

# Mise en équivalence des versions papier-crayon et électronique d'un test de français langue étrangère et mise en perspective des résultats par une procédure d'autoévaluation

DESROCHES Franck, *Responsable du TEF*, Direction des Relations Internationales de l'Enseignement, Chambre de Commerce et d'Industrie de Paris, Paris, France, [fdesroches@ccip.fr](mailto:fdesroches@ccip.fr)

CASANOVA Dominique, *Responsable pédagogique*, Direction des Relations Internationales de l'Enseignement, Chambre de Commerce et d'Industrie de Paris, Paris, France, [dcasanova@ccip.fr](mailto:dcasanova@ccip.fr)

CRENDAL Alexandra, *Responsable pédagogique*, Direction des Relations Internationales de l'Enseignement, Chambre de Commerce et d'Industrie de Paris, Paris, France, [acrendal@ccip.fr](mailto:acrendal@ccip.fr)

RENAUD François, *Responsable pédagogique*, Direction des Relations Internationales de l'Enseignement, Chambre de Commerce et d'Industrie de Paris, Paris, France, [frenaud@ccip.fr](mailto:frenaud@ccip.fr)

MAHY Christian, *Responsable Informatique*, Ecoles Grégoire Ferrandi, Chambre de Commerce et d'Industrie de Paris, Paris, France, [cmahy@ccip.fr](mailto:cmahy@ccip.fr)

DEMEUSE Marc, *Professeur*, Université de Mons-Hainaut, Mons, Belgique, [marc.demeuse@umh.ac.be](mailto:marc.demeuse@umh.ac.be)

**Mots clés** : évaluation ; psychométrie ; français langue étrangère (FLE) ; outil informatique

---

## Résumé

Dans le contexte actuel, les outils d'évaluation doivent notamment s'adapter aux contraintes liées à l'instantanéité de la communication et de la transmission des informations ainsi qu'à l'intensification de la circulation des biens et des personnes. Pour ces raisons, l'attractivité d'une version électronique d'un test établi originellement dans une version papier-crayon est forte, tant pour les candidats que pour les institutions. En dehors des problèmes purement technologiques, l'adaptation d'un test d'un média à l'autre suppose des modifications pédagogiques et administratives. Ces modifications doivent être accomplies avec prudence, de manière à ne pas altérer la nature même de l'épreuve et la comparabilité des résultats dans le temps, quelle que soit le type de test, papier-crayon ou électronique, auquel s'est soumis le candidat.

Cette présentation s'inscrit dans les nouvelles recherches qui se sont développées dans le domaine de l'évaluation du français langue étrangère, où l'apprenant est davantage acteur de son apprentissage. Ces nouvelles approches se conjuguent pour permettre l'évaluation de nouveaux processus cognitifs dans des conditions aussi proches que possible des situations réelles de communication.

Depuis 1998, la base de données mise au point par l'équipe de la *Chambre de Commerce et d'Industrie de Paris* (CCIP) dans le cadre du développement du *Test d'Évaluation de Français* (TEF) permet un suivi précis des candidats à travers les différentes épreuves auxquelles ils se sont soumis. La documentation de chacune des épreuves (compréhension orale et écrite, expression orale et écrite) permet une analyse fine des résultats, y compris jusqu'à la réponse à un item particulier. Cette base de données est enrichie chaque année par les pré-tests et les épreuves en situation réelle qui sont présentées par les candidats. La mise au point d'une forme électronique de testing, à côté de la version papier-crayon initiale, exploite les données de cette base riche de plusieurs dizaines de milliers de résultats. La présente communication est fondée sur cette mise au point et sur l'étude de son équivalence par rapport à la version papier-crayon. Elle aborde également quelques problèmes liés à la sécurisation des données (questions, jeux d'épreuves, communication des résultats) qui exigent différentes étapes de développement informatique.

Le lien entre épreuves à correction objective (les épreuves de compréhension orale et écrite du TEF) et autoévaluation sur la base des échelles DIALANG est aussi mis à l'épreuve empiriquement. Les premiers résultats semblent offrir plus de questions que de réponses, notamment à propos des capacités d'autoévaluation des candidats face à un référentiel non explicité et à l'usage d'un outil dont le pouvoir de discrimination et la consistance interne, du moins sur la base d'un échantillon réduit et homogène, ne sont pas rencontrés.

