



**HAL**  
open science

# La subjectivité de l'évaluation en production orale : le cas des langues vivantes

Chloé Pedelabat-Lartigau

► **To cite this version:**

Chloé Pedelabat-Lartigau. La subjectivité de l'évaluation en production orale : le cas des langues vivantes. Education. 2018. dumas-02090955

**HAL Id: dumas-02090955**

**<https://dumas.ccsd.cnrs.fr/dumas-02090955>**

Submitted on 21 Sep 2020

**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.

## **Année universitaire 2017-2018**

***Master Métiers de l'enseignement, de l'éducation et de la formation***

***Mention Second degré***

***Parcours : Anglais***

# **La subjectivité de l'évaluation en production orale**

**Le cas des langues vivantes**

**Présenté par Chloé PEDELABAT-LARTIGAU**

**Mémoire de M2 encadré par Alain GIRAULT**

*Merci à M. Girault pour son soutien, ses conseils précieux et son aide tout au long de mon parcours de formation au métier d'enseignant.*

*Merci tout particulièrement à Florian, sans qui ce mémoire n'aurait sans doute jamais vu le jour.*

*Merci à Marine, Carolyn et Pauline pour leur disponibilité et leur participation aux expérimentations. Merci encore à Marine qui a été une tutrice remarquable pendant cette année de stagiairisation*

*Merci à François-Xavier et à tous mes amis qui m'ont toujours soutenu.*

*Merci à mes parents.*

*Merci à mes élèves.*

# SOMMAIRE

<b>INTRODUCTION</b> .....	1
<b>ÉTAT DES LIEUX</b> .....	3
1. Évaluer en langues vivantes .....	3
1.1. Bref historique de l'évaluation en langues vivantes .....	3
1.1.1. L'évaluation dans les années 1970 et 1980 .....	3
1.1.2. 1992-2001 : l'évaluation diagnostique en début de seconde .....	4
1.1.3. La banque d'outils d'aide à l'évaluation diagnostique .....	4
1.2. Le Cadre européen commun de référence pour les langues .....	5
1.2.1. Le CECRL, un nouveau cadre en matière d'évaluation en langues vivantes .....	5
1.2.2. Chapitre 9 : évaluer .....	6
2. L'impact de la subjectivité dans l'évaluation .....	8
2.1. Mythe de la « vraie note » et les parasites de l'évaluation .....	8
2.1.1. La docimologie et le mythe de la « vraie note » .....	8
2.1.2. Les « parasites de l'évaluation » .....	9
2.1.3. La « constante macabre » .....	11
2.2. Remèdes à la subjectivité : objectifs opérationnels, critères et grilles d'évaluation .....	12
2.2.1. La relation évaluation / objectifs .....	12
2.2.2. Les critères et grilles d'évaluation .....	13
3. Parler, c'est dire (mots) et faire (gestes) .....	15
3.1. L'évaluation du langage verbal en production orale .....	16
3.1.1. Les compétences mobilisées dans l'activité langagière de production .....	16
3.1.2. Compétences et échelles du CECRL .....	17
3.2. Le langage non verbal .....	19
3.2.1. L'importance de la communication non verbale .....	19
3.2.1. Le langage non verbal dans l'évaluation .....	19
<b>PROBLÉMATIQUE</b> .....	22
<b>MÉTHODE</b> .....	23
1. Participants .....	23
2. Matériel .....	23
3. Procédure .....	25
4. Analyse des résultats .....	26
<b>DISCUSSION et CONCLUSION</b> .....	32
1. Une subjectivité à toute épreuve ? .....	32
2. L'impact du langage non verbal sur l'évaluation .....	34
3. Retour critique .....	36

<b>BIBLIOGRAPHIE</b> .....	38
<b>ANNEXES</b> .....	40

# INTRODUCTION

L'évaluation constitue un moment important dans le cursus scolaire d'un élève. Quelle que soit sa forme, elle suscite des difficultés singulières pour l'évaluateur. Une question légitime et à la fois complexe émerge immédiatement : sur quelle base doit-on évaluer les élèves ? Et comment les évaluer de la manière la plus juste possible ?

La question de la subjectivité semble constituer, ici, un enjeu important. En effet, l'objectivité de l'évaluation apparaît généralement comme un gage du traitement équitable des élèves, par la mesure réelle de leur savoir ainsi que de leur compétence. Mais est-elle seulement possible ? En matière d'évaluation de la production écrite, on compte plusieurs procédés dont l'objectif est d'éviter que la subjectivité du correcteur n'affecte la mesure réelle des acquis et des difficultés de l'élève. On peut notamment citer les doubles corrections, ainsi que l'anonymisation des copies. Mais qu'en est-il du cas particulier de l'évaluation en production orale ?

Certains outils visent bel et bien à réduire la subjectivité de l'évaluateur en matière de production orale. Il s'agira, dans ce mémoire, de revenir sur ces outils mis à disposition des évaluateurs et de questionner leur pertinence au regard des singularités de la production orale. Dans quelle mesure les outils mobilisés permettent-ils effectivement de rationaliser le processus d'évaluation ? On s'intéressera particulièrement, en la matière, à la dimension non verbale du discours. Celle-ci est-elle véritablement pensée en matière d'évaluation de l'oral, et dans quelle mesure peut-elle affecter l'évaluation de l'élève ?

Pour aborder ce problème, le cas des langues vivantes sera étudié. En effet, la question de la place de la subjectivité dans l'évaluation de la production orale est particulièrement sensible dans le cas des enseignements de langues vivantes. D'abord, l'oral y occupe une place de plus en plus importante et, à ce titre, la question de son évaluation se pose avec d'autant plus de force. Ensuite, le Cadre européen de référence pour les langues constitue un document officiel de référence dans l'enseignement des langues vivantes en France, et il fournit des outils et des préconisations précises qui visent à fournir un cadre commun pour tous les enseignants de langues vivantes, notamment en matière d'évaluation.

Nous proposons donc, dans ce mémoire, de nous interroger sur le problème de la subjectivité de l'évaluation dans le cadre de la production orale en langues vivantes. Pour ce faire, nous reviendrons dans un premier temps sur les évolutions récentes en matière d'évaluation en langues vivantes. Nous nous intéresserons ensuite à l'impact de la subjectivité

sur les apprenants et aux principes qui guident la réduction de cette subjectivité en matière d'évaluation. Enfin, nous mettrons à jour certaines spécificités propres à l'oral en langues vivantes, et les difficultés qu'elles sont susceptibles d'entraîner.

L'analyse des résultats obtenus lors des expérimentations nous offrira un recul critique sur nos hypothèses de départ, et précisera si nous pourrions, a minima, réduire la subjectivité par l'adoption de critères supplémentaires, lesquels permettraient une approche plus globale de la situation d'évaluation orale.

# ÉTAT DES LIEUX

Pour comprendre l'évaluation de la production orale et saisir la question de la subjectivité qui en découle, il convient tout d'abord de faire un bref rappel de l'évolution de l'évaluation en langues vivantes. Il est nécessaire, ensuite, de se pencher sur le caractère subjectif de l'évaluation, avant de s'intéresser plus particulièrement à ce problème en matière de production orale.

## **1. Évaluer en langues vivantes**

La question de l'évaluation est devenue omniprésente dans notre société. Celle-ci n'épargne aucun secteur (Hadji, 2012, p. 9). Nous proposons ici de nous intéresser à l'évolution de l'évaluation en langues vivantes, afin de mieux comprendre la transformation du système évaluatif, notamment avec l'introduction du Cadre européen de référence pour les langues (CECRL), publié en 2001 par le Conseil de l'Europe.

### **1.1. Bref historique de l'évaluation en langues vivantes**

Le problème de l'évaluation en langues vivantes a depuis longtemps suscité des débats et orienté les politiques linguistiques d'enseignement des langues. La chronologie de l'évaluation que nous proposons ci-dessous mettra l'accent sur les évolutions les plus récentes, partant des années 1970 pour arriver aux années 2000. Il ne s'agit donc pas d'un historique exhaustif mais simplement d'une aide au lecteur pour mieux contextualiser l'évaluation en langues vivantes aujourd'hui.

#### **1.1.1. L'évaluation dans les années 1970 et 1980**

Pendant longtemps, l'évaluation en langue vivante n'était que sommative. Comme l'indique le rapport de l'inspection générale de l'Éducation nationale de 2007, l'évaluation a longtemps eu pour fonction d'évaluer de manière globale les connaissances acquises par l'élève, sans distinguer chez lui sa maîtrise des différentes activités langagières, alors appelées compétences ou savoir-faire. Les activités langagières telles qu'elles sont définies aujourd'hui correspondent aux modalités de traitement des messages oraux ou écrits, en réception ou en production. Nous différencions à présent cinq activités langagières : la réception ou compréhension orale qui consiste à écouter et comprendre un message audio ; la réception écrite qui permet de lire et comprendre un texte ; l'expression ou production orale qui désigne l'action de parler en interaction ou en continu ; l'expression écrite qui correspond à l'écriture en interaction ou en continu ; et à cela s'ajoute la médiation orale ou écrite qui permet de



traduire et d'interpréter des locuteurs de deux langues différentes qui ne peuvent communiquer directement (Conseil de l'Europe, 2001, p. 40). Nous soulignons que dans le cadre de notre mémoire, nous nous intéresserons plus particulièrement à la question de la subjectivité de l'évaluation en production orale.

L'évaluation orale a ainsi longtemps été fondue dans une logique de mesure du niveau général des élèves. Les notes obtenues servaient d'abord à un classement de ceux-ci :

« On pratiquait une évaluation orale en début de séance et on donnait un devoir écrit périodique ; on évaluait les savoirs plutôt que les savoir-faire et la différence entre exercice d'entraînement et exercice d'évaluation n'était guère marquée. La note unique, globale, servait surtout à classer les élèves ; elle ne permettait pas de rendre compte, de façon fine et fiable, de leur niveau dans les différentes activités langagières. » (Inspection générale de l'Éducation nationale, 2007, p. 7)

Ce n'est que récemment, au cours des années 1980, que cette logique fut infléchie. En 1983 notamment, la réforme du baccalauréat introduisit un modèle d'évaluations dissociées. Ce modèle engageait les enseignants à évaluer trois domaines de façon distincte : la compréhension, l'expression écrite et la compétence linguistique. Cependant, la note unique ne fut pas abandonnée, car on ne prenait toujours en compte que la moyenne des trois notes obtenues. Il faudra attendre les années 1990 pour observer un modèle d'évaluation qui rompe davantage avec la logique de classement, même si celui-ci perdure.

### **1.1.2. 1992-2001 : l'évaluation diagnostique en début de seconde**

En 1992, l'introduction d'une évaluation nationale diagnostique à l'entrée en seconde générale et technologique et en seconde professionnelle devait constituer un « tournant décisif dans les pratiques évaluatives des enseignants en langues » (Inspection générale de l'Éducation nationale, 2007, p. 8). En effet, l'objectif revendiqué par ce modèle d'évaluation était de permettre une réponse adaptée aux besoins des élèves. Notamment, l'introduction de cahiers d'évaluation devait remettre en cause, selon l'Inspection générale de l'Éducation nationale, les traditions évaluatives qui jusqu'ici avaient cours (Inspection générale de l'Éducation nationale, 2007). Cette logique visait donc moins à sanctionner un niveau général de l'élève qu'à apprécier ses compétences et savoir-faire en début d'année scolaire, afin d'anticiper certaines de ses difficultés et lui proposer des réponses adéquates. Il s'agissait de promouvoir un usage de l'évaluation marqué davantage par une posture d'accompagnement.

### **1.1.3. La banque d'outils d'aide à l'évaluation diagnostique**

Au tout début des années 2000, l'évaluation diagnostique nationale en début de seconde disparaît, et une banque de ressources au service de la pédagogie commence à être mise à

disposition des professeurs d'école, de collège et de lycée. Ce changement devait élargir la logique de l'évaluation diagnostique tout au long du cursus scolaire. En effet, ces nouveaux outils devaient permettre aux enseignants de suivre au plus près l'évolution des élèves et cibler les difficultés que ceux-ci étaient susceptibles de rencontrer au cours de l'année scolaire :

« La « banque d'outils d'aide à l'évaluation diagnostique » [...] permet d'évaluer les compétences des élèves en classe, à tout moment de l'année scolaire [...]. La banque d'outils permet aux enseignants d'effectuer, en cours d'apprentissage, un diagnostic des compétences de leurs élèves et des difficultés particulières qu'ils rencontrent dans tel ou tel domaine. » (Inspection générale de l'Éducation nationale, 2007, p. 8)

La Direction de l'évaluation et de la prospective aurait souhaité que ces outils bousculent « quelque peu les habitudes pédagogiques » et suscitent un changement « dans la pratique des enseignants ». (Inspection générale de l'Éducation nationale, 2007, p. 9). Toutefois, en 2007, la banque d'outils restait encore insuffisamment connue des professeurs et n'était que très peu utilisée.

On observe donc une évolution du modèle d'évaluation en langues vivantes. Celui-ci tend à mieux distinguer certains domaines de compétences et à en préciser les modalités d'évaluation. Mais ces changements intervenus en France depuis les années 1980 s'inscrivent surtout dans une dynamique récente d'harmonisation des exigences au niveau européen.

## **1.2. Le Cadre européen commun de référence pour les langues**

Depuis le décret n° 2005-1011 du 22 août 2005, les programmes français d'enseignement des langues vivantes s'appuient sur le Cadre européen commun de référence pour les langues. Le ministère français de l'Éducation nationale en fait état sur son site internet, notant que « l'enseignement des langues vivantes étrangères s'inscrit dans le cadre européen commun de référence pour les langues (CECRL), qui fournit une base commune aux États membres de l'Union européenne pour concevoir les programmes »<sup>1</sup>. Ce Cadre commun constitue donc un document privilégié en matière d'évaluation en langue vivante, et propose des remèdes à la subjectivité de l'évaluation.

### **1.2.1. Le CECRL, un nouveau cadre en matière d'évaluation en langues vivantes**

Le Cadre européen commun de référence pour les langues (Conseil de l'Europe, 2001), publié en 2001 par le Conseil de l'Europe, est l'aboutissement d'un travail collaboratif entre

---

<sup>1</sup> Consulté le 23/02/2018 sur <http://eduscol.education.fr/cid45718/les-langues-vivantes-etrangees-ecole.html>

membres de la profession enseignante en Europe depuis le début des années 90. Ses finalités sont triples : apprendre, enseigner et évaluer. En plus d'uniformiser les attendus en langue au niveau européen, deux objectifs principaux gouvernent la mise en place de cet outil :

« 1. Encourager les praticiens dans le domaine des langues vivantes [...], y compris les apprenants, à se poser un certain nombre de questions.

