

## **Élaboration d'une version électronique équivalente à la version papier-crayon d'un test de français langue étrangère à enjeux critiques**

**Dominique Casanova, Alexandra Crendal, Alexandre Holle<sup>1</sup>**

Chambre de commerce et d'industrie de Paris, Direction des relations internationales de l'enseignement, Centre de langue française

**Marc Demeuse**

Université de Mons

### **Introduction**

Dans le contexte actuel, les outils d'évaluation des apprentissages doivent 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, de manière à répondre aux besoins des candidats et des institutions. Ils doivent, en outre, pouvoir tenir compte de contraintes économiques et parfois d'autres contraintes plus inattendues, par exemple, les contraintes épidémiques, qui limitent les échanges de documents classiques pour des raisons de sécurité sanitaire.

Ainsi, pour les raisons évoquées précédemment, l'attractivité d'une version électronique d'un outil d'évaluation des apprentissages est forte et, en dehors des problèmes purement technologiques, l'adaptation d'une modalité de passation à une autre peut entraîner des modifications pédagogiques, depuis la conception des items jusqu'à l'édition des formes parallèles de l'outil. Elle modifie également le circuit administratif (gestion des sessions de test avec des centres agréés) et la relation candidat – outil d'évaluation, tant d'un point de vue

---

1. Ce travail n'aurait pas été possible sans l'engagement continu des autres membres du Centre de langue française travaillant sur le Test d'évaluation de français.

strictement matériel (par exemple, l'utilisation d'une « souris » versus la transcription des réponses sur une feuille de réponses) 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 en face d'un carnet de test et de la feuille de réponses).

Dans certains cas, la transition de la modalité de passation papier-crayon d'un outil d'évaluation des apprentissages vers la modalité électronique est l'occasion de mettre en œuvre une version adaptative ou de proposer de nouveaux formats d'items. Ce type de modalité électronique n'est pas développé ici, mais le lecteur intéressé pourra se référer à l'ouvrage publié par Blais en 2009, où plusieurs chapitres montrent l'intérêt des tests adaptatifs au format électronique, où les items proposés à un candidat sont choisis en fonction des réussites et des échecs aux items précédents, aboutissant ainsi à une épreuve « sur mesure », constituée d'un ensemble généralement limité d'items, dont la difficulté est proche du niveau de compétence du candidat.

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) (2007) a nécessité la prise en compte de nombreux paramètres évoqués ci-dessus. Le présent chapitre aborde à la fois les problèmes pédagogiques et informatiques posés par le passage d'une version papier-crayon bien documentée et très largement utilisée à une adaptation en ligne du même produit. Après une brève revue de la littérature concernant la mise en équivalence de versions papier-crayon et électroniques de tests, il précise les conditions dans lesquelles a été élaborée la version électronique du TEF, les choix opérés pour sa mise au point et les expérimentations entreprises pour valider l'équivalence de cette version électronique avec la version papier-crayon du même test.

## **1 Du papier-crayon à l'ordinateur**

### *La comparabilité des résultats*

Il existe de nombreux cas d'adaptation électronique de tests conçus initialement dans une version papier-crayon, et plusieurs chapitres de Blais (2009) font référence à des études de comparaison entre deux versions d'un même test, dont la méta-analyse de Mead et Drasgow (1993). Dans le domaine des langues, les versions électroniques des tests ont commencé à être élaborées à la fin des années 1990 et ont connu un nouvel essor avec l'avènement d'Internet.

Dans son rapport de recherche, Bennet (2003) cite différentes études qui montrent que, pour des tests de puissance (traduction de *power test*), on constate une bonne corrélation des résultats et des différences très faibles de moyenne et d'erreur de mesure entre les versions électronique et papier-crayon dans le cas de populations adultes à qui on a soumis des questionnaires à choix multiples. Les études de comparaison évoquées par Burton et collab. (2009) et Houssemand et collab. (2009) confirment que l'on peut obtenir des résultats équivalents avec les deux versions semblables d'un même test dans le cas de questionnaires à choix multiples:

Nous pouvons donc conclure que les passations par ordinateur sont largement équivalentes aux passations papier-crayon pour tout public, du moment qu'on a rendu la version informatisée semblable à la version papier (Houssemand et collab. 2009).

En revanche, Bennett (2003) fait état de différentes études qui montrent que des variations plus importantes peuvent être constatées dans le cas d'épreuves de production écrite, dont une des sources supposées est la familiarité d'utilisation d'un clavier. La banalisation de l'ordinateur est en effet récente et on peut constater un facteur générationnel dans son utilisation. Laurier et Diarra (2009) relatent des expérimentations qui montrent que les élèves habitués à utiliser l'ordinateur pour rédiger leurs écrits ont de meilleures notes lorsqu'ils sont soumis à des épreuves utilisant le traitement de texte que dans une évaluation en mode papier-crayon.