## Introduction

Dans le contexte actuel, les outils d'évaluation doivent, de manière à répondre aux besoins des candidats et des institutions, s'adapter aux contraintes liées à l'instantanéité de la communication et de la transmission des informations ainsi qu'à l'intensification de la circulation des biens et des personnes. Ils doivent, en outre, pouvoir tenir compte de contraintes économiques et parfois d'autres contraintes fort inattendues, comme par exemple, l'apparition et le développement d'une épidémie comme le SRAS (Syndrome respiratoire aigu sévère), qui limite les échanges de documents classiques pour des raisons évidentes de sécurité sanitaire alors même que des candidats étrangers doivent préparer leur dossier d'admission dans des universités françaises.

Pour les raisons évoquées précédemment, l'attractivité d'une version électronique est forte et, en dehors des problèmes purement technologiques, l'adaptation du test d'un média à l'autre suppose des modifications pédagogiques, depuis la conception des items jusqu'à l'édition des formes parallèles du test. Elle modifie également le circuit administratif (gestion des sessions de tests avec les centres agréés) et la relation candidat - outil d'évaluation, tant d'un point de vue strictement matériel (par exemple, l'utilisation d'une souris versus la transcription des réponses sur une feuille de réponse) que symbolique (par exemple, présence d'un « tiers inconnu » en la « personne » de l'ordinateur en face duquel se trouve le candidat versus l'isolement de celui-ci face à un carnet de test et la feuille de réponse).

La mise au point de la version électronique du *Test d'Évaluation de Français* (TEF) de la *Chambre de Commerce et d'Industrie de Paris* (CCIP) (2005) a nécessité la prise en compte de ces nombreux paramètres. La présente communication aborde à la fois les problèmes scientifiques, pédagogiques et informatiques posés par le passage d'une version papier bien documentée et très largement utilisée à une adaptation en ligne du même produit.

## 1. Adaptation technique

Le e-tef, lancé le 10 mars 2005 par la CCIP dans les centres agréés en Chine, a été généralisé, dans un premier temps, dans un nombre limité de centres, dès le 15 septembre 2005. Cette version, entièrement développée au sein de ses services informatiques, permet de raccourcir les délais de passation, faciliter l'administration et assurer une sécurisation accrue des épreuves. Mais la standardisation liée, de manière générale, à la passation d'un test, implique de s'intéresser également à l'environnement qui est offert au candidat lorsqu'il passe le test : contraintes logistiques, acoustiques, distales (« *données environnementales : institutionnelles, culturelles, politiques, économiques, géographiques, écologiques* », et proximales (« *...agencement spatial, organisation temporelle, et organisationnelle* »). Ces contraintes ont des répercussions sur le fonctionnement des centres agréés. Ceux-ci sont confrontés à de nouvelles modalités d'organisation qui ne sont pas sans conséquences, par exemple, sur la sécurisation du test depuis la réception des épreuves jusqu'à la communication des résultats, mais aussi sur le niveau de formation des surveillants. Ainsi, dans la version classique du TEF, les surveillants distribuent, au début de chaque épreuve, les livrets correspondants, qu'ils récupèrent à l'issue de l'épreuve. Ces instants, au cours desquels il y a du mouvement, de la circulation de livrets et du bruit font courir le risque de voir des candidats tenter de s'échanger des réponses, malgré la surveillance rapprochée. Par contre, dans la version électronique, lorsqu'il débute le test, le candidat se voit proposer l'épreuve de compréhension écrite (CE). Après validation (par le candidat ou à l'issue du temps maximal imparti<sup>1</sup>), l'épreuve de compréhension orale (CO) est lancée automatiquement. La fin du temps réglementaire imparti à cette épreuve provoque le lancement de l'épreuve lexique/structure (LS). Lors de la validation d'une épreuve (par le candidat ou à l'issue du temps réglementaire imparti), les réponses sont communiquées automatiquement au site Internet e-tef<sup>2</sup>. Il n'y a donc pas de temps de latence entre deux épreuves, si ce n'est pour le candidat la possibilité de relire les consignes générales pendant 5 minutes. De plus, des jeux de questions différents sont présentés à des candidats qui sont physiquement dans la même salle de test, les empêchant ainsi totalement de communiquer, ce qui n'était pas possible dans la version classique du TEF, à cause de la compréhension orale.