2. Faciliter les échanges d'informations entre les praticiens et les apprenants. » (Conseil de l'Europe, 2001, p. 4)

Le but revendiqué par le Cadre européen est de donner des pistes de réflexion. Il ne s'agit donc pas, initialement d'un document qui se voudrait exhaustif, ni d'un document ayant pour but de prescrire certaines méthodes. Toutefois, il faut noter que le CECRL n'a pas seulement valeur de recommandation. En effet, depuis son intégration aux programmes de langues vivantes en 2005, le CECRL a une force contraignante en France<sup>2</sup>. A ce titre, le chapitre 9 du CECRL, consacré aux modes d'évaluation, présente un intérêt particulier pour notre objet d'étude. En effet, il expose « les diverses finalités de l'évaluation et les types d'évaluation qui y correspondent à la lumière de la nécessité de réconcilier les critères concurrents d'exhaustivité, de précision et de possibilité opératoire. » (Conseil de l'Europe, 2001, p. 8). Six niveaux de compétences sont préconisés et permettent d'harmoniser, pour toutes les langues, les pratiques dans les pays membres du Conseil de l'Europe. L'objectif est ici de parvenir à une meilleure cohérence et une meilleure transparence en matière d'évaluation et d'enseignement des langues entre les différents États membres, mais également entre les différents acteurs, enseignants et apprenants (Tagliante, 2001, p. 35). Or, cette transparence constitue un aspect essentiel à prendre en compte lorsque l'on traite de la subjectivité dans l'évaluation. Le Cadre européen commun propose, à cet effet, certaines préconisations.

### **1.2.2. Chapitre 9 : évaluer**

Mettant en regard différents types d'évaluation, le CECRL pose clairement le problème de la subjectivité dans l'évaluation et tente de lui apporter des réponses structurées. Tout d'abord, il oppose l'évaluation subjective à l'évaluation objective. La première se définit comme le « jugement [d'examinateur] sur la qualité de la performance » (Conseil de l'Europe, 2001, p. 142). La seconde, l'évaluation objective, est définie comme l'opposé du jugement propre à l'examinateur. Selon les auteurs, elle vise en effet à écarter la part de subjectivité de l'évaluateur généralement par « l'utilisation d'un test indirect dans lequel une seule réponse est possible, par exemple un QCM » (Conseil de l'Europe, 2001, p. 142).

---

<sup>2</sup> Décret n° 2005-1011 du 22 août 2005 – JO du 25 août 2005, BOEN n° 31 du 1<sup>er</sup> septembre 2005

Reconnaissant la complexité de cette opposition subjectivité/objectivité, les auteurs du CECRL proposent une démarche permettant de réduire cette part de subjectivité, en augmentant ainsi la validité et fiabilité de l'évaluation. Ils recommandent notamment de :

- « développer une **spécification du contenu** de l'évaluation fondée, par exemple, sur un cadre de référence commun au contexte en question
- utiliser la négociation ou les **jugements collectifs** pour sélectionner le contenu et/ou noter les performances
- adopter des **procédures normalisées** relatives à la passation des tests
- fournir des **grilles des corrections précises** pour les tests indirects et baser l'évaluation des tests directs sur des critères de correction clairement définis
- exiger des **jugements multiples** et/ou l'**analyse de différents facteurs**
- mettre en place une **double correction** ou une **correction automatique** lorsque c'est possible
- assurer une **formation** relative aux **barèmes de correction**
- vérifier la qualité de l'évaluation (validité, fiabilité) par l'**analyse des résultats**. » (Conseil de l'Europe, 2001, p. 143)

Ces recommandations s'appuient ainsi sur l'explicitation des critères mobilisés pour toute évaluation et l'utilisation de références collectives, sinon standardisées, en matière de sélection des contenus et de procédures d'évaluation. Les mêmes principes guident les recommandations du CECRL concernant la subjectivité du jugement. En effet, les auteurs opposent le jugement fondé sur l'impression, défini comme un « jugement entièrement subjectif fondé sur l'observation de la performance de l'apprenant en classe, sans aucune référence à des critères particuliers relatifs à une évaluation spécifique » (Conseil de l'Europe, 2001, p. 143), au jugement dit « guidé ». Le second est un « jugement dans lequel on réduit la subjectivité propre à l'examineur en ajoutant à la simple impression une évaluation consciente en relation à des critères spécifiques » (Conseil de l'Europe, 2001, p. 143).

\*

Le Cadre européen souligne ainsi la nécessité de recourir à des critères précis et communs. Ce document tend à cadrer les nouvelles pratiques en matière d'évaluation et il encourage la mise en place de critères et de pratiques communs au niveau européen. Nous reviendrons de façon plus approfondie sur cette question qui a trait aux critères d'évaluation, afin d'en saisir les enjeux en matière de subjectivité. Avant cela, il s'agirait de s'intéresser de façon plus précise à l'impact de la subjectivité et de comprendre les difficultés qui lui sont liées.

## **2. L'impact de la subjectivité dans l'évaluation**

L'évaluation selon une normativité humaine n'échappe pas à la question de la subjectivité. Nous nous interrogerons d'abord sur l'impact de cette subjectivité dans l'évaluation, notamment au travers des parasites de l'évaluation et de la constante macabre. Puis nous tenterons de présenter des remèdes à cette subjectivité.

### **2.1. Mythe de la « vraie note » et les parasites de l'évaluation**

« Pourquoi les notes diffèrent-elles d'un correcteur à l'autre ? Pourquoi un même travail peut-il être évalué différemment, à quelques mois ou semaines d'intervalle, par un même correcteur ? » (Tagliante, 2005, p. 11). Les questions soulevées par Tagliante sont autant de questions que se posent les enseignants lorsqu'ils se retrouvent face à leur devoir d'évaluateur. Or, certains biais d'évaluation sont déjà bien connus de la littérature. Sans prétendre à l'exhaustivité, il s'agit ici de synthétiser quelques-uns de ces mécanismes. Nous soulignons que les études que nous mentionnerons dans cette partie portent essentiellement sur l'évaluation écrite, mais les principes qui les fondent peuvent, malgré certaines spécificités, s'appliquer au champ de l'évaluation orale.

#### **2.1.1. La docimologie et le mythe de la « vraie note »**

Dès le début du XX<sup>ème</sup> siècle, les travaux de Piéron et d'autres chercheurs ont donné naissance à une nouvelle discipline, la docimologie, en s'interrogeant sur « la fidélité des notes, en cas de répétition de l'examen ou de changement de l'examineur » (Piéron, 1969, p. 7). La docimologie correspond à la science de la mesure de l'apprentissage, de l'évaluation et des examens, elle fait l'étude critique des examens et des concours. Selon les chercheurs de cette discipline, les travaux des élèves pourraient être objectivement mesurables et quantifiables. Dès lors, il serait possible d'obtenir une « vraie » note, c'est-à-dire une note représentative de la véritable valeur du travail fourni :

« [...] les différences de notation pouvaient être éliminées en multipliant le nombre de correcteurs et en appliquant aux diverses notes obtenues une procédure statistique de mesure permettant de déterminer la « vraie note ». » (Tagliante, 2005, p. 11)

Il ressort ainsi des travaux menés en docimologie, l'idée selon laquelle il est possible de gommer entièrement la subjectivité de l'évaluation en ayant recours à plusieurs correcteurs. Cette « vraie note » paraît constituer un idéal bien ancré pour l'évaluateur, les enseignants aspirant généralement à appliquer une réelle objectivité dans l'évaluation des performances de leurs élèves. Toutefois, bien qu'intéressants et bien qu'ayant apportés des modifications

pédagogiques, les résultats des travaux menés en docimologie se sont avérés inapplicables dans leur ensemble, car trop coûteux.

Dans les années 1970, les résultats issus des travaux en docimologie furent remis en question. Hormis les épreuves présentées sous forme de questions à choix multiples (QCM), qui peuvent être mesurées de manière objective, pour Tagliante, la « vraie note » tient davantage du mythe que de la réalité :

« La notation [...], n'est pas un instrument de mesure exacte, elle ne contribue pas à la fidélité de l'évaluation. Elle ne peut être qu'une estimation donnée à une performance, en fonction de critères d'appréciation. » (Tagliante, 2005, p. 12).

Récemment, dans son ouvrage sur la notation des élèves, Dubus est revenu sur la question de la vraie note, qu'il définit comme « la valeur authentique et intrinsèque d'une copie, qui ne dépendrait plus du correcteur, et qu'on atteindrait au moyen d'une correction multiple » (Dubus, 2006, p. 271). L'auteur souligne que la mise en œuvre d'un tel niveau de correction multiple dans des situations réelles d'examens semble impossible. Il souligne notamment que, pour une copie corrigée par différents évaluateurs, « l'analyse statistique de la variation de l'ensemble des notes montre qu'elle est attribuable pour 40% à la copie et pour 60% au correcteur » (Dubus, 2006, p. 93). Barlow ajoute, à ce propos, que chaque professeur a « ses habitudes de notation, son équation personnelle » (Barlow, 2003, p. 33). Certains notent de façon « généreuse », d'autres de façon plus « sévère ». Certains utilisent « tout l'éventail » de notes compris entre 0 et 20, tandis que d'autres tendent à réduire ce panel de notes disponibles, de sorte qu'en fonction de son correcteur, la même note aura une signification toute différente. La question de la subjectivité dans l'évaluation paraît ainsi aussi centrale que complexe.

La docimologie aura permis de montrer que la notation de copies n'est jamais stable ni constante, ce qui signifie que la notation de copies ne peut jamais être fidèle. De nombreux facteurs connus sous le terme de « critères parasites de fidélité » (Tagliante, 2005, p. 12) ou « parasites de l'évaluation » (Riba et Mègre, 2014, p. 72) contribuent à la subjectivité dans la notation. Voyons plus précisément ces biais et les principes qui les sous-tendent.

### **2.1.2. Les « parasites de l'évaluation »**

Riba et Mègre précisent que les « parasites de l'évaluation » sont des « facteurs humains qui échappent à la standardisation des grilles d'évaluation et qui ont une influence directe sur l'attribution d'une note » (Riba et Mègre, 2014, pp. 72-73). Les enseignants connaissent souvent implicitement ces critères parasites de la fidélité, mais ne savent pas toujours les

nommer. Pourtant, il est important que les correcteurs en soient pleinement conscients pour essayer de les éliminer, ou au moins de les réduire. De nombreux chercheurs rapportent principalement onze de ces critères (Barlow, 2003, p. 34 ; de Peretti et al., 2005, pp. 50-52 ; Tagliante, 2005, pp. 12-13 ; Guimard, 2010, p. 44 ; Riba et Mègre, 2014, pp. 72-73 ;). On trouve d'abord des mécanismes qui ont trait à l'ordre des corrections des travaux en question. *L'effet d'ordre*, notamment, consiste à déprécier les notes attribuées aux élèves évalués en dernier, indépendamment de la qualité de leurs travaux. On trouve également, dans une veine un peu différente, *l'effet de contraste*, qui consiste à minorer la note d'une copie de niveau moyen lorsqu'elle est corrigée dans la suite immédiate d'une copie excellente : le contraste entre les deux copies tend alors à desservir la seconde copie, de moins bonne qualité que la première ou inversement, à sur-noter cette copie en fonction de la qualité de la copie précédente.

On trouve également des mécanismes liés à la perception que l'enseignant a de l'élève. *L'effet de favoritisme*, à ce titre, a trait aux affects de l'enseignant. Ce dernier trouvera rebutant d'attribuer une mauvaise note à certains élèves qu'il apprécie et qui lui renvoient une image valorisante de lui-même. Sur la base de critères qui n'ont pas trait aux compétences évaluées par le test, la note est ainsi rehaussée. De plus, on peut noter un *effet de stéréotype*, qui renvoie à la constitution d'un niveau de référence de l'élève lors des premières évaluations. L'enseignant tendra ici à évaluer les performances ultérieures de l'élève en référence à ce niveau initialement attribué. *L'effet de contamination*, encore, consiste à évaluer un élève en fonction du niveau généralement attribué à celui-ci. À copie égale, la note sera supérieure chez un élève dont le niveau est habituellement bon, et inférieure chez un élève dont le niveau est généralement faible.

On trouve encore des effets liés à certaines qualités ou défauts présents dans le travail fourni et qui affectent l'appréciation du correcteur. *L'effet halo*, notamment, consiste à rehausser la note attribuée à un élève en raison de la qualité de son expression et de sa présentation, qu'elle soit écrite ou orale. La qualité de la présentation affecte alors l'appréciation du contenu même : un contenu de qualité tendra à être déprécié lorsque la qualité de l'expression de l'élève, son écriture, etc. est mauvaise. De façon analogue, *l'effet choc* consiste à valoriser ou dévaloriser le travail évalué en raison de la qualité remarquable ou, au contraire, de la médiocrité exceptionnelle d'une part restreinte et non représentative du rendu final. *L'effet goutte d'eau*, encore, renvoie à la dévalorisation générale de certains travaux, sur la base d'une succession de petites erreurs qui, en s'accumulant au sein d'une copie ou d'une série de copies, finissent par susciter le désinvestissement de l'évaluateur.

Enfin, on peut noter des facteurs externes qui peuvent affecter l'évaluation. *L'effet de fatigue*, d'abord, constitue un facteur incompressible qui affecte les performances de l'évaluateur, celui-ci étant davantage en mesure d'évaluer équitablement lorsqu'il est pleinement disponible pour cet exercice complexe. *L'effet écart-type*, ensuite, consiste, chez un enseignant, à n'utiliser qu'un panel réduit de notes lors de la correction : par exemple, noter entre 4 et 16, alors qu'un collègue ne descendra que rarement en dessous de 6, etc. *L'effet évaluation externe*, enfin, est un mécanisme de pression professionnelle et institutionnelle implicite. L'enseignant tendrait ici à ne pas attribuer de notes trop basses, de peur d'entacher sa réputation vis-à-vis de ses collègues et de son institution : il pourrait passer pour un « mauvais enseignant ».

Ces critères parasites démontrent la nécessité pour les enseignants de fournir un effort constant pour atteindre plus d'objectivité. Effort d'autant plus important que l'évaluation est un acte qui peut avoir des conséquences sur l'avenir social de l'élève évalué. La responsabilité professionnelle de l'enseignant est donc ici engagée. Pour autant, ces biais sont difficiles à corriger, si bien que certains effets paraissent s'imposer aux professionnels, comme c'est le cas, notamment, de la « constante macabre ».

### **2.1.3. La « constante macabre »**

La « constante macabre » est un terme utilisé pour la première fois par Antiby, à la fin des années 1980. Selon ce chercheur en didactique, « sous la pression de la société, les enseignants se sentent obligés de mettre un certain nombre de mauvaises notes pour que leur évaluation soit crédible » (Antiby, 2007, p. 15). Selon cet auteur, la proportion de mauvaises notes dans la notation d'examens serait indépendante du sujet de l'examen, indépendante du correcteur et indépendante de la véritable qualité des réponses données par les élèves : la proportion de « mauvaises notes » serait toujours similaire. C'est ce pourcentage de mauvaises notes qui est appelé « constante macabre ». Il convient de préciser que dans l'ouvrage d'Antiby, le mot « évaluation » n'est utilisé que lorsque la note attribuée à l'élève est prise en compte dans sa scolarité et dans son orientation.