Une autre source de variation possible pour les épreuves d'expression écrite est liée aux différences de format de présentation des productions. Plusieurs études montrent que les différences de notation sont très faibles entre une copie manuscrite et des copies numérisées. En revanche, on constate une tendance à attribuer un score plus élevé à une copie manuscrite qu'à son équivalent saisi sur ordinateur. Selon Russell et Tao (2004), les correcteurs sont sensibles au format de présentation des productions : les erreurs semblent plus faciles à percevoir sur un texte dactylographié et le niveau d'exigence plus élevé, alors que les correcteurs ont tendance à montrer plus d'empathie envers les candidats lorsque les textes sont manuscrits et que la moindre lisibilité des textes manuscrits peut rendre des erreurs plus difficiles à déceler. Le lecteur trouvera une présentation détaillée de la comparabilité entre ces modalités de production pour les épreuves d'écriture dans Laurier et Diarra (2009) de même que dans le présent ouvrage au chapitre 13, de Godin, Blais et Laurier.

#### *Standardisation des caractéristiques de présentation des items*

Bennett (2003) évoque le fait que, en format électronique, des différences de caractéristiques de présentation des items à l'écran peuvent apparaître d'un ordinateur à

l'autre, ce qui peut se traduire par un manque de standardisation et donc un manque d'équité envers certains candidats.

Bennet recense comme principales sources de variation le débit Internet, qui peut rendre plus long le passage d'une question à une autre, la résolution de l'écran et la taille des fontes, qui entraînent un confort de lecture plus ou moins satisfaisant selon l'ordinateur sur lequel le candidat passe le test et la nécessité de recourir ou non à des barres de défilement. Il se réfère pour cela à l'étude de Bridgeman, Lennon et Jackenthal (2003), qui conclut notamment qu'une approche prudente doit être adoptée pour la délivrance informatique de tests à enjeux critiques afin que le recours aux barres de défilement soit semblable d'un ordinateur à l'autre.

Donc, en plus de veiller à la comparaison de variables globales comme les scores aux épreuves selon les différentes modalités, les concepteurs de tests doivent s'attacher à garantir la standardisation des épreuves électroniques, d'une part, et à ce que les caractéristiques de présentation de certains items ne les rendent pas sensiblement plus faciles ou plus difficiles d'un média à l'autre (papier-crayon ou écran-clavier), d'autre part.

#### *Différences dans la passation des épreuves*

Bennet mentionne également d'autres différences qui peuvent être introduites par le changement du format du test, même lorsque la version électronique conserve une organisation des items identique à celle du format papier-crayon.

Tout d'abord, la réponse aux items de la version électronique peut nécessiter, de la part du candidat, une manipulation plus ou moins complexe de la souris, pour effectuer des

actions comme le pointage, la sélection (clic) et le glisser-déposer. Dans une étude ayant précédé la mise en place de la première version électronique du TOEFL® (Test of English as a Foreign Language™, ETS), Taylor et collab. (1999) montrent qu'aucune différence de performance au TOEFL CBT ne peut être constatée entre un groupe de sujets familiers avec l'informatique et un groupe de sujets non familiers avec l'informatique, lorsque les sujets ont profité, avant la passation du test assisté par ordinateur, d'une séance de tutoriel. Ensuite, alors que dans la version papier-crayon il est usuel que les candidats attendent que le temps imparti pour une épreuve du test soit écoulé avant de pouvoir commencer l'épreuve suivante, l'administration électronique peut leur permettre de passer d'une épreuve à l'autre dès qu'ils s'estiment prêts. Enfin, du fait que la lecture à l'écran de texte longs peut nécessiter plus de temps que leur lecture sur support papier, le test électronique peut involontairement introduire une contrainte de vitesse plus forte que le test papier-crayon

## **2 Contexte d'élaboration d'une version électronique**

### *Le Test d'Évaluation de Français de la Chambre de commerce et d'industrie de Paris*

Jusqu'à la fin des années 1990, il n'existait en France que des *certifications* pour l'évaluation des compétences en langue française, qui validaient « un niveau global de compétences langagières acquises lors d'un apprentissage » (Noël-Jothy et Sampsonis, 2006).