L'une des différences essentielles, pour le candidat, réside dans la gestion du temps rendue beaucoup plus autonome par rapport à la version papier. En effet, la durée du test papier-crayon, pour les épreuves de compréhension écrite, de compréhension orale et de lexique/structure, est identique pour tous les candidats et s'étale sur deux heures dix minutes. Quand un candidat a fini une épreuve, il doit attendre le lancement collectif de l'épreuve suivante, ce qui est parfois un peu long pour les candidats les plus faibles ou les plus forts, la volonté des concepteurs étant de proposer un test de puissance, aussi peu contaminé que possible par la composante vitesse. Dans la version électronique, le candidat a, par contre, la possibilité de valider les épreuves de compréhension écrite (CE) et de lexique/structure (LS) avant la fin du temps imparti. Pour la compréhension orale (CO), la validation s'effectue automatiquement quelques secondes après l'écoute du dernier enregistrement car cette partie s'effectue en « temps bloqué », suivant en cela l'écoute des messages, sans possibilité, comme dans la vie réelle, de procéder à des réécoutes successives.

<sup>1</sup> Ce temps est identique à celui de la version papier-crayon.

<sup>2</sup> Cela permet également de s'assurer du fait que l'épreuve est réalisée en temps réel, dans le centre, ce qui diminue encore les possibilités de fraude, tout comme la suppression de l'expédition, même sécurisée, des carnets de test. Pour contrôler les possibilités de fraude informatique par interception des flux sur le réseau, un protocole sécurisé (SSL) d'échange des données a été mis en œuvre. Ne circule sur le réseau qu'une partie de l'information liée à la question en cours. L'interception éventuelle de cette information, même décryptée, ne permet donc pas à de reconstituer la question dans son intégralité.

L'ergonomie de la version électronique améliore aussi la situation du candidat en limitant les possibilités d'erreur. Par rapport à la version papier-crayon, le candidat ne doit donc plus saisir ses réponses sur une feuille distincte destinée à la lecture optique, mais simplement cocher sur l'écran, à l'aide de la souris, l'une des possibilités offertes pour la seule question à laquelle il répond. Il passe ensuite simplement à la question suivante. Les modifications successives sont aussi simplifiées – il suffit de cocher une nouvelle réponse – par rapport à la version classique où il faut recourir à une ligne de repentir sur la feuille optique et se limiter à une seule correction ou recourir à du correcteur liquide. Cette amélioration de l'ergonomie se marque principalement dans la simplification de la synchronisation du candidat dans l'épreuve de compréhension orale (CO). Alors que, dans la version classique, le candidat doit rester synchronisé avec la bande son diffusée collectivement tout en lisant les consignes ou les illustrations dans le carnet de test et en reportant ses réponses sur la feuille de réponse, dans la version électronique, l'affichage des questions est synchronisé à la bande son et le candidat dispose de toute l'information visuelle à travers un seul canal, l'écran de l'ordinateur, sur lequel il enregistre ses réponses. Par ailleurs, la diffusion des éléments sonores s'effectue à travers un casque audio, ce qui améliore encore le confort d'écoute et assure à tous les candidats une situation comparable, quelque soit le nombre de candidats et leur position dans la salle de test.

Les concepteurs et, de manière générale, l'ensemble de l'équipe pédagogique, ont également été confrontés à des adaptations rendues nécessaires par la version électronique de l'outil, tant au niveau de la présentation et de l'organisation iconique que de certaines modalités de réponse ou de consigne. La comparaison des figures 1 et 2 met en évidence l'amélioration de l'ergonomie du test dans sa version électronique. La figure 1 présente 5 questions de la version papier-crayon et renvoie à la feuille de réponse pour indiquer chacune des solutions choisies. La figure 2 présente la première de ces questions dans sa forme électronique. Le texte est présenté dans une fenêtre et les propositions – il faut identifier une forme erronée parmi quatre extraits, soulignés dans le texte – figurent sous la fenêtre. Pour les cinq questions, le même texte est présenté successivement, mais les parties soulignées sont celles qui correspondent à la question en cours. C'est le candidat qui décide de passer d'une question à l'autre en sélectionnant le bouton « suivant », qu'il décide de répondre ou de s'abstenir.

## SECTION D

## QUESTIONS 146 À 150

◆ Dans certaines phrases du texte suivant, des parties (A, B, C ou D) ont été soulignées. L'une de ces parties est incorrecte.

→ Cochez sur la fiche optique la réponse correspondant à la partie incorrecte.