Tout comme les autres parasites de l'évaluation, il s'agit d'un phénomène inconscient qui a pour conséquence « l'échec artificiel injuste » (Antiby, 2007, p. 15 et p. 27) de nombreux élèves. En effet, ces derniers peuvent aléatoirement se retrouver dans le « mauvais » tiers de la classe et ne pas réussir à augmenter leurs résultats scolaires, malgré leur travail. Cela signifie donc que, dans une classe, les élèves « moins bons » sont condamnés à être en situation d'échec, malgré leurs efforts et malgré le fait qu'ils puissent avoir tout compris. La



difficulté réside dans le fait qu'il s'agisse d'un phénomène de société, qui ne dépend pas uniquement des enseignants et sur lequel ceux-ci n'ont pas prise. Or, cet échec n'est pas sans conséquence sur le parcours des élèves.

Selon Antibi, « tant que la constante macabre existera, notre système éducatif sera défaillant, paralysé » (Antibi, 2007, p. 15) et ses répercussions sont nombreuses. On peut citer notamment la détérioration du climat de confiance entre les professeurs et les élèves, la perte de confiance en soi des jeunes confiés à l'institution, le mal-être à l'école ou encore un processus d'orientation inadapté pour certains élèves dans un contexte où les notes jouent un rôle très important en France. Plusieurs chercheurs soulignent également que « l'évaluation peut générer une menace au niveau individuel, sous forme d'anxiété ou de crainte de l'échec, mais aussi au niveau relationnel, sous forme de tensions entre camarades ou avec les enseignants » (Butera et al., 2011, p. 11). Selon certains psychologues scolaires, on trouverait même une relation entre suicide et constante macabre (Antibi, 2007, p. 31). Dès la première moitié du vingtième siècle, Piéron soulignait déjà que le hasard jouait un rôle important dans les succès et les échecs et que cela restait accepté comme une norme à laquelle il serait difficile, voire impossible, de déroger (Piéron, 1969, p. 5). L'impact que peut avoir encore aujourd'hui la subjectivité de l'évaluation sur les apprenants paraît donc bien réel, et pour le moins préoccupant. On comprend, dès lors, la nécessité de trouver des solutions structurées pour répondre au problème de la subjectivité de la notation.

## **2.2. Remèdes à la subjectivité : objectifs opérationnels, critères et grilles d'évaluation**

Selon Tagliante, la fidélité de la notation n'est possible que si cette dernière est toujours stable et constante (Tagliante, 2005, p. 23). En la matière, si l'objectivité paraît difficile à atteindre, il est cependant possible de tendre vers la fidélité de la note en restreignant au maximum la subjectivité de l'évaluateur (Tagliante, 2005, p. 24). C'est ce que nous proposons d'aborder dans la prochaine sous-partie.

### **2.2.1. La relation évaluation / objectifs**

Comme le souligne Tagliante, la formulation des objectifs constitue un premier moyen pour encadrer le travail demandé par l'enseignant et organiser les critères d'évaluation. Elle permet d'optimiser l'enseignement et l'apprentissage, et donc l'évaluation des acquis, dans tous les domaines. Mager précise qu'un objectif « décrit le résultat que veut atteindre un cours plutôt que le processus d'enseignement ou les moyens mis en œuvre » (Mager, 2005, p. 7). Cela signifie qu'un objectif permet à l'enseignant de définir ce qu'il souhaite que ses élèves

soient capables de faire à la fin de son enseignement. Il convient de distinguer, en la matière, les objectifs généraux, les objectifs spécifiques et les objectifs opérationnels.

Les objectifs généraux, d'abord, sont centrés sur l'élève : il s'agit des résultats attendus à la fin d'une séquence. Par exemple, « Être capable de faire un discours, argumenter, utiliser un langage formel, convaincre. », sont des attentes générales directement portées sur l'élève. Les objectifs spécifiques sont également centrés sur l'élève et ont pour but d'affiner l'objectif général en le démultipliant en autant d'objectifs spécifiques à faire acquérir. Enfin, les objectifs opérationnels, ou « objectifs spécifiques opérationnels », sont ceux qu'il est indispensable d'acquérir pour accéder à l'objectif suivant. Ces objectifs ont la particularité d'être opérationnalisés, c'est-à-dire qu'ils sont formulés de façon à ce qu'on puisse les évaluer. Ainsi, un objectif opérationnel formulé correctement doit être directement évaluable.

Suite aux travaux de Mager notamment (Mager, 2005, p. 21), quatre qualités ont été fixées pour qu'un objectif soit opérationnel et évaluable : l'univocité, le comportement observable, les conditions et les critères (Tagliante, 2005, p. 22) :

« 1. Pour qu'une intention pédagogique tende à devenir opérationnelle, son contenu doit être énoncé de la manière la moins équivoque possible.

2. Pour qu'une intention pédagogique tende à devenir opérationnelle, elle doit décrire une activité de l'apprenant identifiable par un comportement observable.

3. Pour qu'une intention pédagogique tende à devenir opérationnelle, elle doit mentionner les conditions dans lesquelles le comportement escompté doit se manifester.

4. Pour qu'une intention pédagogique tende à devenir opérationnelle, elle doit indiquer le niveau d'exigence auquel l'apprentissage est tenu de se situer, et les critères qui serviront à l'évaluation de cet apprentissage. » (Hameline, 2005, p. 62)

Le nombre d'objectifs opérationnels doit être limité afin d'être représentatif de l'enseignement et correspondre à la séquence d'apprentissage (Tagliante, 2005, p. 23). L'évaluation dépend ainsi toujours des objectifs que l'enseignant a retenus ainsi que des critères qu'il a décidé d'appliquer. Voyons plus précisément l'importance de ces critères.

### **2.2.2. Les critères et grilles d'évaluation**

Selon Tagliante, l'enseignant dispose de deux outils essentiels pour restreindre au maximum la subjectivité de l'évaluateur : une grille comportant des critères d'appréciation, ainsi qu'un barème chiffré pour l'accompagner (Tagliante, 2005, p. 24). Comme le précise Dubus, « l'usage de barèmes pour la correction et la notation constitue un pas vers un minimum d'objectivité » (Dubus, 2006, p. 95). Toutefois, il ne peut garantir une rigueur

absolue, dès lors que l'activité fait appel au jugement humain. Pour permettre de dresser certains « garde-fous », l'auteur propose d'observer quelques règles de base.

La grille de correction doit être composée de critères à partir desquels la performance sera appréciée et évaluée. Cette grille de correction doit également permettre à différents évaluateurs de se référer aux mêmes critères et de différer la notation dans le temps. Nous rappelons que les critères d'évaluation sont toujours définis en fonction d'objectifs opérationnels préalablement sélectionnés par l'enseignant-évaluateur, objectifs qui doivent être indépendants. Pour avoir du sens, la notation doit donc être rapportée à des objectifs et des critères.

Il convient de faire quelques rappels sur certaines définitions avant de poursuivre notre réflexion. Un critère est un « caractère ou propriété d'après lequel/laquelle on porte une mesure ou une appréciation sur une performance. » (Tagliante, 2005, p. 189) Il ne faut pas confondre critère et descripteur. Le rapport de l'inspection de 2007 (Inspection générale de l'Éducation nationale, 2007, pp. 19-20) rappelle que les descripteurs globaux permettent de positionner l'élève sur un niveau du CECRL (de A1 à C1) mais ne peuvent pas tenir lieu de critères d'évaluation. En effet, les descripteurs ne suffisent pas pour apprécier la qualité des performances d'un élève dans les différentes activités langagières. Pour cela, il est nécessaire de sélectionner des critères pertinents à mettre dans une grille d'évaluation pour apprécier la qualité de la production ou compréhension d'un élève, à l'écrit comme à l'oral.

Dans son chapitre 5, le CECRL (Conseil de l'Europe, 2001, pp. 86-101) propose un inventaire des composantes de chacune des compétences (linguistique, sociolinguistique, pragmatique) que l'élève doit maîtriser à chaque niveau (de A1 à C1). Cet inventaire est accompagné de la formulation de critères qualitatifs : étendue linguistique générale, étendue du vocabulaire, maîtrise du vocabulaire, correction grammaticale, maîtrise du système phonologique, maîtrise de l'orthographe, correction sociolinguistique, souplesse, tours de parole, développement thématique, cohérence et cohésion, aisance à l'oral, précision. Si ces critères sont nombreux, les concepteurs du CECRL préconisent de faire des choix et éventuellement de hiérarchiser les critères, notamment en fonction de la finalité de l'apprentissage.

\*

Les remèdes à la subjectivité cités précédemment n'ont pas pour ambition de dresser une liste exhaustive de solutions. Nous tenons notamment à citer un autre aspect important à prendre en compte lorsqu'il s'agit d'évaluer un élève : la question de la posture de l'évaluateur. En effet, Baillat et al. Soulignent que « la fonction d'évaluateur appelle

l'adoption d'une posture particulière vis-à-vis de l'évalué. Il doit se mettre en position d'extériorité et d'objectivité vis-à-vis de l'autre » (Baillat et al., 2008, p. 61). La difficulté, pour l'évaluateur, consiste à garder une posture objectivante et à ne pas confondre jugement sur la performance et jugement sur autrui. Pour cela, il est recommandé de définir des objectifs clairs et précis à garder tout au long de la séquence d'apprentissage qui permettront ensuite de constituer une grille d'évaluation et un barème chiffré permettant d'accompagner l'évaluateur dans l'acte d'évaluation.

### **3. Parler, c'est dire (mots) et faire (gestes)**

Comme nous venons de le voir, l'évaluation en langues vivantes est un exercice complexe. L'introduction du CECRL constitue, à ce titre, une référence centrale qui fournit des principes communs en matière d'évaluation ainsi que des critères explicites. Ses préconisations touchent notamment aux questions de subjectivité et sont susceptibles de réduire le poids de celle-ci dans l'évaluation faite par les enseignants.

Dans ce cadre, le cas de l'évaluation orale est assez singulier. D'abord, la place de la production orale en matière d'évaluation semble, dans la pratique, mineure. En effet, selon le rapport de 2007 réalisé par l'Inspection générale de l'Éducation nationale, les compétences orales sont peu évaluées par les enseignants (Inspection générale de l'Éducation nationale, 2007), les notes d'oral ne représentant généralement que 25% de la note globale du trimestre. L'importance de l'écrit demeure ainsi centrale dans les pratiques d'évaluation, alors même que les professionnels se disent généralement prêts à faire évoluer ces pratiques. Pourtant, les préconisations du ministère de l'Éducation nationale insistent sur la place prédominante que devrait occuper l'oral. On a vu une évolution à ce titre avec la création en 2013 d'épreuves anticipées en langues vivantes pour le diplôme du baccalauréat. Depuis cinq ans, les enseignants doivent désormais évaluer la compréhension orale et la production orale de leurs élèves en cours d'année<sup>3</sup>. Ensuite, et surtout, si les préconisations relatives à l'évaluation de l'oral peuvent être communes à l'évaluation de la production écrite, celles-ci présentent toutefois certaines spécificités. Nous nous intéressons, dans cette dernière section, aux spécificités de l'évaluation orale et à ses difficultés.

---

<sup>3</sup> Consulté le 30/04/2018 sur <http://eduscol.education.fr/langues-vivantes/enseigner/ressources-pour-les-evaluations-et-les-examens/baccalaureats-generaux-et-technologiques/ressources-pour-les-epreuves-orales.html>

### **3.1. L'évaluation du langage verbal en production orale**

L'activité langagière de production orale répond, comme toute forme de production, à un besoin de communication :

« En didactique des langues, la *production* désigne la mise en œuvre de l'aptitude d'expression dans le cadre d'une situation de communication. » (Robert et Rosen, 2010, p. 215)

Cette production orale se présente sous plusieurs formes, telles que les annonces publiques, les exposés, la lecture d'un texte à voix haute ou encore le chant. Quelle que soit sa forme, la production orale requiert de l'élève plusieurs types de compétences qu'il convient de cibler.

#### **3.1.1. Les compétences mobilisées dans l'activité langagière de production**

La bonne réalisation de toute production orale nécessite la mobilisation de trois types distincts de compétences : 1) la compétence linguistique, 2) la compétence sociolinguistique et 3) la compétence pragmatique (Conseil de l'Europe, 2001, p. 86).

Le CECRL définit la compétence linguistique comme l'ensemble des savoirs et savoir-faire relatifs aux différentes dimensions d'un système de langue, tels que le lexique, la phonétique ou la syntaxe. Cette compétence est considérée « indépendamment de la valeur sociolinguistique de ses variations et des fonctions pragmatiques de ses réalisations. » (Conseil de l'Europe, 2001, p. 17). La compétence linguistique correspond ainsi à l'ensemble des compétences lexicales, grammaticales, sémantiques, phonologiques, orthographiques et orthoépiques. Par exemple, la réalisation d'une lettre de motivation est une activité langagière de production écrite qui nécessite un usage adéquat des temps (compétence grammaticale), ou encore la mobilisation d'un vocabulaire adapté au domaine professionnel en question (compétence lexicale).

La compétence sociolinguistique, quant à elle, correspond à l'usage social de la langue : elle renvoie « aux paramètres socioculturels de l'utilisation de la langue » (Conseil de l'Europe, 2001, p. 18), et « porte sur la connaissance et les habiletés exigées pour faire fonctionner la langue dans sa dimension sociale » (Conseil de l'Europe, 2001, p. 93). Il convient donc, au-delà de la maîtrise des savoirs et savoir-faire linguistiques, que l'apprenant connaisse les règles des relations sociales, telles que le choix des salutations, les règles de politesse adaptées, le registre de langue approprié ou encore les expressions idiomatiques. Pour reprendre l'exemple précédemment cité, la réalisation d'une lettre de motivation nécessite l'usage d'un registre formel et l'utilisation de formules de politesse adaptées.

Enfin, la compétence pragmatique renvoie à la cohérence de la production au regard de son contexte de mise en œuvre :

« La compétence pragmatique [...] recouvre l'utilisation fonctionnelle des ressources de la langue (réalisation de fonctions langagières, d'actes de parole) en s'appuyant sur des scénarios ou des scripts d'échanges interactionnels. Elle renvoie également à la maîtrise du discours, à sa cohésion et à sa cohérence, au repérage des types et genres textuels, des effets d'ironie, de parodie. Plus encore pour cette composante que pour la composante linguistique, il n'est guère besoin d'insister sur les incidences fortes des interactions et des environnements culturels dans lesquels s'inscrit la construction de telles capacités. » (Conseil de l'Europe, 2001, p. 18)

Cette compétence pragmatique comprend la compétence discursive, la compétence fonctionnelle et la compétence de conception schématique. Dans le cas de notre lettre de motivation, la compétence pragmatique renvoie à la connaissance des codes d'une lettre de motivations, c'est-à-dire à la structure attendue de ce type de lettre.

