Ile-de-France (2008). *Référentiel de programmes pour l'Alliance Française*. Paris : Clé international.
- Beacco J-C. & al., (2007). *Niveau A1 pour le français - un référentiel*. Paris : Didier.
- Beacco J-C. & al., (2008). *Niveau A2 pour le français – un référentiel*. Paris : Didier.
- Beacco J-C. & al., (2011). *Niveau B1 pour le français – un référentiel*. Paris : Didier.
- Bonboir, A. (1972). *La docimologie*. Paris : Presses universitaires de France.
- Bento, M. (2012). L'évaluation des langues dans une perspective actionnelle. *Les Langues*
- Conseil de l'Europe. (2001). *Cadre européen commun de référence*. Paris : Les Éditions Didier.
- Doucet, P. (2001). Pour un test utile. *ASp*, 34, 13-33. Repéré le 11/12/2014 à <http://asp.revues.org/1696>
- Ebel, R. L. (1972). *Essentials of Education Measurement*. Englewood Cliffs, NJ : Pertinence Hall.
- Hadji, C. (1997). *L'évaluation démystifiée*. ESF éditeur : Paris.
- Horner, C. (2012). L'évaluation de la performance : comment être objectif ? *Les cahiers de l'asdifl : évaluation et certifications*, 23. Didier : Paris.
- Huver, H. Springer, C. (2011). *L'évaluation en langues*. Paris : Les Éditions Didier.
- Lussier, D. (1992). *Évaluer les apprentissages dans une approche communicative*. Paris : Hachette.
- Lussier, D. Turner, C. (1995). *Le point sur l'évaluation en didactique des langues*. Centre Éducatif et Culturel inc. : Anjou, Québec.
- McNamara, T. (2000). *Language Testing*. Oxford : Oxford University Press.
- Mothe, J.D. (1975). *L'évaluation par les tests dans la classe de français*. Paris : Hachette.
- Nadeau, M. (1988). *L'évaluation de programmes*. Québec : Les presses de l'Université Laval.
- Noël-Jothy, F. Sampsonis, B. (2006). *Certifications et outils d'évaluation en FLE*. Hachette : Paris.
- Tagliante, C. (1998). *Tests de Niveau*. Paris : Clé international.
- Tagliante, C. (1991). *L'évaluation*. Paris : Clé international.
- Tagliante, C. (2005). *L'évaluation et le cadre européen commun*. Paris : Clé international.

Sterne, H. (1983). *Fundamental Concepts of Language Teaching*. Oxford : Oxford University Press.

Veltcheff, C. Hilton, S. (2003). *L'évaluation en FLE*. Hachette : Paris.

Modernes, N. 106. APLV : Paris.

Sitographie

Clé International (s.d.). Repéré le 19/1/2015 à http://www.cle.fr/centre_linguistique-fr-idm-105-n-Exercices-idh-102.html (dernière consultation : 27/1/2015).

L'Alliance Française de Bruxelles (s.d.). Repéré le 19/1/2015 à <http://placement.alliancefr.be/index2.html>

ALTISSIA (s.d.). Repéré le 19/1/2015 à <http://www.altissia.com/fr/lp/test-de-niveau-francais-en-ligne-altissia>

Lardez, S (2014). Un campus à Paris pour Grenoble École de Management. Repéré le 18 février 2013 à <http://www.ecoles2commerce.com/1243-un-campus-a-paris-pour-grenoble-ecole-de-management>

Grenoble École de Mangagement (2015). Site web de GEM. Repéré le 18 février, 2013 à <http://www.grenoble-em.com/classements>

Le Figaro Étudiant (s.d.). Site web du journal Figaro. Repéré le 2/5/2015 à <http://etudiant.lefigaro.fr/les-news/actu/detail/article/220-millions-de-personnes-parlent-francais-hors-de-france-1468/>

Table des matières

Remerciements	3
Sommaire	5
Introduction	6
PARTIE 1. PRÉSENTATION DU CONTEXTE	9
1. L'organisme d'accueil	9
1.1. Histoire et philosophie	9
1.2. Organisation	9
2. Les cours de langue, et spécifiquement de FLE, à GEM	10
3. Les acteurs	11
3.1. Le personnel	11
3.2. Le public d'apprenants	11
4. Conclusions sur le contexte et le public	14
5. La commande	15
5.1. L'origine de la commande	15
5.2. La première formulation de la commande	16
5.3. La commande finale	17
PARTIE 2 : L'ÉTAT DE L'ART.....	18
1. L'acte d'évaluer en langues : les destinataires et les finalités	18
1.1. « Évaluer » : une définition	18
1.2. Qui évalue qui et pourquoi	19
2. Les différents types d'évaluations et leurs finalités	20
2.1. Après l'apprentissage : l'évaluation sommative	20
2.2. Pendant l'apprentissage : l'évaluation formative	21
2.3. Avant l'apprentissage : l'évaluation prédictive	22
2.4. L'importance des objectifs d'apprentissage en évaluation	24
2.5. D'autres types d'évaluations	24
2.5.1. L'évaluation critériée normative	25
2.5.2. L'évaluation holistique et analytique	25
2.5.3. L'évaluation de la performance et des connaissances	25
2.5.4. L'évaluation du savoir et de la capacité	26
2.5.5. Maîtrise ou continuum	26
2.5.6. L'évaluation continue ou ponctuelle	26
2.5.7. L'évaluation formative ou sommative	26
2.5.8. L'évaluation directe et indirecte	27
2.5.9. L'évaluation objective et subjective	27
2.5.10. L'évaluation sur échelle ou liste de contrôle	27
2.5.11. L'évaluation sur une impression ou un jugement guidé	28
2.5.12. L'évaluation par série ou par catégorie	28
2.5.13. L'auto-évaluation	28
2.6. Une conclusion sur les différentes évaluations	29
3. Les quatre étapes du processus évaluatif et les enjeux associés	30
3.1. Étape 1 : l'intention de l'évaluation	30
3.1.1. La praticabilité	30
3.1.2. L'équité	31
3.2. Étape 2 : la mesure	32
3.2.1. La validité	32
3.3. Étape 3 : le jugement	33
3.3.1. L'objectivité et la subjectivité	34
3.3.2. La fiabilité	37
3.3.3. Les effets parasites de la fiabilité	38
3.4. Étape 4 : la décision	40