<p><b>EXEMPLE</b></p> <p>Vous devez cocher " D " car la formulation correcte est " se compte ".</p>	<p style="text-align: center;"><u>La chaussure</u></p> <p>Se protéger <u>les pieds</u> est un souci qui ne <u>date pas</u> d'hier. Mais de l'<u>objet chaussant</u> à la chaussure, le pas franchi <u>ce compte</u> en millénaires.</p> <p style="text-align: center;">A B C D</p>
<p><b>QUESTION 146</b></p>	<p>Les peuples de l'Antiquité portaient <u>surtout</u> des sandales <u>a semelles</u> de cuir ou de bois, tenues par des <u>lanières croisées</u>. Toutes sortes de modèles <u>furent imaginés</u> par les Grecs, puis les Romains.</p> <p style="text-align: center;">A B C D</p>
<p><b>QUESTION 147</b></p>	<p>L'époque médiévale voit les sandales <u>remplacées</u> par des chaussures fermées par <u>des lacets</u>. Au X<sup>e</sup> siècle, la chaussure <u>au sens</u> moderne du terme apparaît, <u>façonnée</u> par cuir.</p> <p style="text-align: center;">A B C D</p>
<p><b>QUESTION 148</b></p>	<p>Au XV<sup>e</sup>me siècle, avec la <u>mode italienne</u>, les chaussures <u>s'affinant</u> et prennent des formes élégantes. Mais la chaussure <u>est restée</u>, jusqu'au XIX<sup>e</sup> siècle, réservée <u>aux riches</u>.</p> <p style="text-align: center;">A B C D</p>
<p><b>QUESTION 149</b></p>	<p>La Première guerre mondiale joue <u>un rôle</u> déterminant dans le <u>développement</u> industriel de la chaussure. Dès 1919, la mode à prix abordable <u>s'affiche</u> dans les magasins <u>des fabricants</u>.</p> <p style="text-align: center;">A B C D</p>
<p><b>QUESTION 150</b></p>	<p>Aujourd'hui, la mode <u>ressemble</u> avec une superposition <u>de tendances</u> et la chaussure présente une <u>grande variété</u>, tant du point de vue des formes que des <u>matières utilisées</u>.</p> <p style="text-align: center;">A B C D</p>

Figure 1 – Présentation d'une unité extraite de l'épreuve lexique/structure (LS) dans sa forme papier-crayon (TEF).

The screenshot shows the 'etest' logo in the top left and the title 'Test d'unités' in the top right. The main content is a text passage titled 'La chaussure' with the following text:

Se protéger les pieds est un souci qui ne date pas d'hier. Mais de l'objet chaussant à la chaussure, le pas franchi ce compte en millénaires.

Les peuples de l'Antiquité portaient surtout des sandales à semelles de cuir ou de bois, tenues par des lanières croisées.

Toutes sortes de modèles furent imaginés par les Grecs, puis les Romains.

L'époque médiévale voit les sandales remplacées par des chaussures fermées par des lacets. Au Xe siècle, la chaussure au sens moderne du terme apparaît, façonnée par cuir.

Au XVIème siècle, avec la mode italienne, les chaussures s'affinent et prennent des formes élégantes. Mais la chaussure est restée, jusqu'au XIXe siècle, réservée aux riches.

La Première guerre mondiale joue un rôle déterminant dans le développement industriel de la chaussure. Dès 1919, la mode à

Below the text is a question titled 'QUESTION 1' with the subtext 'Partie incorrecte'. The question has five radio button options:

- surtout
- à semelles
- lanières croisées
- furent imaginés
- Je ne sais pas.

In the bottom right corner, there is a blue circular button with a play icon and the text 'Suivant'.

Figure 2 – Présentation d'une unité extraite de l'épreuve lexique/structure (LS) dans sa forme électronique (e-tesf)

D'autres facilités techniques, non encore exploitées jusqu'ici, pourront être mises en œuvre grâce à l'informatisation du test, comme par exemple, l'enregistrement des temps de réponse, de manière à procéder à des analyses complémentaires.

## 2. Analyse des résultats

Depuis 1998, la base de données qui a été mise au point permet un suivi précis des candidats à travers les différentes épreuves auxquelles ils se sont soumis. La documentation de chacune des épreuves (compréhension orale et écrite, lexique et structure, expression orale et écrite) permet une analyse fine des résultats, y compris jusqu'à la réponse à un item particulier, comme l'illustre une autre communication (Demeuse *et al.*, 2005) faite à ce colloque.

Cette base de données est enrichie par les pré-tests et les épreuves en situation réelle qui sont présentées par les candidats. Par ailleurs, d'autres informations sont collectées et des tests spécifiques sont utilisés, comme dans le cas du passage de la version papier à la version électronique du test (Test d'Évaluation de Français). Les analyses qui sont présentées reposent sur l'exploitation de cette base de données et des données complémentaires, comme les auto-évaluations des candidats. L'exposé s'appuie sur des enquêtes et des résultats de candidats ayant passé la version papier et la version électronique de deux formes parallèles lors de pré-test.