Toutes les compétences ici mentionnées constituent la base d'un découpage des activités langagières de production, qui s'applique notamment au cas de la production orale. Ce découpage autorise les auteurs du CECRL à définir certains aspects particuliers des compétences et à proposer l'utilisation d'échelles de niveaux pour mesurer la maîtrise de ces différents aspects. Ces échelles permettent donc aux utilisateurs du CECRL de se situer ou de situer les apprenants dans des niveaux de compétence. Le CECRL souligne que ces échelles ne peuvent être utilisées par les évaluateurs sans que ceux-ci ne s'interrogent sur les outils dont les apprenants auront besoin pour être performants dans la tâche qui leur est demandée (Conseil de l'Europe, 2001, p. 51).

### **3.1.2 Compétences et échelles du CECRL**

Les échelles du CECRL délimitent six niveaux différents, allant du niveau A1 au niveau C2 (Conseil de l'Europe, 2001, pp. 49-71). Elles permettent de délimiter les attendus vis-à-vis de certains aspects d'une compétence donnée. La compétence lexicale, par exemple, est définie par le CECRL comme « la connaissance et [la] capacité à utiliser le vocabulaire d'une langue qui se compose 1. d'éléments lexicaux et 2. d'éléments grammaticaux » (Conseil de l'Europe, 2001, p. 87). Cette compétence renvoie à deux critères centraux, à savoir l'étendue du vocabulaire et la maîtrise du vocabulaire. Pour chacun de ces critères, le CECRL propose des niveaux de maîtrise attendus allant de A1 à C2. La compétence grammaticale est définie, quant à elle, comme « la connaissance des ressources grammaticales de la langue et la capacité de les utiliser » (Conseil de l'Europe, 2001, p. 89). Cette fois, un seul critère est proposé par les auteurs du CECRL, celui de la correction grammaticale qui correspond

notamment à la structure des énoncés, la cohérence des temps ou encore l'accord des verbes. Une échelle allant de A1 à C2 est, encore une fois, proposée, permettant de décrire les attendus en la matière. La compétence sémantique, encore, « traite de la conscience et du contrôle que l'apprenant a de l'organisation du sens » (Conseil de l'Europe, 2001, p. 91). Si le CECRL souligne certains aspects centraux de cette compétence (comme la relation du mot et du contexte ou les relations inter-lexicales), il ne propose pas d'échelles associées permettant de détailler des niveaux.

Dans le cas spécifique de la production orale, nous ne nous intéresserons pas à la compétence orthographique, laquelle renvoie à l'écrit. La compétence phonologique, sur laquelle repose la prononciation correcte d'une langue occupe, en revanche, une place particulière dans l'évaluation de l'expression orale, puisque cette compétence ne peut être évaluée dans le cadre d'une production écrite. Le CECRL propose, en la matière, une échelle relative à la maîtrise du système phonologique. La compétence orthoépique touche également à la question de l'expression orale. Elle a trait à la capacité de produire une prononciation correcte à partir d'un écrit lu pour la première fois : le CECRL ne propose pas d'échelle précise en la matière.

Malgré des difficultés avouées, le CECRL aborde également la compétence sociolinguistique, et propose une échelle décrivant les attendus en matière de correction sociolinguistique. Il aborde enfin la compétence discursive et la compétence de conception schématique et propose des échelles détaillées relatives à la souplesse du discours, à la maîtrise des tours de parole, à la gestion du développement du discours, à la cohérence et la cohésion de celui-ci, à l'aisance orale ou encore à la précision du propos.

Le CECRL propose donc des outils détaillés permettant de préciser les compétences communicatives langagières (linguistiques sociolinguistiques et pragmatiques) nécessaires à la réalisation d'une production, qu'elle soit orale ou écrite. On a présenté, ici, les critères relatifs aux compétences mobilisables en production orale. Le CECRL propose ainsi des outils permettant de cibler certaines dimensions précises, et de décrire six niveaux de maîtrise relatifs à ces dimensions. L'évaluateur dispose alors d'indicateurs détaillés lui permettant de juger d'une production en fonction des objectifs qu'il a lui-même définis : l'existence de ces critères communs répond bien, semble-t-il, aux exigences d'explicitation des critères permettant de réduire la part de subjectivité dans l'évaluation. Mais, ces développements concernent essentiellement l'évaluation du langage verbal. Or, il convient de prendre en considération la question du langage non verbal.

## **3.2. Le langage non verbal**

La communication non verbale, ou langage non verbal, repose sur l'expression corporelle de manière générale : les gestes, les postures du corps ou encore l'expression du visage. Mais il existe également d'autres signes dits « non verbaux ». On peut citer notamment la tenue vestimentaire, l'apparence physique, l'élocution, le débit de parole ou le ton de la voix. Le CECRL aborde explicitement cette question du langage non verbal. Cependant, il ne consacre que peu de développement à ce sujet et la question de l'évaluation ne paraît centrale en la matière.

### **3.2.1 L'importance de la communication non verbale**

Le langage non verbal occupe une place non négligeable dans les méthodes d'enseignements. S'il est assez largement pré-conscient pour celui qui l'exprime, il est clairement perceptible pour les autres, dont l'attention peut alterner entre la dimension de contenu (les mots) et les indices de la relation.

La gestuelle, notamment, paraît constituer une composante importante dans l'apprentissage en langues vivantes chez les jeunes enfants, affectant les capacités de mémorisation, de compréhension et de production (Boué, 2013, p. 12 ; Buit, 2015, p. 11). Le type d'enseignement promu par le CECRL et le ministère de l'Éducation nationale invite précisément les élèves, aujourd'hui, à réussir à s'exprimer et se faire comprendre par les autres, davantage qu'à maîtriser totalement la langue :

« On passe d'une logique de maîtrise quasi totale d'une ou plusieurs langues à une logique d'interaction entre différentes langues, quel que soit le niveau de maîtrise de ces dernières. »<sup>4</sup>

La place de la gestuelle, des mimiques, des intonations n'est alors pas négligeable. Si l'on veut communiquer et qu'on ne maîtrise pas parfaitement les outils linguistiques, le langage non verbal (gestuelle et comportement paralinguistique) peut s'avérer ici primordial. Or, si les auteurs du CECRL explicitent bien les compétences nécessaires à la production orale et proposent des échelles spécifiques en la matière, la prise en compte du langage non verbal dans les questions d'évaluation ne paraît pas centrale.

### **3.2.1 Le langage non verbal dans l'évaluation**

Le CECRL consacre une partie relativement restreinte à la communication non verbale dans son chapitre 4 sur l'utilisation de la langue. Les auteurs distinguent, dans ce chapitre,

---

<sup>4</sup> Consulté le 25/04/2018 sur : <http://eduscol.education.fr/cid45678/cadre-europeen-commun-de-referenc-eccl.html>



deux composantes du langage non verbal qui ont trait à la production orale (Conseil de l'Europe, 2001, pp. 72-73).

Tout d'abord, ils rapportent les gestes et actions qui accompagnent les activités langagières. Ceux-ci comprennent la désignation, la démonstration et les actions clairement observables. Dans ces trois cas, sans le geste ou l'action, on ne peut comprendre l'énoncé. Cela signifie donc que le geste accompagne l'action et s'avère indispensable pour en comprendre le sens.

Le CECRL aborde ensuite le comportement paralinguistique. Celui-ci est composé 1) du langage du corps, dont le sens peut varier d'une culture à l'autre, 2) de l'utilisation d'onomatopées qui permettent de véhiculer un sens codé, par exemple pour marquer le dégoût et 3) de l'utilisation de traits prosodiques, par exemple le ton de la voix, son volume ou son intensité. La manière dont s'exprimera la personne, en chuchotant ou en criant, changera alors le sens véhiculé par le locuteur.

Enfin, les auteurs du CECRL s'intéressent aux éléments paratextuels, qui assurent en fait une fonction paralinguistique dans le cadre d'une production écrite : ils ont trait aux éléments d'un texte (illustrations ou éléments typographiques) qui affectent le sens de l'énoncé. Dans le cadre d'une production orale, donc, seules les deux premières composantes nous intéressent : les gestes et actions qui complètent le discours et les comportements paralinguistiques tels que le ton de la voix, son intensité, son débit, etc.

La place du langage non verbal dans l'évaluation n'est pas explicitement abordée par le CECRL. On pourrait pourtant s'interroger sur sa place dans la bonne réalisation d'une production orale et sur ses rapports avec l'évaluation des compétences communicatives langagières. En effet, si le langage non verbal accompagne l'apprentissage en langues vivantes et qu'il est susceptible d'affecter le sens de ce qui est dit dans une production orale, ne doit-il pas être pris en considération spécifiquement dans le cadre d'une évaluation de la production orale ? La moindre attention portée à cette dimension pourrait favoriser le caractère subjectif de l'évaluation orale. La prise en compte du langage non verbal pourrait ne s'effectuer que sur la base d'un jugement subjectif de l'évaluateur : celui-ci, sans avoir explicité la question du langage non verbal, pourrait valoriser ou dévaloriser une production en s'appuyant sur des critères personnels.

\*

Il semble ainsi difficile de dissocier langage verbal et langage non verbal dans l'activité de production orale. Pourtant, si le CECRL préconise des critères d'évaluation pour la production orale afin d'encadrer l'évaluation, que fait-on du langage non verbal qui peut

précisément affecter le sens du discours, perturber ou améliorer sa cohérence vis-à-vis du sujet donné, ou parasiter l'appréciation globale, notamment à travers l'effet halo ou l'effet goutte d'eau ? Dans quelle mesure, l'utilisation de critères communs peut-elle réduire la subjectivité de l'évaluation de la production orale, compte tenu de la dimension spécifique du langage non verbal en production orale ? Le cadre européen rappelle que le geste, par exemple, peut permettre de contourner la parole en cas de difficulté, mais ne peut-on pas donner une part plus significative au langage corporel et à la dimension non verbale présente dans tout acte de communication ? Cette question semble montrer une limite du CECRL qu'il conviendrait d'approfondir pour enrichir un outil déjà très complet en matière d'évaluation.

# PROBLÉMATIQUE

Nous avons vu que l'évaluation échappe difficilement à la notion de subjectivité. De nombreux parasites viennent en effet influencer la notation. Face à cette difficulté, les travaux de nombreux chercheurs ont proposé des outils permettant de diminuer la part de subjectivité présente dans l'évaluation. Dans le domaine des langues vivantes, le Cadre européen commun de référence pour les langues, document de référence, met à la disposition des enseignants de nombreux outils permettant de préciser des critères communs d'évaluation afin de mesurer les acquis et les compétences de l'élève et de réduire la part d'arbitraire dans le jugement de l'évaluateur. On peut citer notamment les grilles d'évaluation par niveau (de A1 à C2) dans les différentes compétences nécessaires à l'activité de production langagière : compétence linguistique, compétence sociolinguistique et compétence pragmatique.

Cependant, dans quelle mesure l'efficacité des outils mis à disposition par le CECRL permet-elle de gommer la subjectivité de l'évaluation ? Plus particulièrement, ces critères sont-ils efficaces dans le cas de l'évaluation de la production orale en langue vivante, qui présente des singularités non négligeables ? La dimension non verbale de cette production orale, notamment, pourrait bien renforcer ces difficultés à trouver une note juste.

Dans le cadre de nos expérimentations, nous tenterons de mettre en regard les résultats attribués par différents évaluateurs à une même production orale. Deux questions distinctes guident ces expérimentations. Y a-t-il des écarts importants entre les évaluateurs malgré la présence d'une grille d'évaluation s'inspirant des critères mentionnés dans le CECRL ? La notation, sur support audio, c'est-à-dire sans accès à l'image et à la gestuelle de l'élève, est-elle significativement différente par rapport à celle d'un évaluateur ayant accès au document vidéo lorsque l'on mobilise cette même grille ?

Deux hypothèses ont été formulées pour répondre à ces questions. D'une part, le CECRL gomme la subjectivité de l'évaluation en expression orale. D'autre part, la dimension gestuelle parasite l'objectivité de l'évaluation en expression orale.

# MÉTHODE

## 1. Participants

Pour conduire nos expérimentations, nous avons fait une sélection de six élèves (quatre filles et deux garçons) auprès de deux classes de secondes générales, au lycée Pablo Neruda à Saint-Martin-d'Hères, un lycée polyvalent en proche banlieue du centre-ville de Grenoble. Le choix des élèves s'est fait en se basant sur leur différence de niveau d'anglais, l'objectif étant d'obtenir un échantillon représentatif de l'hétérogénéité présente dans ces deux classes de seconde. Nous avons pu apprécier le niveau d'anglais de ces élèves depuis le début de l'année et nous avons estimé que leur niveau général en production orale correspondait aux niveaux A1, A2 et B1 du Cadre européen commun de référence pour les langues.

Les auteurs du CECRL proposent des descripteurs pour pouvoir apprécier les trois niveaux cités précédemment en production orale générale. Les niveaux A1, A2 et BA décrivent dans quelle mesure l'apprenant est capable de produire une présentation ou une description simple, et d'organiser un discours également simple grâce à des expressions ou des connecteurs logiques. Le niveau A1 correspond au niveau le plus faible :

PRODUCTION ORALE GÉNÉRALE	
<b>C2</b>	Peut produire un discours élaboré, limpide et fluide, avec une structure logique efficace qui aide le destinataire à remarquer les points importants et à s'en souvenir.
<b>C1</b>	Peut faire une présentation ou une description d'un sujet complexe en intégrant des arguments secondaires et en développant des points particuliers pour parvenir à une conclusion appropriée.
<b>B2</b>	Peut méthodiquement développer une présentation ou une description soulignant les points importants et les détails pertinents.
	Peut faire une description et une présentation détaillées sur une gamme étendue de sujets relatifs à son domaine d'intérêt en développant et justifiant les idées par des points secondaires et des exemples pertinents.
<b>B1</b>	Peut assez aisément mener à bien une description directe et non compliquée de sujets variés dans son domaine en la présentant comme une succession linéaire de points.
<b>A2</b>	Peut décrire ou présenter simplement des gens, des conditions de vie, des activités quotidiennes, ce qu'on aime ou pas, par de courtes séries d'expressions ou de phrases non articulées.
<b>A1</b>	Peut produire des expressions simples isolées sur les gens et les choses.