4.	Les items possibles dans un test autocorrectif	41
4.1.	Les QCM en détails et les enjeux liés	42
4.1.1.	Les QCM et la validité	42
4.1.2.	Les QCM, l'objectivité et la fiabilité	43
4.1.3.	Les QCM, le hasard et la divination	43
4.2.	Conclusions sur les QCM.....	45
5.	L'histoire de l'évaluation en langue : de 1950 à aujourd'hui.....	45
5.1.	1950 à 1970 : la période psychométrique structuraliste	46
5.2.	1970 à 1990 : la période communicative métacognitive	47
5.3.	La période actuelle : 1990 au présent.....	48
5.4.	Remarques sur l'histoire de l'évaluation en langues	50
PARTIE 3. LA MISE EN PLACE D'UN DISPOSITIF HYBRIDE DE POSITIONNEMENT À GEM		52
1.	L'élaboration du nouveau test de positionnement.....	52
1.1.	L'analyse d'un échantillon de tests de positionnement	53
1.2.	Questionnaire auprès des étudiants sur le test existant	57
1.3.	Questionnaire de satisfaction auprès des enseignants sur le test existant	59
1.4.	L'analyse du test existant.....	61
1.4.1.	L'analyse de la forme.....	61
1.4.2.	L'analyse du contenu linguistique	63
1.4.3.	L'analyse de quatre tests faits lors de la rentrée de janvier 2015	67
1.5.	La conception du nouveau test de positionnement	69
1.5.1.	Étape 1 : la décision de fournir un test ou un examen	70
1.5.2.	Étape 2 : la planification	70
1.5.3.	Étape 3 : la conception du nouveau test	71
1.5.3.1.	Les compétences évaluées et le contenu linguistique du nouveau test.....	73
1.5.3.2.	L'échelle de notation	76
1.5.4.	Étape 4 : l'expérimentation et l'analyse des résultats.....	77
1.5.4.1.	L'analyse de la capacité du nouveau test à déterminer un niveau de langue.....	78
1.5.4.2.	La vérification de la difficulté du nouveau test de positionnement	81
1.5.4.3.	Questionnaire sur le nouveau test auprès des étudiants.....	82
1.5.5.	Étape 5 : Spécifications du test final	84
2.	La conception de la fiche pour le test oral	85
2.1.	L'analyse de la fiche existante pour le test oral.....	86
2.2.	L'élaboration de la nouvelle fiche pour le test oral	87
2.3.	L'élaboration du guide pour les enseignants	89
2.4.	Conclusion sur la conception de la nouvelle fiche pour le test oral.....	90
3.	La conception de l'auto-évaluation.....	90
Conclusion.....		93
Bibliographie.....		95
Sitographie		97
Table des matières		98

MOTS-CLÉS : Évaluation, test de positionnement, FLE, école de management, dispositif hybride de positionnement, niveau, validité, fiabilité

RÉSUMÉ

Ce mémoire de fin de master 2 professionnel, spécialité Française Langue Étrangère, a pour ambition d'examiner la notion de l'évaluation en langue dans le contexte de Grenoble École de Management. Après une présentation du contexte dans lequel s'insère ce stage, et qui permet de comprendre les particularités du terrain, suivra une explication des théories pertinentes à l'évaluation en langue afin de clarifier les concepts qui constituent ce champ d'études. Enfin, nous évoquerons l'élaboration d'un dispositif hybride de positionnement pour Grenoble École de Management.

La question qui sous-tend ce travail est la suivante : comment élaborer un dispositif de positionnement hybride en français pour des étudiants étrangers inscrits en école de management? Cette étude s'efforce de répondre à cette question en prêtant une attention aux diverses techniques et processus évaluatifs tels qu'ils sont décrits dans le CECRL. De plus, en prenant en compte les particularités du public présent à GEM, ce mémoire vise à rendre compte de la nécessité de connaître les destinataires d'un produit pédagogique avant et pendant sa conception. Enfin, les notions clés de ce domaine, telles que la validité et la fiabilité, ainsi que leurs rôles dans l'élaboration d'une épreuve en langue sont également une partie importante de ce travail.

KEYWORDS : language testing, assessment, level, positioning test, hybrid positioning process, validity, reliability

ABSTRACT

This end of Masters study, specialty French as a Foreign Language, examines the notion of language testing in the context of a management school, in this case Grenoble School of Management. Following a presentation of the institution in which this internship was carried out to bring to light the specificities of this school, an explication of relevant theories to language assessment will follow in order to define and clarify this field. The creation of a placement process, which determines the level in French of foreign students studying at Grenoble Business School, will then be described.

The guiding question for this project is the following: how does one create a hybrid positioning process in French for foreign students enrolled in a management school? This work strives to respond to all aspects of this question by attentively examining the different evaluation techniques and procedures as they are described in the CECRL. In addition, by taking into account the unique characteristics of the students at Grenoble School of Management, it aims to underline the importance of being familiarized with the subjects of such a testing process before and during its creation. Moreover, this paper pays particular attention to the notions of the validity and reliability of language evaluation.