Pour rappel (Demeuse *et al.*, 2005), les corrélations entre deux formes parallèles papier-crayon, soumises aux mêmes candidats, à quelques jours d'écart, calculée sur la base de 5 séances pré-tests regroupant chacune de 57 à 81 candidats (avant tout ajustement des formes et suppression/remplacement d'items), s'élèvent à 0,88. Le même type de dispositif, impliquant cette fois une forme papier-crayon, puis une forme électronique, conduit à un résultat comparable : 0,89 et hautement significatif (à 0,0001). Cette fidélité, établie par corrélation des résultats aux deux formes parallèles, électronique d'une part, papier-crayon d'autre part, sur un nombre limité de candidats (n=22) repose également sur une situation de type expérimental, ce qui conduit sans doute à une sous-estimation. Les corrélations entre les différentes épreuves des deux formes parallèles s'élèvent respectivement à 0,81 pour la compréhension écrite, 0,77 pour la compréhension orale et 0,75 pour le lexique/structure. Toutes ces corrélations sont hautement significatives (à 0,0001), malgré la faiblesse des effectifs.

De manière générale, en termes de points au test, les candidats semblent avantagés par la version électronique, sauf dans le domaine lexique/structure où la situation est également partagée entre ceux qui améliorent leur score et ceux qui subissent une régression. Le gain moyen observé est de 34,64 points sur 900 pour l'ensemble du test, en faveur de l'e-tef, il est de 11,36 points sur 300 à la compréhension écrite, de 20,64 points sur 360 à la compréhension orale et de seulement 2,64 points sur 240 à l'épreuve lexique/structure. En termes de niveaux<sup>3</sup>, un écart d'un seul niveau, au maximum, peut-être observé, sauf dans le cas de la compréhension orale où un candidat bénéficie d'une amélioration de deux niveaux et dans le cas du lexique structure où un candidat bénéficie d'un progrès de deux niveaux, alors qu'un autre régresse de deux niveaux.

Deux facteurs peuvent expliquer en partie ces gains :

- pour les 22 candidats le TEF papier a eu lieu avant le e-tef. Les candidats ont pu tirer profit de cette familiarisation avec la structure générale du test pour améliorer leur score. De nouvelles études sont donc actuellement entreprises de manière à contrôler un effet de l'ordre de présentation qui viendrait s'ajouter à un éventuel effet lié à la nature du support.
- Pour la moitié des candidats, l'enjeu était plus fort pour le second test, puisqu'il conduisait à la délivrance d'une attestation, alors que la première version ne présentait aucun enjeu véritable.

Ce qui est relativement rassurant, c'est que la suspicion d'un gain, en compréhension orale et en lexique/structure, gain lié à l'amélioration de l'ergonomie de ces deux parties dans le e-tef, s'il existe, joue en principe dans le même sens que l'ordre de présentation.

---

<sup>3</sup> Les sept niveaux du TEF et de l'e-tef sont en étroite correspondance avec les 6 niveaux du *Cadre européen commun de référence pour les langues* (CECR) du Conseil de l'Europe (2000). Les niveaux du TEF sont par ailleurs indexés sur les Standards Linguistiques Canadiens (Demeuse *et al.*, à paraître ; Crendal, 2005).

### 3. Quel sens donner aux résultats pour le candidat : tentative de confrontation hétéro-évaluation / auto-évaluation ?

A côté de l'intérêt technique de la mise en équivalence de deux types de présentations, il est aussi intéressant d'examiner la manière dont les résultats d'un test peuvent être compris par les candidats qui s'y soumettent. De manière à répondre à cette question, nous avons présenté à 36 candidats hispanophones, inscrits dans un centre de langue de Vichy, avant une passation de e-tef, la grille d'auto-évaluation mise au point par DIALANG (tableaux 3 et 4)<sup>4</sup>. Cette grille, la version courte de l'instrument, soit 36 items relatifs à la compréhension orale et écrite pour lesquels le candidat doit indiquer si oui ou non il est capable de réaliser la performance mentionnée, a été administrée dans sa version espagnole, de manière à éviter tout problème d'interprétation de la part des candidats.