Source : Conseil de l'Europe, 2001, p. 49

## 2. Matériel

Afin de mener à bien nos expérimentations, plusieurs documents ont été nécessaires. Tout d'abord, pour s'inscrire dans une logique de respect des élèves, une fiche d'autorisation d'enregistrement de l'image et de la voix leur a été distribuée. Les élèves ont pu ainsi être

informés que leur professeure faisait des recherches sur la subjectivité de l'évaluation en production orale. Il a fallu ensuite regrouper tous les documents nécessaires au bon déroulement des expérimentations. Notamment, la consigne de la tâche finale, la grille d'évaluation accompagnée d'un barème avec des indicateurs précis et enfin, les supports vidéo et audio fournis par les élèves pour l'évaluation finale. Afin d'expliquer notre démarche, il convient de préciser les principes qui ont présidé à leur construction et à leur sélection.

Pour que tous les évaluateurs prenant part à l'expérimentation soient en mesure de noter les élèves de la manière la plus « juste » possible, il a d'abord fallu mettre à leur disposition la consigne de la tâche finale qui était demandée aux élèves, ainsi qu'une liste non exhaustive du vocabulaire étudié en classe. La partie la plus délicate a ensuite été de construire une grille d'évaluation et un barème qui permettent aux évaluateurs de noter les élèves avec la marge d'interprétation la plus faible possible. À cet effet, nous avons accompagné le barème choisi de critères précis correspondant à une note chiffrée. Cette grille, distribuée aux évaluateurs, a été construite selon des principes à la fois inspirés du Cadre européen commun, mais également de l'ouvrage de Tagliante sur l'évaluation (2005). Selon cette auteure, il serait plus facile pour l'évaluateur d'avoir à entourer une note déjà inscrite sur une grille, plutôt que de réfléchir à la note qu'il voudra donner pour tel ou tel critère :

« Il est plus simple, surtout lors de l'évaluation de la production orale, et lorsque l'enseignant a bien assimilé les critères sur lesquels il devra se prononcer, d'entourer un chiffre sur une grille, que de réfléchir à la note qu'on va attribuer pour un des critères. » (Tagliante, 2005, p. 67)

Voici donc la grille à partir de laquelle les évaluateurs ont noté les élèves :

#### Grille d'évaluation Expression Orale – 2de – Niveau visé B1

##### **S'adresser à un auditoire**

Peut faire un exposé simple et direct, préparé, sur un sujet familier.

Respect de la consigne	0	0,5	1	1,5	2		
Performance	0	0,5	1	1,5	2	2,5	3
Morphosyntaxe	0	0,5	1	1,5	2	2,5	3
Lexique approprié	0	0,5	1	1,5	2	2,5	3
Pertinence	0	0,5	1	1,5	2	2,5	3
Cohérence	0	1	2	3			
Correction phonétique	0	1	2	3			
Note :							

Le barème chiffré accompagnant cette grille d'évaluation figure en annexe (Annexe 1).

Pour pouvoir tester nos hypothèses de départ, nous nous intéresserons aux notes chiffrées sur 20 des six productions d'élèves récoltées auprès des différents évaluateurs. En tout, quatre évaluateurs auront participé à l'étude : deux enseignantes stagiaires, une tutrice académique et une enseignante en fin de carrière. Chaque évaluateur est titulaire du CAPES d'anglais.

### **3. Procédure**

Deux expérimentations ont été mises en place pour tester chacune des hypothèses de départ : d'une part, le CECRL gomme la subjectivité de l'évaluation en expression orale ; d'autre part, la dimension gestuelle parasite l'objectivité de l'évaluation en expression orale.

Pour la première expérimentation, trois élèves ont été évalués par quatre évaluateurs différents. Afin de gommer la subjectivité de l'évaluation au maximum, un certain nombre d'invariantes ont été retenues. Les évaluateurs s'étaient mis d'accord au préalable sur les attentes de l'évaluation grâce à une même grille utilisée par chacun pour la notation. La grille d'évaluation a été conçue, comme on l'a dit, en s'appuyant sur le Cadre européen commun de référence pour les langues. Celle-ci était également accompagnée d'un barème critérié permettant aux enseignants évaluant les productions de bien assimiler les critères sur lesquels ils devaient se prononcer. Les trois productions choisies étaient donc évaluées quatre fois : une fois par chaque évaluateur, l'objectif étant de comparer les notes attribuées pour les trois productions.

Dans cette expérimentation, les productions évaluées étaient sous format vidéo. Ici, deux facteurs pouvaient influencer la notation. D'une part, les évaluateurs ont chacun leurs « habitudes de notation, [leur] équation personnelle » (Barlow, 2003, p. 33) et constitue une part de subjectivité propre au jugement humain. D'autre part, un des évaluateurs connaissait les élèves, ce qui peut jouer sur la subjectivité de la notation, alors que les autres évaluateurs ne les connaissaient pas. Nous insistons ici sur le fait que les évaluations ont été faites « à l'aveugle » par les trois correcteurs qui ne connaissaient pas les élèves. Aucune information n'a été fournie sur les élèves pour ne pas influencer la notation et éviter « l'effet de contamination » qui consiste à évaluer un élève en fonction du niveau qui lui est généralement attribué.

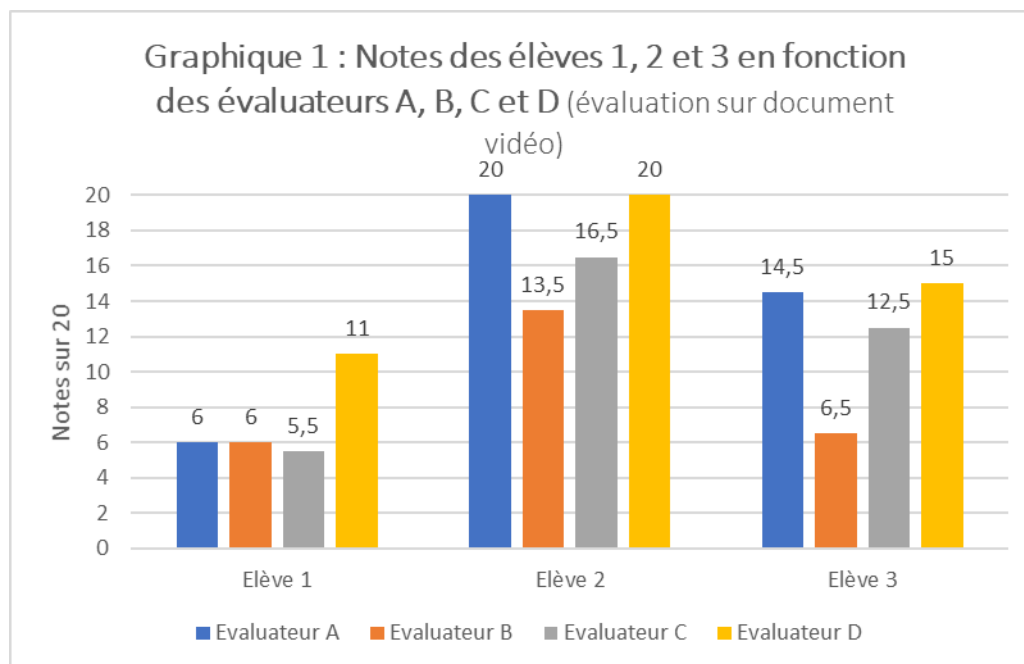
La seconde expérimentation était sensiblement identique. Trois autres productions étaient à nouveau évaluées. Afin de tester la deuxième hypothèse, cette fois, trois correcteurs ont fait l'évaluation à partir d'enregistrements audio uniquement, alors que j'avais effectué la notation à partir des vidéos. Cela permettait ainsi de mettre en avant l'influence ou non de la

dimension gestuelle dans les expressions orales des élèves. Du reste, la procédure de cette seconde expérimentation était identique à la première.

Les deux expérimentations se sont déroulées sur un mois, le temps de récolter les données de tous les évaluateurs. Il nous semble important de noter que les évaluations n'ont pas été faites au même moment ni dans les mêmes conditions. Certains évaluateurs ont parfois regardé la vidéo ou écouté l'enregistrement plusieurs fois, d'autres se sont contentés d'un seul visionnage ou d'une seule écoute. Ces paramètres n'ont pas pu être pris en compte dans le déroulé des expérimentations mais c'est un élément qu'il faudrait prendre en compte à l'avenir.

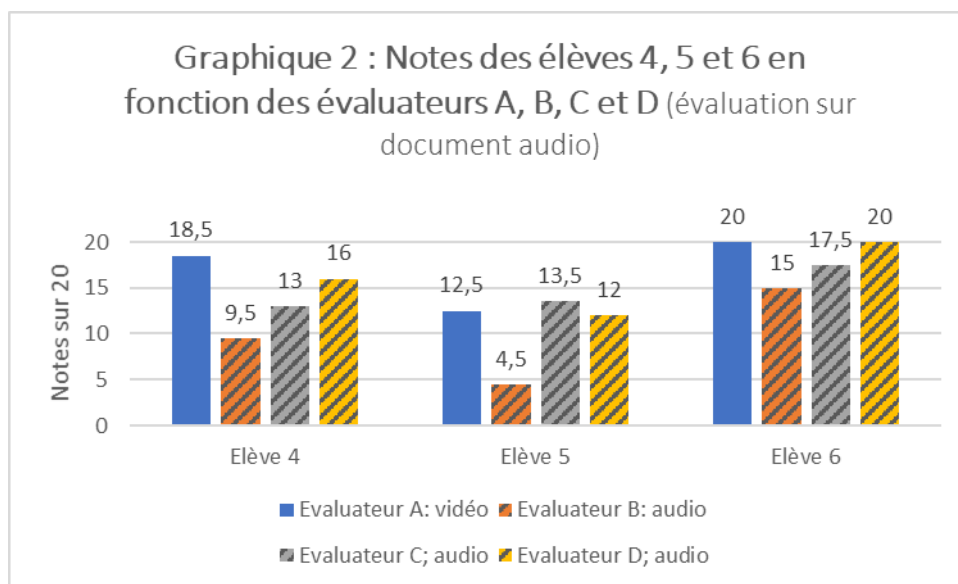
#### 4. Analyse des résultats

Les résultats obtenus lors des deux expérimentations ont été rassemblés sous forme de graphiques pour permettre une meilleure visualisation.



Dans le graphique 1, pour chaque élève 1, 2 et 3, on compte quatre notes sur 20. Les notes de l'évaluateur A sont en bleu, celles de B en orange, celles de C en gris et celles de D sont en jaune. Ces notes correspondent à l'appréciation, pour chaque élève, de 7 compétences requises : respect de la consigne, performance, morphosyntaxe, lexique, pertinence, cohérence, correction phonétique (ces compétences ont été évaluées selon le barème précédemment indiqué). Toutes les notes des élèves 1, 2 et 3 ont été attribuées à partir d'un document vidéo.

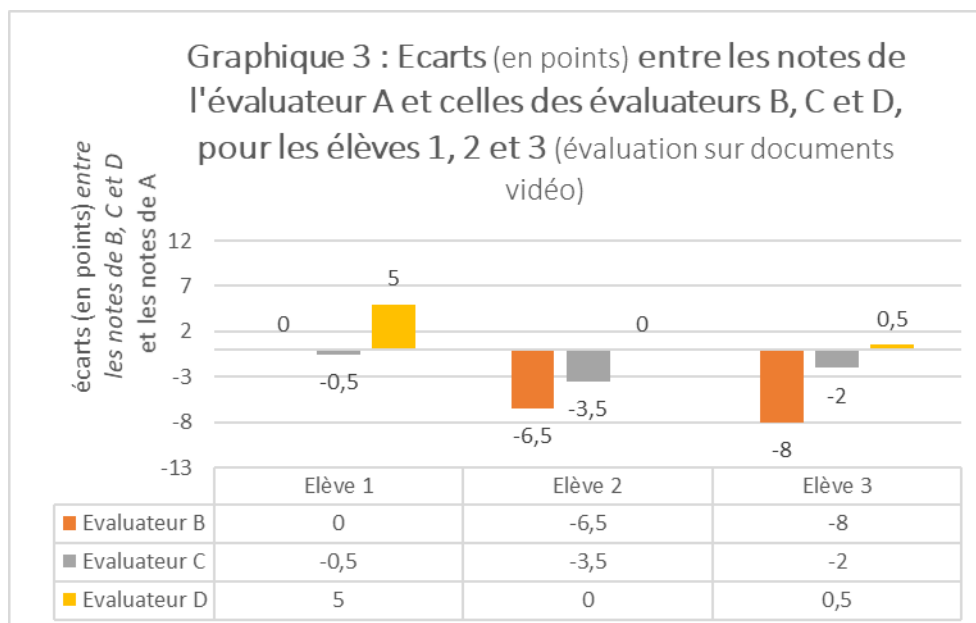
On constate, de façon générale, une disparité des notes attribuées par chaque évaluateur pour un même élève.



Dans le graphique 2, pour chaque élève 4, 5 et 6, on compte à nouveau quatre notes sur 20. De la même manière que précédemment, les notes de l'évaluateur A sont en bleu, celles de B en orange, celles de C en gris et celles de D en jaune. Ces notes correspondent à l'appréciation, pour chaque élève, de 7 compétences requises mentionnées plus haut : respect de la consigne, performance, morphosyntaxe, lexique, pertinence, cohérence, correction phonétique. Les notes attribuées aux élèves 4, 5 et 6 par l'évaluateur A ont été attribuées à partir d'un document vidéo. En revanche, les évaluateurs B, C et D ont noté les élèves 4, 5 et 6 à partir d'un document audio. Il s'agissait des mêmes documents mais les évaluateurs B, C et D n'avaient pas accès à l'image pour ces trois élèves.

On constate, à nouveau, une disparité des notes attribuées par chaque évaluateur pour un même élève.





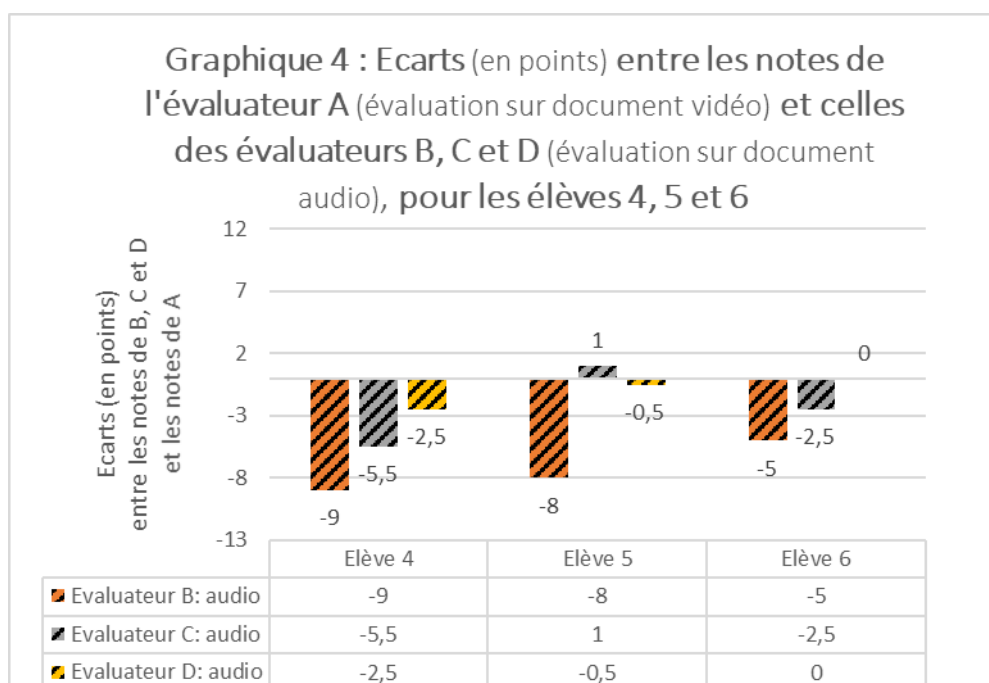
Le graphique 3 représente les écarts de notes, pour les élèves 1, 2 et 3, entre l'évaluateur A et les évaluateurs B, C et D. Ces notes ont été attribuées sur la base d'un document vidéo. Les écarts de points négatifs (barres descendantes) signifient que l'évaluateur B, C ou D a donné une note inférieure à celle que l'évaluateur A a donné pour les élèves 1, 2 et 3. À l'inverse, les écarts positifs (barres montantes) signifient que l'évaluateur B, C ou D a donné une note supérieure à A pour la production orale d'un même élève<sup>5</sup>. L'absence d'écart (0) signifie que l'évaluateur B, C ou D a donné la même note que l'évaluateur A. Les écarts avec l'évaluateur B sont représentés en orange, ceux avec l'évaluateur C sont en gris et les écarts avec l'évaluateur D sont en jaune. Les mêmes codes couleur ont été repris afin faciliter la lecture des résultats.

Ici, on constate que seul l'évaluateur D a donné des notes supérieures à celles données par l'évaluateur A, et il n'a donné aucune note inférieure. À l'inverse, les évaluateurs B et C ont systématiquement donné des notes inférieures ou égales aux notes de l'évaluateur A, mais jamais de note supérieure. De plus, les notes de l'évaluateur D sont quasi identiques à celles de l'évaluateur A pour les élèves 2 et 3 (0.5 point d'écart ou aucun écart), tandis que la note attribuée à l'élève 1 montre un écart important de 5 points : l'évaluateur D a été beaucoup plus généreux que A pour l'élève 1, et les notes qu'il a attribuées aux élèves 2 et 3 sont quasi identiques à celles attribuées par A. En d'autres termes, les notes attribuées par les évaluateurs

<sup>5</sup>Par exemple, si l'écart de points entre la note de l'évaluateur A et celle de l'évaluateur B pour l'élève 1 est de 0 point, alors l'évaluateur A et l'évaluateur B ont donné la même note à l'élève 1 ; si l'écart de points entre la note de l'évaluateur A et celle de l'évaluateur C pour l'élève 2 est de -3,5 points, alors l'évaluateur C a donné 3,5 points de moins que l'évaluateur A à l'élève 2 ; si l'écart de points entre la note de l'évaluateur A et celle de l'évaluateur D pour l'élève 1 est de 5 points, alors l'évaluateur D a donné 5 points de plus que l'évaluateur A à l'élève 1.

A et D sont très proches pour les élèves 2 et 3, tandis que l'évaluateur D a été beaucoup plus généreux pour l'élève 1.

À l'inverse, les notes attribuées par l'évaluateur B et C sont inférieures à celles de A pour les élèves 2 et 3 (respectivement -6,5 ; -8 ; -3,5 ; -2), tandis que les notes attribuées à l'élève 1 présentent des écarts très faibles (0 ; 0,5). Les notes de l'évaluateur B pour les élèves 2 et 3 présentent davantage d'écarts avec celles de l'évaluateur A que les notes de l'évaluateur C par rapport à A. En d'autres termes, l'évaluateur B tend à juger plus sévèrement les élèves 2 et 3 que les évaluateurs A et C ; et l'évaluateur C tend à juger plus sévèrement que l'évaluateur A pour ces mêmes élèves. En revanche, les notes des évaluateurs B et C pour l'élève 1 sont très proches des notes attribuées par l'évaluateur A.

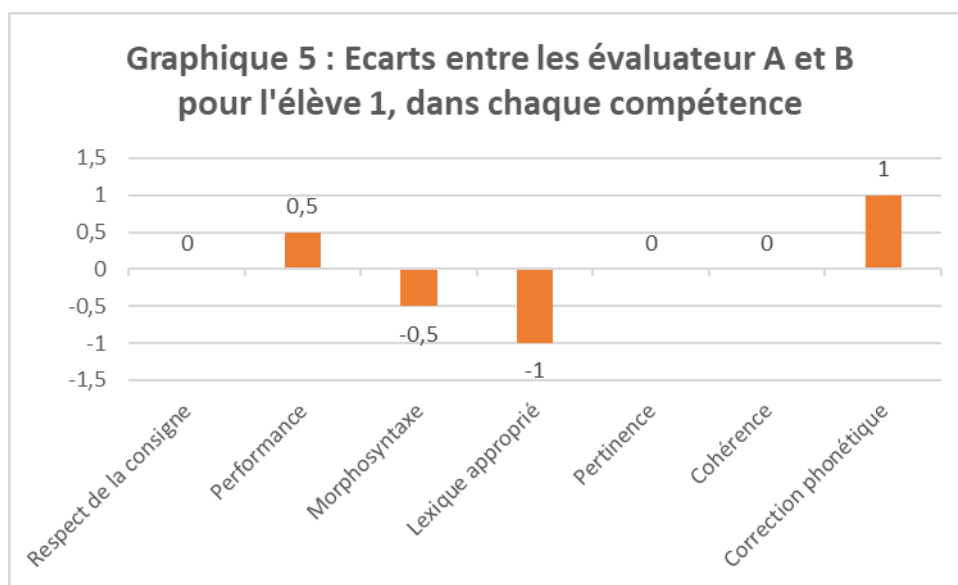


Le graphique 4 représente les écarts de notes, pour les élèves 4, 5 et 6, entre l'évaluateur A et les évaluateurs B, C et D. Ces notes ont été attribuées sur la base d'un document audio par les évaluateurs B, C et D (les rayures sur les barres indiquent cette spécificité). Comme dans le graphique précédent, les écarts de points négatifs (barres descendantes) signifient que l'évaluateur B, C ou D a donné à une note inférieure à celle que l'évaluateur A a donné aux élèves 4, 5 ou 6. À l'inverse, les écarts positifs (barres montantes) signifient que l'évaluateur B, C ou D a donné une note supérieure à A.

Dans l'ensemble, l'évaluateur A tend à noter de manière plus généreuse que l'évaluateur C. Les notes attribuées par l'évaluateur C montrent en effet des écarts négatifs relativement importants par rapport aux notes de l'évaluateur A (-5,5 points pour l'élève 4 et -2,5 points

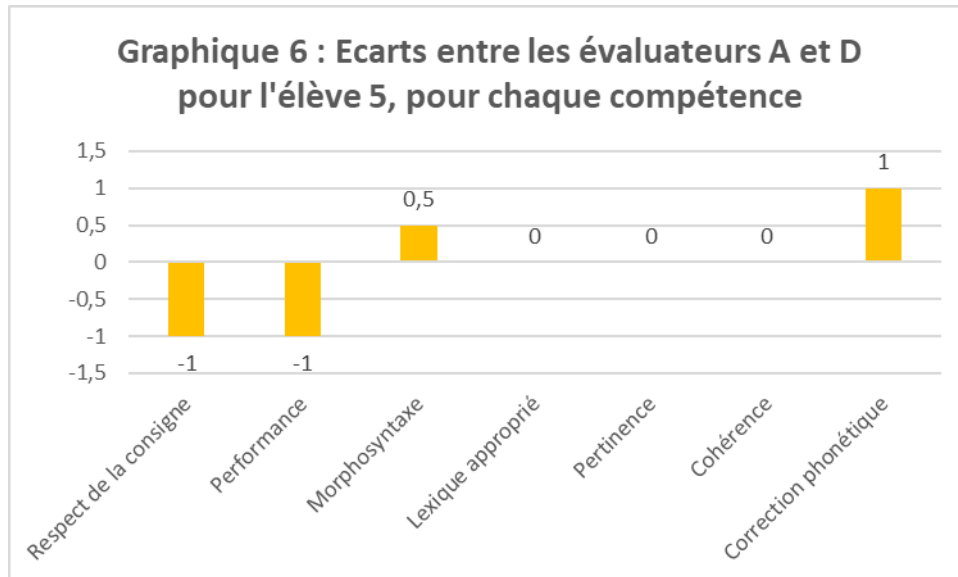
pour l'élève 6). En revanche, il n'y a qu'un point de différence entre l'évaluateur C et l'évaluateur A pour l'élève 5. On constate, ici, que seul l'évaluateur C a donné une note supérieure à l'évaluateur A pour l'élève 5. L'évaluateur B, pour sa part, a systématiquement donné des notes inférieures aux notes de l'évaluateur A, comme c'était le cas dans la première expérimentation, avec des écarts allant jusqu'à 9 points (pour l'élève 4) : les écarts sont relativement importants pour les trois élèves (-9 points pour l'élève 4 ; -8 points pour l'élève 5 et -5 points pour l'élève 6).. L'évaluateur D, quant à lui, a attribué des notes très proches de celles attribuées par l'évaluateur A pour les élèves 5 et 6 (0,5 point d'écart ou aucun écart), tandis qu'il a été un peu plus sévère à l'égard du travail de l'élève 4 : la note attribuée à l'élève 4 présente un écart négatif de 2,5 points.

Les deux prochains graphiques ont été sélectionnés afin d'illustrer deux situations spécifiques qu'il nous a paru intéressant de souligner. Parmi les écarts faibles (entre 0 et 0,5 point), les points attribués par les évaluateurs ne sont pas les mêmes dans chaque critère évalué.



On observe ainsi, dans le graphique 5, les écarts entre les notes attribuées par l'évaluateur A et l'évaluateur B pour l'élève 1. Celui-ci a obtenu des deux évaluateurs la note de 6 sur 20. En revanche, on constate que les points n'ont pas été répartis de la même manière selon les critères. L'écart global de 0 correspond au fait que les écarts de notes se sont annulés en s'additionnant. L'évaluateur B a donné 0,5 point de plus que l'évaluateur A pour la performance et 1 point de plus pour la correction phonétique. En revanche, il a donné 0,5 point de moins que l'évaluateur A pour la morphosyntaxe et 1 point de moins pour le lexique. Sur la note globale, on ne perçoit pas ses différences de notation puisque la note globale est la

même pour les deux évaluateurs. Il nous a semblé intéressant de pointer que même pour les évaluations avec un écart faible ou nul, la notation n'a pas toujours été la même si l'on s'intéresse aux critères contenus dans la grille d'évaluation : en valeur absolue, la somme des écarts, c'est-à-dire la différence absolue d'appréciation, vaut tout de même 3 points.



Dans le graphique 6, nous avons procédé de la même manière, en considérant les notes de l'élève 5 pour les évaluateurs A et D. L'élève 5 a obtenu une note de 12,5 sur 20 par l'évaluateur A et une note de 12 par l'évaluateur B. Ici, l'écart global entre les deux notes est de 0,5. C'est un écart faible mais on constate, en regardant le graphique 6, que les points n'ont pas été attribués de la même façon par les deux évaluateurs. L'évaluateur D a donné 1 point de moins que l'évaluateur A pour le respect de la consigne et la performance. En revanche, l'évaluateur D a donné 0,5 point et 1 point de plus que l'évaluateur A, respectivement pour la morphosyntaxe et la correction phonétique.

# DISCUSSION et CONCLUSION

L'objectif de cette étude est d'apprécier les limites de l'objectivité que chaque enseignant peut porter à l'évaluation de ses élèves. Deux expérimentations ont été mises en place afin de tester deux hypothèses de départ : d'une part, le Cadre européen commun de référence pour les langues comme la subjectivité de l'évaluation en production orale ; d'autre part, la dimension gestuelle parasite l'objectivité de cette évaluation. Pour chacune des expérimentations, quatre évaluateurs ont noté trois élèves à partir d'une grille d'évaluation élaborée en s'inspirant du CECRL.

Dans la première expérimentation, trois élèves ont été évalués sur support vidéo par les quatre évaluateurs. Pour la deuxième expérimentation, trois autres élèves ont été évalués : seul l'évaluateur A a noté les élèves à partir d'une vidéo, tandis que les trois autres évaluateurs ont noté les mêmes élèves à partir d'un support audio. Les résultats obtenus suggèrent que la question de la subjectivité dans l'évaluation ne peut être résolue entièrement tant qu'elle se fonde sur une normativité humaine.

## 1. Une subjectivité à toute épreuve ?

Si le CECRL propose des pistes de réflexions et des outils pertinents visant à diminuer la part de subjectivité dans l'évaluation de la production orale, force est de constater que les données de l'expérimentation n°1 montrent des disparités importantes dans les notes attribuées par les évaluateurs A, B, C et D pour les élèves 1, 2 et 3. Dans l'idéal, les notes attribuées par les quatre évaluateurs auraient dû être identiques pour chaque élève.

Certes, le CECRL ne prétend pas fournir des solutions permettant d'atteindre à une objectivité totale. Cependant, au regard des écarts constatés, on est loin de trouver dans l'usage de critères communs un remède adéquat au problème de la subjectivité. Les écarts de points entre les notes globales attribuées par l'évaluateur A et celles attribuées par l'évaluateur B pour les élèves 2 et 3, par exemple, sont respectivement de 6,5 et 8 points. Si les écarts relevés entre les notes attribuées par les évaluateurs A et B sont les plus grands, d'autres écarts par rapport à A ne sont pas moins significatifs : on trouve notamment un écart de 5 points avec l'évaluateur D, et un écart de 3,5 points avec l'évaluateur C.

Ces différences vont généralement dans le même sens : l'examineur A a eu tendance à noter plus généreusement ses élèves. Cependant, les écarts constatés ne traduisent pas simplement une trop grande générosité de A à l'égard de ses élèves. L'évaluateur D a souvent

attribué des notes proches de celles attribuées par l'évaluateur A, et il a parfois noté bien plus généreusement que celui-ci (5 points d'écart pour l'élève 1). De plus, des écarts entre les évaluateurs B, C ou D indiquent que la subjectivité ne se résume pas, dans ce cas-ci, à la connaissance des élèves, puisque ces trois évaluateurs ne les connaissaient pas. On notera, par exemple, un écart de 6 points entre les notes attribuées par les évaluateurs B et C pour l'élève 3, et un écart de 8,5 points entre B et D pour ce même élève. Certains facteurs relatifs aux caractéristiques mêmes des évaluateurs pourraient être en cause.