**Tableau 1** – Extrait de l'échelle d'auto-évaluation linguistique de DIALANG dans le domaine de la compréhension écrite. La première colonne indique le niveau du CECR associé à la compétence décrite dans la deuxième colonne

A1	Je peux comprendre l'idée générale de textes simples donnant des informations et des descriptions courtes et simples, surtout s'ils contiennent des images qui facilitent la compréhension.
A1	Je peux comprendre des instructions simples et courtes surtout si elles sont illustrées.
A1	Je peux comprendre des messages courts, simples, par exemple sur des cartes postales.
A2	Je peux comprendre des textes courts et simples écrits dans un langage quotidien.
A2	Je peux comprendre des lettres personnelles simples et courtes.
B1	Je peux chercher des informations spécifiques dont j'ai besoin pour accomplir une tâche dans un long texte ou dans plusieurs textes courts.
B1	Je peux suffisamment bien comprendre la description d'événements, de sentiments et de souhaits dans des lettres personnelles pour correspondre avec un ami ou une connaissance.
B1	Je peux comprendre des instructions claires pour l'utilisation d'un appareil.
B2	Je peux lire différents écrits sur ce qui m'intéresse et facilement comprendre les points essentiels.
B2	Je peux rapidement lire des longs textes complexes en sachant repérer les détails importants.
B2	Dans mon domaine, je peux acquérir des informations, des idées et des opinions de sources très spécialisées.
B2	Je peux comprendre des articles et rapports traitant de problèmes actuels dans lesquels les auteurs adoptent des positions ou des points de vue particuliers.
C1	Je peux comprendre en détail un grand nombre de longs textes complexes si je peux relire les passages difficiles.
C1	Dans des longs textes complexes traitant de la vie sociale, professionnelle et universitaire, je peux identifier des détails comprenant des attitudes et opinions qui ne sont pas explicitement décrites.
C1	Je peux comprendre différents écrits en utilisant occasionnellement un dictionnaire.
C1	Je peux comprendre en détail des longues instructions complexes sur une nouvelle machine ou procédure même en dehors de mon domaine, si je peux relire les passages difficiles.
C2	Je peux comprendre et interpréter pratiquement toutes sortes de textes écrits, que ce soient des textes abstraits, à structures complexes ou des textes littéraires avec des traits familiers ou des textes documentés.
C2	Je peux comprendre les subtilités de style et les différents niveaux de sens d'un grand nombre de longs textes complexes.

<sup>4</sup> Les échelles ont été téléchargées à partir du site DIALANG : <http://www.dialang.org/project/>.



**Tableau 2** – Extrait de l'échelle d'auto-évaluation linguistique de DIALANG dans le domaine de la compréhension orale. La première colonne indique le niveau du CECR associé à la compétence décrite dans la deuxième colonne

A1	Je peux suivre ce qui est dit à débit très lent et bien articulé avec de longues pauses qui me permettent d'en comprendre le sens.
A1	Je peux comprendre des questions et des instructions et suivre des consignes courtes et simples.
A2	Je peux suffisamment comprendre pour échanger des propos habituels sans trop d'efforts.
A2	Je peux identifier le sujet d'une discussion que j'écoute quand elle est lente et claire.
A2	Je peux suffisamment comprendre la langue pour être capable de me débrouiller dans la vie de tous les jours si cela est dit clairement et lentement.
A2	Je peux me débrouiller dans les situations simples dans un magasin, un bureau de poste ou une banque.
B1	Je peux saisir les idées principales d'une longue discussion que j'écoute s'il s'agit d'un langage est clair et standard.
B1	Je peux suivre une conversation claire bien que dans une situation réelle je doive quelquefois demander à mon interlocuteur de répéter un mot ou une expression.
B1	Je peux comprendre ce qui est dit sur la vie quotidienne ou le travail, d'une façon générale et en détail, lorsque les locuteurs parlent clairement et avec un accent que je connais.
B1	Je peux comprendre l'essentiel des programmes diffusés qui traitent de sujets connus ou intéressants pour moi lorsque le débit est relativement lent et distinct.
B2	Je peux comprendre un langage parlé standard dans des situations réelles ou des émissions diffusées concernant des sujets familiers ou non que je rencontre dans la vie personnelle, universitaire ou professionnelle. Seuls des bruits de fond extrêmes, une structure peu claire et/ ou un parler familier me posent quelques problèmes.
B2	Je peux suivre l'essentiel de conférences, cours et rapports utilisant des idées et une langue complexes.
B2	Je peux comprendre des annonces et des messages sur des sujets concrets ou abstraits quand ils sont exprimés en langage standard à un débit normal.
B2	Je peux comprendre presque tous les documents diffusés ou enregistrés en langage standard et je peux saisir l'humeur et le ton de celui qui parle.
C1	Je peux suivre une conversation animée avec des personnes de langue maternelle.
C1	Je peux assez aisément suivre la plupart des conférences, discussions et débats.
C2	Je n'ai aucune difficulté à comprendre la langue parlée quelle qu'elle soit, dite à débit rapide par un locuteur natif dans une situation réelle ou dans un document diffusé.
C2	Je peux suivre des conférences ou des exposés sur des sujets spécialisés dans un langage parlé, régional, avec une terminologie non familière.