C'est ce que l'on déduit, notamment, des écarts constatés entre l'évaluateur B et l'évaluateur A, qui sont les plus fréquents et les plus importants : les notes attribuées par B sont par ailleurs systématiquement en deçà des notes données par A. Or, l'évaluateur B est professeur en classe préparatoire. On peut supposer que ses attentes sont plus élevées et que, de ce fait, il a noté plus « sévèrement » les élèves. Il y aurait là un effet de contraste lié aux habitudes de B, généralement confronté à des élèves de meilleur niveau. À l'inverse, l'évaluateur D, qui présente le moins d'écart avec l'évaluateur A, est enseignant en collège. Il est, par ailleurs, le seul à avoir attribué une note significativement supérieure à celle attribuée par A (5 points d'écart pour l'élève 1). Il se peut, ici, que l'évaluateur D ait noté les élèves de manière plus généreuse par effet de contraste avec ses propres élèves, ceux-ci étant généralement d'un niveau plus faible.

Malgré les difficultés à départir les effets attribuables à l'existence d'une relation préalable entre l'évaluateur et l'élève et ceux liés aux habitudes des évaluateurs, ces résultats suggèrent que le jugement propre à chaque évaluateur pèse sur l'évaluation. Les écarts observés ne permettent pas d'affirmer, dès lors, que des critères inspirés du Cadre européen de référence pour les langues sont suffisants pour réduire la subjectivité de l'évaluation. L'hypothèse n°1, selon laquelle la mobilisation de critères gomme la subjectivité, paraît infirmée. Cependant, au-delà des facteurs associés aux caractéristiques de chaque évaluateur (connaissance des élèves ou attentes spécifiques de l'évaluateur), on constate des différences liées à l'usage même des grilles d'évaluation.

En effet, les écarts très importants entre l'évaluateur A et l'évaluateur B peuvent s'expliquer par une utilisation spécifique de la grille par l'évaluateur B. En observant celui-ci lors de son évaluation des élèves, on a constaté qu'il notait d'abord les élèves en se basant sur son impression générale. Au moment de noter chaque critère sur la grille, cet évaluateur a ainsi entouré les notes du barème afin que leur somme corresponde à un total proche de sa première impression. En d'autres termes, l'utilisation de la grille était subordonnée, ici, au

jugement initial de l'évaluateur, et l'usage du barème vise essentiellement à retrouver la note initialement attribuée.

Une autre explication concerne moins l'usage différencié de la grille que les différences dans l'appréciation de chaque compétence, quand bien même des consignes précises visaient à assurer une compréhension commune des critères. En effet, lorsqu'on s'intéresse aux écarts faibles entre les notes attribuées à un même élève, on constate que la notation de chaque compétence par les évaluateurs n'est pas tout à fait la même. En d'autres termes, pour un même élève, les évaluateurs n'ont pas réparti les points de manière strictement identique selon les différents critères utilisés. Un écart nul entre les notes globales d'un élève peut ainsi masquer en réalité des écarts dans l'appréciation de chaque critère. C'est le cas des écarts entre les évaluateurs A et B pour l'élève 1 : les évaluateurs ont attribué à l'élève 1 une même note globale (6/20), mais ils ont apprécié différemment 4 critères sur 7.

\*

On trouve donc, dans les résultats de l'expérimentation n°1, des différences notables en matière de notation, qui ne permettent pas de sanctionner l'efficacité et la validité de la grille d'évaluation et du barème utilisés. Qu'en est-il des résultats de l'expérimentation n°2, qui visait à tester l'impact de la communication non verbale sur l'évaluation de la production orale, en contrôlant notamment la dimension gestuelle de la production ?

## **2. L'impact du langage non verbal sur l'évaluation**

L'expérimentation n°2 avait pour but de distinguer la manière dont un évaluateur évaluait une production orale avec ou sans avoir accès à la dimension gestuelle de l'élève, et d'apprécier ainsi la dimension non verbale de la production. Certes, la notion de communication non verbale ne se limite pas uniquement à la gestuelle, mais englobe également des éléments tels que le débit de la voix ou le ton. En abordant ainsi la question du langage non verbal, il aurait donc été difficile de dissocier d'une part, ce qui avait trait à la gestuelle et qui n'était perceptible que pour les évaluations sur support vidéo et, d'autre part, les éléments relatifs à l'expression (débit de la voix, ton) et qui étaient communs à l'évaluation sur support audio et vidéo. Ceci étant, l'expérimentation permettait d'apprécier le poids de la gestuelle dans la note de l'élève et de montrer en quoi cette dimension peut affecter sa note.

Les évaluateurs B, C et D ont ainsi noté les élèves 4, 5 et 6 à partir d'un document audio, sans avoir accès aux images, en mobilisant les grilles d'évaluations de l'expérimentation n°1.

L'évaluateur A, pour sa part, procédait à l'évaluation de ces mêmes élèves à partir d'un document vidéo.

Il est difficile d'interpréter les résultats de cette seconde expérimentation, en raison des grandes disparités de notations entre les quatre évaluateurs lors de la première expérimentation. N'ayant pas validé la première hypothèse, il n'est pas possible de distinguer, ici, la part de subjectivité propre à chaque évaluateur et le rôle de la gestuelle dans les écarts. Les écarts constatés entre l'évaluateur A (qui disposait d'un support vidéo) et l'évaluateur B (qui ne disposait que du support audio) sont relativement prononcés, comme c'était déjà le cas dans l'expérimentation n°1. De même, les écarts entre l'évaluateur A et l'évaluateur D sont relativement faibles, comme dans la première expérimentation, même si, ici l'évaluateur D n'a pas eu tendance à noter plus généreusement que A.

Il est donc complexe de tirer de cette expérimentation des éléments répondant à notre hypothèse. On peut toutefois noter que ces résultats vont dans le même sens que ceux de la première expérimentation.

D'abord, l'évaluateur A a eu tendance à noter plus généreusement ses propres élèves. Ici, le cas de l'élève 4 est particulièrement parlant, et il autorise une interprétation plus précise. En effet, les écarts entre les notes attribuées par l'évaluateur A à l'élève 4 et celles attribuées par les évaluateurs B, C et D sont très importants (-9 ; -5,5 ; -2,5 respectivement). Or, ici, l'élève 4 est généralement introverti et présente des difficultés à s'exprimer à l'oral. L'évaluateur A est au fait de cette caractéristique puisqu'il connaît bien l'élève. La note généreuse attribuée à l'élève 4 par l'évaluateur A pourrait alors s'expliquer par la connaissance que A a de l'élève : l'évaluateur a pu vouloir que les efforts de l'élève 4 soient récompensés, compte tenu de son manque d'assurance initial. Il s'agit là d'une part de subjectivité liée à la connaissance des élèves et qui, malgré la grille, a influencé l'évaluation. L'anonymisation de la production étant une affaire complexe à l'oral, les effets de la relation entre l'évaluateur et son élève sur la notation semblent difficiles à maîtriser.

De même, dans cette seconde expérimentation, lorsque les notes globales attribuées par deux évaluateurs à un même élève sont proches, on trouve des différences d'appréciations pour chaque compétence. C'est ce qu'on a pu observer lors de la première expérimentation. C'est le cas notamment pour l'élève 5. En effet, l'écart entre les notes globales attribuées par les évaluateurs A et D est faible (0,5 point). Cependant, les évaluateurs ont apprécié différemment 4 critères sur 7. Ces résultats suggèrent donc que l'appréciation de chaque critère tend à différer selon les évaluateurs, et ce, même lorsqu'il s'agit d'une production évaluée sur support audio.



### 3. Retour critique

Notre étude présente plusieurs limites sur lesquelles il est nécessaire de revenir. La première de ces limites tient au fait que seulement quatre évaluateurs ont participé aux expérimentations. Pour une étude détaillée des relations précédemment rapportées, il aurait été nécessaire de disposer d'un échantillon plus conséquent. Cependant, les limites de cette étude touchent davantage à l'organisation du protocole mis en place.

La composition de la grille d'évaluation qui a été retenue pour les expérimentations pose quelques difficultés qui ont pu altérer la pertinence de notre expérience. En effet, les auteurs du Cadre européen commun de référence pour les langues précisent « qu'au-delà de quatre ou cinq catégories on est cognitivement saturé et que sept catégories constituent un seuil psychologique à ne pas dépasser » (Conseil de l'Europe, 2001, p. 145). Il faut comprendre ici le mot « catégorie » comme critère d'évaluation. Or, dans notre grille d'évaluation, nous avons pris en compte sept critères pour évaluer les productions des élèves : respect de la consigne, performance, morphosyntaxe, lexique, pertinence, cohérence, correction phonétique. Le but du barème chiffré, qui contient des indicateurs et le détail d'attribution des points pour chaque critère, était précisément de réduire au maximum cette marge d'appréciation personnelle.

Cependant, il se peut qu'il ait eu l'effet inverse de celui escompté. En effet, le choix de retenir sept critères dans le cadre de notre expérimentation constitue la limite haute du nombre acceptable de critères selon le CECRL. Or, si le barème visait avant tout à guider les évaluateurs au maximum pour ne laisser que peu de place à l'interprétation des critères, il se peut que le nombre d'informations à contrôler pour chaque critère ait mené à une surcharge cognitive. En d'autres termes, la quantité importante de critères ainsi que les précisions relatives à leur mesure ont pu perturber l'appréciation de l'évaluateur, sa disponibilité et son discernement. L'évaluateur peut bien avoir perdu de vue les éléments centraux dans le cadre de l'évaluation de la production en question, en raison du nombre élevé d'éléments à contrôler.

Il aurait peut-être fallu, en conséquence, faire des choix qui limitent encore davantage le nombre de critères présents dans la grille d'évaluation. L'idéal aurait été de construire le scénario d'apprentissage de la séquence pour laquelle les élèves ont eu à produire un discours en co-enseignement. Cela aurait permis un double regard pour la mise en place des activités, mais également pour la création de la grille d'évaluation qui s'avère un exercice complexe. Opérer des choix parmi une quinzaine de critères peut parfois poser des problèmes quand

l'objectif de la séquence pédagogique intègre plusieurs de ces critères mais qu'il est nécessaire d'en écarter certains.

Une autre limite concerne le déroulement des deux expérimentations. Pour que l'évaluation se fasse dans les mêmes conditions pour chaque évaluateur, il aurait été intéressant de faire faire la notation au même moment, avec les mêmes consignes orales et écrites. Cela aurait peut-être gommé l'impact de facteurs extérieurs sur la notation des élèves.

\*

\*      \*

L'évaluation est un aspect complexe de l'enseignement. Professeurs comme élèves la redoutent et, cependant, y tiennent. Au cours d'une enquête par questionnaires menée en 1995, Odile et Jean Veslin (1995) relevaient déjà cette ambiguïté. Les professeurs interrogés reléguent l'exercice parmi ceux qui suscitaient le moins d'enthousiasme, et redoutaient l'angoisse du fameux « tas de copies » ainsi que les potentielles injustices qu'ils pouvaient y commettre. Pour autant, ils reconnaissent son caractère d'« évidence ». Les élèves, quant à eux, étaient forts demandeurs de notes, sans pour autant cacher leur rancœur, et leur sentiment d'être parfois jugés « à la tête du client ».

Ce malaise, qui entoure souvent l'activité d'évaluation, est bien légitime. Il est donc nécessaire de mettre en place des outils efficaces d'évaluation permettant d'assurer une équité optimale. Nos expérimentations avaient pour objectifs d'interroger ces outils en matière d'évaluation de la production orale en langues vivantes. Elles avaient également pour but de mesurer le poids de la gestuelle dans la notation des élèves dans ce type d'activité, considérant le manque de critères relatifs à la dimension non verbale de la communication dans le CECRL. Nos résultats montrent la difficulté à établir une grille et un barème permettant d'atteindre une certaine objectivité dans l'évaluation de la production orale.

# BIBLIOGRAPHIE

## Œuvres littéraires

- Antibi, A. (2007). *Pour en finir avec la Constante Macabre ou Les notes : la fin du cauchemar*. Toulouse: Math'Adore.
- Baillat, G., De Ketele, J.-M., Paquay, L., & Thélot, C. (2008). *Evaluer pour former : Outils, dispositifs et acteurs*. Bruxelles: Groupe De Boeck.
- Barlow, M. (2003). *L'évaluation scolaire, mythes et réalités*. Issy-les-Moulineaux: ESF éditeur.
- Butera, F., Buchs, C., & Darnon, C. (2011). *L'évaluation, une menace?* Paris: Presses Universitaires de France.
- Conseil de l'Europe. (2001). *Un Cadre européen de référence pour les langues : Apprendre, enseigner, évaluer*. Paris: Les Editions Didier.
- Dubus, A. (2006). *La notation des élèves : Comment utiliser la docimologie pour une évaluation raisonnée*. Paris: Armand Colin.
- Guimard, P. (2010). *L'évaluation des compétences scolaires*. Rennes: Presses universitaires de Rennes.
- Hadji, C. (2012). *Faut-il avoir peur de l'évaluation ?* Bruxelles: Groupe De Boeck.
- Hameline, D. (2005). *Les objectifs pédagogiques : En formation initiale et en formation continue* (14<sup>e</sup> éd.). Issy-les-Moulineaux: ESF éditeur.
- Mager, R. F. (2005). *Comment définir des objectifs pédagogiques* (2<sup>e</sup> éd.). Paris: Dunod.
- France. Ministère de l'Education nationale. Organisation de l'enseignement des langues vivantes étrangères dans l'enseignement scolaire, réglementation applicable à certains diplômes nationaux et commission académique sur l'enseignement des langues vivantes étrangères : Décret n° 2005-1011 du 22-8-2005. Bulletin Officiel de l'Education nationale [en ligne], 1<sup>er</sup> septembre 2005, n°31, [consulté le 16 avril 2018]. Disponible sur <http://www.education.gouv.fr/bo/2005/31/MENE0501621D.htm>

Peretti (de), A., Boniface, J., & Legrand, J.-A. (2005). *Encyclopédie de l'évaluation en formation et en éducation : Guide pratique* (3<sup>e</sup> éd.). Issy-les-Moulineaux: ESF éditeur.

Piéron, H. (1969). *Examens et docimologie* (2<sup>e</sup> éd.). Paris: Presses Universitaires de France.

Riba, P., & Mègre, B. (2014). *Démarche qualité et évaluation en langues*. Paris: Hachette.

Robert, J.-P., & Rosen, E. (2010). *Dictionnaire pratique du CECR*. Paris: Editions OPHRYS.

Tagliante, C. (2005). *L'évaluation et le Cadre européen commun*. Paris: CLE International.