Dans la première analyse qui est présentée ici, le niveau de chaque candidat, à partir de son autoévaluation, a été calculé de manière simple, en totalisant le nombre de réponses affirmatives et en confrontant ce score à la position occupée par chacun des six niveaux du CECR sur cette échelle. La table de correspondance est fournie ci-dessous (tableau 3). Cette approche repose sur l'idée que chaque échelle fonctionne de manière parfaite par rapport à un modèle parfait de Guttman (Alexandre, 1971 ; Matalon, 1965 ; McIver & Carmines, 1981), ce qui ne correspond malheureusement pas aux données collectées sur les 36 candidats de l'étude<sup>5</sup>.

<sup>5</sup> L'alpha de Cronbach calculé pour les deux échelles s'élève à 0,70 pour la compréhension orale, mais de seulement 0,43 pour la compréhension écrite. Nous ne disposons malheureusement pas de ces valeurs pour les échantillons qui ont permis de valider les échelles.

**Tableau 3** – Table de correspondance entre le nombre de réponses positives « oui, je peux » aux 18 items de l'échelle d'autoévaluation DIALANG et les six niveaux du CECR.

Niveau du CECR	Nombre de réponses positives à l'échelle d'autoévaluation (compréhension écrite)	Nombre de réponses positives à l'échelle d'autoévaluation (compréhension orale)
A1	De 0 à 3	De 0 à 2
A2	De 4 à 5	De 3 à 6
B1	De 6 à 8	De 7 à 10
B2	De 9 à 12	De 11 à 14
C1	De 13 à 16	De 15 à 16
C2	De 17 à 18	De 17 à 18

Cette approche ne conduit pas à des résultats très satisfaisants, comme le montrent les tableaux 4 et 5 puisque seuls 9 classements concordants, sur les 72 (36 candidats en CE et en CO), sont observés. La corrélation entre le nombre de réponses affirmatives (autoévaluation) et les résultats obtenus au TEF, si elle permet d'observer un lien significatif ( $r=0,38$  à  $0,0206$ ) dans le sens attendu dans le cas de la compréhension orale, conduit à une absence de lien entre les deux modes d'évaluation ( $r=-0,05$  à  $0,7653$ ) pour la compréhension écrite. Ces résultats particulièrement curieux doivent conduire à réexaminer la manière d'interpréter les échelles d'autoévaluation, en regard de l'évaluation par un test à réponse objective. En effet, si la cohérence de ces deux instruments est aussi faible, voire inexistante, et si cela reflète bien une différence réelle entre la manière dont se perçoivent honnêtement les candidats et la manière dont ils sont évalués de manière externe, il convient de s'interroger sur la capacité qu'auront ces candidats à exploiter les résultats du test, à les comprendre et à les accepter. Naturellement, la capacité d'autoévaluation de candidats en début de formation par rapport à leur maîtrise d'une langue étrangère et sur la base d'un référentiel externe (CECR) non explicité mérite d'être approfondie.

Une autre explication, plus technique cette fois, peut également, et de manière complémentaire, être avancée. Elle questionne l'instrument d'autoévaluation des compétences en compréhension écrite et en particulier son pouvoir de discrimination : la dispersion observée des réponses y est en effet plus faible (écart-type de 1,73, score minimum de 8, score maximum de 16) que dans le cas de la compréhension orale (écart-type de 2,38, score minimum de 4, score maximum de 17). Cela peut contribuer à sa faible consistance interne. Cette différence s'accompagne, par ailleurs, au niveau des deux épreuves du test, d'un mouvement dans le même sens, bien que d'amplitude moindre (écart-type de 46,23 points pour l'épreuve de compréhension écrite contre 48,41 points pour l'épreuve de compréhension orale). Cette approche technique n'est naturellement pas à négliger dans le cas d'échantillons de relativement petite taille relativement homogènes quant à leur composition.

**Tableau 4** – Table de correspondance entre le nombre de réponses positives « oui, je peux » aux 18 items de l'échelle d'autoévaluation DIALANG et les six niveaux du CECR pour la compréhension écrite.

		Niveau du CECR d'après l'autoévaluation					
		A1	A2	B1	B2	C1	C2
Niveau du CECR d'après le TEF	A1			1	6	8	
	A2						
	B1				6	9	
	B2				3	3	
	C1						
	C2						

**Tableau 5** – Table de correspondance entre le nombre de réponses positives « oui, je peux » aux 18 items de l'échelle d'autoévaluation DIALANG et les six niveaux du CECR pour la compréhension orale.

		Niveau du CECR d'après l'autoévaluation					
		A1	A2	B1	B2	C1	C2
Niveau du CECR d'après le TEF	A1		1	1	7		
	A2						
	B1			1	11	6	
	B2				5	1	1
	C1				2		
	C2						

## Conclusion

L'exposé s'est centré sur la possibilité de mettre en place un test de français électronique équivalent à sa version papier-crayon à présent bien documentée. Ce développement est possible à condition de respecter un certain nombre de critères (scientifique, pédagogique, informatique). La réussite d'une version électronique du TEF est intrinsèquement liée à la sécurisation des données (questions, jeux d'épreuves, communication des résultats), ce qui a donné lieu à différentes étapes de développement informatique. La mise en équivalence des deux formes de présentation, telle que nous l'avons montré ici, continue à faire l'objet d'analyses et d'un suivi permanent, notamment de manière à contrôler tous les effets parasites. Si l'amélioration de l'ergonomie du test ne semble pas avoir bouleversé les résultats obtenus sur un premier échantillon de candidats, l'analyse de données complémentaires sera très utile pour notamment expliquer les différences entre candidats, liées par exemple à une plus grande familiarité avec l'outil électronique. L'étude de la correspondance entre une auto-évaluation et une hétéro-évaluation instrumentée en langue étrangère, comme on l'a constaté, n'est pas sans poser question. La communication a mis en lumière le très grand intérêt de ce domaine, dans la mesure où on souhaite que les candidats puissent s'appropriier, en les comprenant, les résultats d'un test, mais a aussi pointé les limites de cette tentative.

Il faut encore ajouter aux réflexions entreprises lors de cette communication qu'il existe d'autres paramètres à prendre en compte. Par exemple, la décision de la CCIP de maintenir pour une période sans doute encore longue les deux versions, s'appuie sur le manque d'équipement et l'impossibilité d'assurer des liaisons fiables et stables avec certains centres agréés. La CCIP souhaite en effet pouvoir offrir le service du TEF, quel que soit la localisation du candidat, en essayant par exemple de minimiser pour lui les problèmes de transports et de déplacement, ce qui justifie le nombre élevé de centres agréés et la cohabitation des deux formules papier-crayon et électronique. Par ailleurs, la CCIP reconnaît que tous les candidats ne sont pas encore nécessairement familiarisés avec le média électronique et elle ne souhaite pas introduire des obstacles supplémentaires et sans rapport avec les compétences évaluées auprès des candidats ayant par exemple un plus faible accès à la technique, pour des raisons économiques ou de formation.

## Bibliographie

ALEXANDRE, V. (1971), *Les échelles d'attitude*. France: Editions universitaires.

CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS (2005). *Le Test d'Evaluation de Français de la Chambre de Commerce et d'Industrie de Paris*. Paris : Chambre de Commerce et d'Industrie de Paris, Centre de Langue de la Direction des relations internationales de l'Enseignement.

CONSEIL DE L'EUROPE (2000). *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*, Paris : Didier.

CRENDAL, A. (2005). « Vers la multiréférentialisation du TEF », *Points communs*, 24, 7-13.

DEMEUSE, M., CRENDAL, DESROCHES, F. , A., RENAUD, F., CASANOVA, D., ARTUS, F. (2005). « La fiabilité de l'évaluation des compétences linguistiques pour des adultes non francophones : présentation d'un protocole d'évaluation », In *Actes du 18<sup>e</sup> colloque international de l'Association pour le Développement des Méthodologies d'Evaluation en Education (ADMEE-Europe)*, Reims, 24-26 octobre 2005.

DEMEUSE, M., CRENDAL, A., DESROCHES, F., OSTER, P., RENAUD, F., & LEROUX, X. (à paraître). « L'évaluation des compétences linguistiques des adultes en français langue étrangère dans une perspective de multi-référentialisation. L'exemple du Test d'Evaluation de Français (TEF) de la Chambre de Commerce et d'Industrie de Paris (CCIP) », In *Actes du 17<sup>e</sup> colloque international de l'Association pour le Développement des Méthodologies d'Evaluation en Education (ADMEE-Europe)*, Lisbonne, 18-20 novembre 2004.

MATALON, B. (1965), *L'analyse hiérarchique*. Paris: Gauthier-Villars.

MC IVER, J.P. & CARMINES, E.G. (1981), *Unidimensional scaling*. London: Sage Publications, Coll. Sage University papers, Quantitative Applications in the social Sciences, N°24.