Veslin, O., & Veslin, J. (1995). *Corriger des copies : évaluer pour former*. Paris: Hachette éducation.

### **Mémoires**

Boué, M. (2013). *La communication non verbale et la place du corps en classe de langues à l'école primaire*. Mémoire de Master 2 EFE-ESE, IUFM Midi-Pyrénées.

Buit, A. (2015). *Faire ses premiers pas en anglais à l'aide de la gestuelle*. Mémoire de Master 2 MEEF, ESPE de Grenoble.

### **Documents internet**

*Baccalauréats généraux et technologiques - Ressources pour les épreuves orales*. (s.d.). Consulté 30 avril 2018, à l'adresse <http://eduscol.education.fr/langues-vivantes/enseigner/ressources-pour-les-evaluations-et-les-examens/baccalaureats-generaux-et-technologiques/ressources-pour-les-epreuves-orales.html>

*CECRL : le Cadre européen commun de référence pour les langues*. (18 novembre 2014). Consulté le 25 avril 2018, à l'adresse <http://eduscol.education.fr/cid45678/cadre-europeen-commun-de-referance-cecrl.html>

*Enseigner les langues vivantes à l'école élémentaire*. (10 avril 2017). Consulté le 23 février 2018, à l'adresse <http://eduscol.education.fr/cid45718/les-langues-vivantes-etranangeres-ecole.html>

# ANNEXES

## Annexe 1 : barème chiffré / Document à disposition des évaluateurs

### Rappel TF :

You are the UN gender equality ambassador. You make a speech in front of the United Nations' representatives and you explain why gender equality should be a priority today.

Focus on: (éléments en italique donnés par les élèves)

*Structure "It is said that ..." / "... are said to..."*

*Body language (eye-contact / body gestures and posture )*

*Accentuation / pronunciation / speed of the voice*

*Arguments*

*Speech = 2 minutes (minimum)*

Nom :

Grille d'évaluation Expression Orale – 2de – Niveau visé B1

### **S'adresser à un auditoire**

Peut faire un exposé simple et direct, préparé, sur un sujet familier.

<b>Respect de la consigne</b>		- Temps respecté (2 minutes minimum) → 0,5 pt - Discours → 0,5 pt - Adresse aux représentants des NU → 0,5 pt - Ambassadeur/drice de l'égalité femme-homme → 0,5 pt	2
<b>Performance</b>	B1	- Assez clair(e) pour être suivi(e) sans difficulté la plupart du temps + Points importants expliqués avec assez de précision (au moins 5) → 1 pt - Objectif maîtrisé = convaincre → 1 pt - maîtrise de la « forme » : débit de la voix, ton, pauses, langage non-verbal (posture, contact visuel, gestuelle) → 1 pt	3
	A2	- Intelligible malgré quelques hésitations + points importants assez précis malgré quelques erreurs (3-4 points) → 1 pt - Objectif en voie d'acquisition = auditoire presque convaincu → 0,5 pt - maîtrise satisfaisante de la forme mais 1 ou 2 aspects ne sont pas utilisés (parmi débit de la voix, ton, pauses, langage non-verbal) → 0,5 pt	2
	A1	- peut faire un bref exposé élémentaire, répété, sur un sujet familier + 1 ou 2 points importants intelligibles → 0,5 pt - Objectif insuffisant = difficultés à convaincre + forme insuffisamment exploitée (seulement 2 ou 3 éléments pris en compte) → 0,5	1
		- exposé peu clair, nombreuses hésitations - aucune mention de points importants - objectif non maîtrisé = exposé non convaincant - forme non exploitée	0

<b>Morphosyntaxe / correction grammaticale</b>	B1	- utilisation fréquente de la structure « it is said that » → 1 pt - Utilise de façon assez exacte un répertoire de structures et « schémas » fréquents (phrases complexes) → 1 pt - cohérence des temps → 1 pt	3
	A2	- utilisation régulière de la structure « it is said that » - Utilise des structures simples correctement mais commet encore systématiquement des erreurs élémentaires - quelques confusions dans les temps	2
	A1	- utilisation ponctuelle de la structure « it is said that » - A un contrôle limité de quelques structures syntaxiques et de formes grammaticales simples appartenant à un répertoire mémorisé - confusion des temps	1
		- structure « it is said that » non-utilisée - structures élémentaires non maîtrisées, nombreuses erreurs élémentaires - confusion dans les temps	0
<b>Lexique approprié</b>	B1	- réinvestissement du lexique introduit en cours + recherche de vocabulaire adapté à son discours - précision des mots	3
	A2	- réinvestissement du lexique introduit en cours - mots parfois imprécis	2
	A1	- réinvestissement partiel du lexique introduit en cours - manque de précision dans l'utilisation de certains mots	1
		- absence du vocabulaire introduit en cours - imprécision des mots	0
<b>Pertinence</b>	B1	- pertinence des arguments présentés dans le discours	3
	A2	- arguments simples et clairs mais parfois peu pertinents dans le discours	2
	A1	- arguments simples mais parfois sans pertinence par rapport au sujet	1
		- arguments non pertinents (hors sujet)	0
<b>Cohérence</b>	B1	- Peut relier une série d'éléments en une suite linéaire de points qui s'enchaînent (logique du déroulement de la pensée)	3
	A2	- peut relier des groupes de mots avec des connecteurs simples tels que « et », « mais » et « parce que »	2
	A1	- peut relier des mots ou groupes de mots avec des connecteurs très élémentaires tels que « et » ou « alors »	1
		- les mots et expressions ne sont pas reliés entre eux	0
<b>Correction phonétique</b> (échelle de Harris)		Peu de trace d'accent étranger Toujours intelligible malgré accent spécifique	3
		Difficultés de prononciation qui exigent une attention soutenue et conduisent quelques fois au malentendu	2
		Très difficile à comprendre à cause de la prononciation	1
		Difficultés de prononciation qui rendent le discours pratiquement inintelligible	0

Nom :

Grille d'évaluation Expression Orale – 2de – Niveau visé B1

**S'adresser à un auditoire**

Peut faire un exposé simple et direct, préparé, sur un sujet familier.

Respect de la consigne	0	0,5	1	1,5	2		
Performance	0	0,5	1	1,5	2	2,5	3
Morphosyntaxe	0	0,5	1	1,5	2	2,5	3
Lexique approprié	0	0,5	1	1,5	2	2,5	3
Pertinence	0	0,5	1	1,5	2	2,5	3
Cohérence	0	1	2	3			
Correction phonétique	0	1	2	3			

Note :

---

**Vocabulaire vu en cours :**

- Stereotypes:

Women: nice body curves / perfect make-up / thin / attractive / pretty / bombshell / big breasts / short / long hair / emotional / sensitive / pink / take care of the husband and children

Men: strong / intelligent / brutal / aggressive / muscular / tall / short hair / blue / like cars / watch TV / drink beer / take care of the garden

- (In)equality / gender / male vs. female / man vs. men + woman vs. women / domestic behaviors / personality traits / sexual harassment

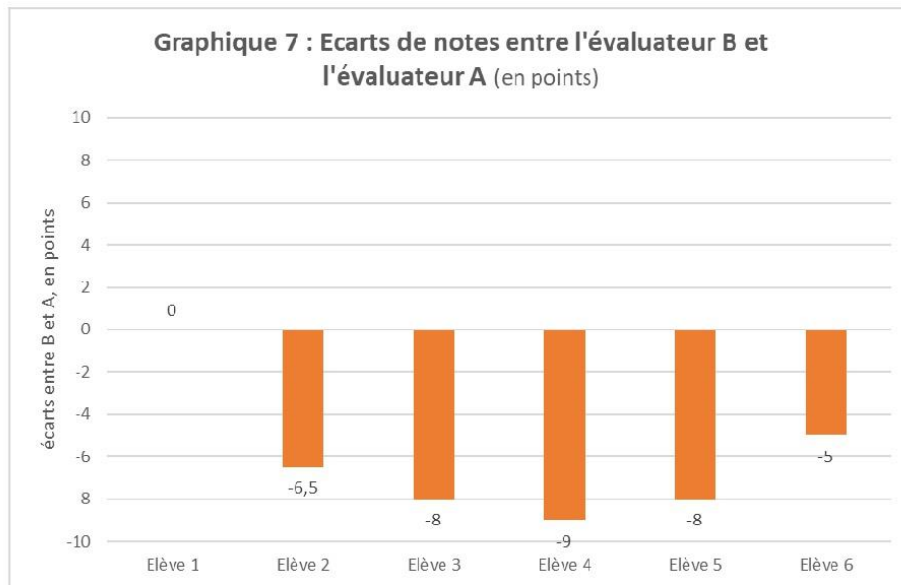
- The representation of women in:

Ads / advertisement / commercial / media / superhero movies / TV series

Women → less covered / almost naked / object of desire

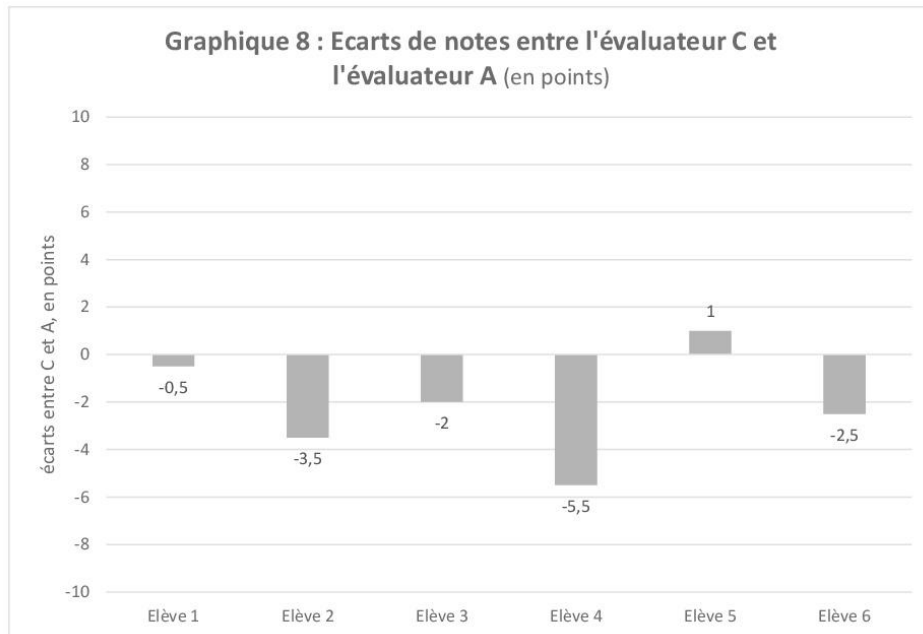
Men → muscular / brave / more covered

**Annexe 2 : Graphique des écarts de notes entre l'évaluateur B et l'évaluateur A (Graphique 7)**

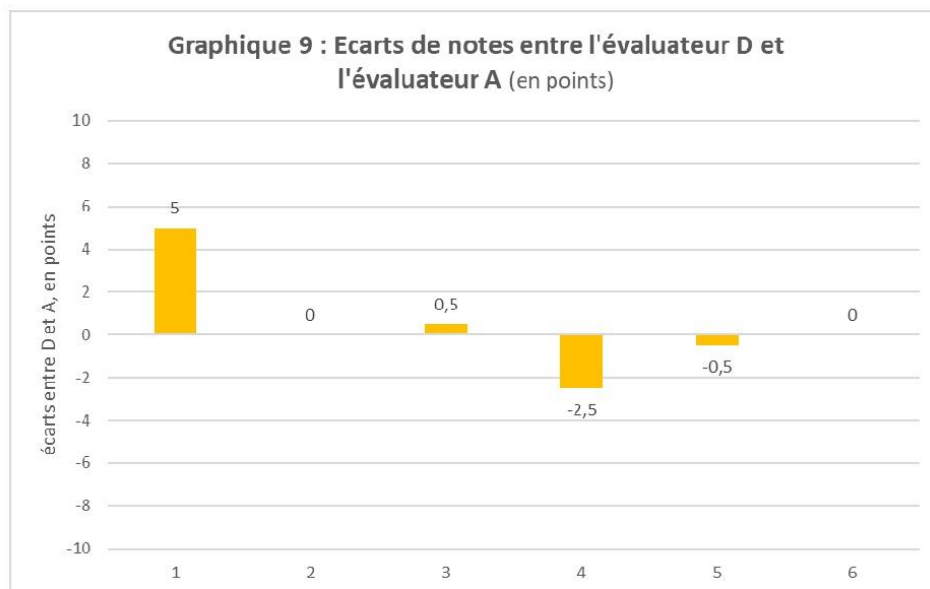




**Annexe 3 : Graphique des écarts de notes entre l'évaluateur C et l'évaluateur A (Graphique 8)**



**Annexe 4 :** Graphique des écarts de notes entre l'évaluateur D et l'évaluateur A (Graphique 9)



**Année universitaire 2017-2018**

**Master 2 Métiers de l'enseignement, de l'éducation et de la formation**  
**Mention Second degré**  
**Parcours : Anglais**

**Titre du mémoire : La subjectivité de l'évaluation en production orale – Le cas des langues vivantes**

**Auteur : Chloé PEDELABAT-LARTIGAU**

**Résumé :**

L'acte d'évaluation en langues vivantes est un exercice complexe. Les évaluateurs sont en effet confrontés à la question que tout enseignant se pose : comment évaluer ses élèves de la manière la plus juste possible ? Les travaux de la recherche montrent que l'évaluation selon une normativité humaine n'échappe pas à la question de la subjectivité et ses impacts sur les apprenants sont nombreux. Il apparaît dès lors comme indispensable de répondre à cette question par des outils adéquats permettant d'atténuer, si ce n'est gommer entièrement, cette part de subjectivité. Ce mémoire s'attache plus précisément à la question de la subjectivité de l'évaluation dans la production orale qui est particulièrement sensible dans l'enseignement des langues vivantes. L'oral y occupe une place de plus en plus importante et, à ce titre, la question de son évaluation se pose avec d'autant plus de force. L'objectif de cette étude est donc notamment de mettre en avant l'influence de la communication non verbale dans l'évaluation de la production orale.

**Mots clés : évaluation, subjectivité, communication non verbale, CECRL, langues vivantes**

**Summary :**

Language assessment is a complex activity. When evaluating their students, teachers ask themselves: how can I grade my students as fairly and equitably as possible? Research has shown that evaluating with a human normativity cannot bypass the issue of subjectivity. Furthermore, its impacts on learners are numerous. It is therefore necessary to tackle the problem of biased evaluation with suitable tools allowing for a higher rate of objectivity. This thesis concentrates particularly on the question of subjectivity in oral production. Oral plays an important role in modern language teaching and, for this reason, the question of its evaluation is also more and more important. The purpose of this study is to put forth the

influence of nonverbal communication in the evaluation of oral production.

**Key words: evaluation, subjectivity, nonverbal communication, CEFR, modern languages**