La CCIP a créé le TEF en 1998 pour offrir un instrument de référence internationale destiné à mesurer de façon précise, objective et fiable le niveau en langue française des personnes dont la langue maternelle n'est pas le français. Test à enjeux critiques, le TEF bénéficie notamment de la reconnaissance du ministère français de l'Éducation nationale, pour l'évaluation en langue française des étudiants étrangers candidats à une première inscription

en premier cycle d'études universitaires, du gouvernement fédéral canadien (Citoyenneté et Immigration Canada) pour l'évaluation linguistique des candidats à l'immigration vers ce pays dans la catégorie des travailleurs qualifiés ou la catégorie de l'expérience canadienne, et du ministère de l'Immigration et des Communautés culturelles du Québec, dans une version adaptée (le TEFAQ), pour les procédures d'immigration vers le Québec.

La mise en œuvre de la version papier-crayon du TEF nécessite plusieurs envois postaux. En amont d'une session, le service de gestion administrative adresse le matériel de test au centre de passation. À l'issue de la passation, le centre expédie au service administratif du TEF, pour traitement des réponses, les documents de la session, les fiches optiques de réponses et les productions écrites des candidats, d'une part, et le matériel de test ayant servi à l'organisation de la session, d'autre part. La durée de ces échanges postaux et le temps nécessaire à la préparation des colis dans le cas de sessions comportant un grand nombre de candidats se traduisent par des délais d'organisation des sessions et de délivrance des résultats parfois perçus comme excessifs par des candidats qui ont un besoin rapide d'évaluation de leur niveau de français, notamment dans le cas de la constitution d'un dossier de candidature. Enfin, les mesures mises en place dans le cadre du SRAS (Syndrome respiratoire aigu sévère) en 2003 et 2004, qui limitaient les échanges de documents classiques pour des raisons évidentes de sécurité alors même que des candidats étrangers devaient préparer leur dossier d'admission dans des universités françaises ont convaincu la CCIP de la nécessité de développer une version électronique du TEF, qui a été expérimentée au début de l'année 2005 auprès des centres TEF en Chine avant d'être mise à la disposition de l'ensemble des centres agréés.

Parce qu'elle permet la génération immédiate des résultats aux épreuves obligatoires (compréhension écrite, compréhension orale et lexiques et structure), conçues sous la forme de questionnaires à choix multiples, la version électronique réduit la nécessité d'échanges postaux et offre ainsi aux candidats la possibilité d'une évaluation rapide de leur niveau en langue française.

### *Le choix du format électronique*

Il ne s'agissait pas, pour la CCIP, de remplacer un test en version papier-crayon par un test en version électronique mais bien de faire coexister les deux versions du TEF. En effet, le TEF est proposé par plus de 450 centres agréés répartis dans 100 pays et certains centres ne disposent pas de l'équipement informatique approprié pour l'organisation de la version électronique. Les candidats n'ont donc pas toujours le choix de la modalité de passation du test.

Par souci d'équité, la CCIP a donc choisi de minimiser les différences entre les versions électroniques et papier-crayon, et rejeté l'option d'un test adaptatif ou le choix de nouveaux formats d'items. La version électronique a par ailleurs été limitée aux épreuves obligatoires, sous la forme de questions à choix multiple, et les actions à entreprendre par les candidats lors de la passation du test se résument à des manipulations simples de la souris d'ordinateur (pointer, sélectionner, faire défiler).

Les candidats souhaitant également être évalués en expression écrite et en expression orale peuvent se soumettre à ces deux épreuves qui ont été conservées dans leur format original. Les copies de l'épreuve d'expression écrite font l'objet d'une double correction à la Chambre de commerce et d'industrie de Paris (2004) et l'épreuve d'expression orale est



évaluée localement par un jury constitué de deux examinateurs du centre agréé ne connaissant pas personnellement les candidats (Demeuse et collab. 2008).

### **3 Particularités de la version électronique du TEF**

De par la nature du support utilisé, le passage d'une version papier-crayon d'un test à une version électronique fait nécessairement apparaître des différences, tant en ce qui a trait à la présentation des items qu'au rapport général du candidat au test, où l'évolution du candidat s'effectue à travers une interface informatique et non plus par la manipulation de livrets. Par ailleurs, 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.

#### *Adaptation de l'interface pour la passation des épreuves obligatoires*

La version électronique du TEF (e-tef) répond aux préoccupations soulevées par Bridgeman et collab. (2003). Ainsi, le logiciel de passation affiche les images et textes à des tailles identiques quelle que soit la résolution de l'ordinateur, et des caractéristiques techniques minimales sont requises pour que les centres puissent organiser des sessions au moyen de la version électronique.

Pour tenir compte des qualités inégales de connexion et de débit Internet, un pré-téléchargement est effectué par le responsable du centre agréé en amont de la session, ce qui limite les échanges Internet à l'envoi de fichiers textes comportant, en début d'épreuve, les

questionnaires (sans images ni bandes son), et les réponses des candidats en fin d'épreuve.

Une même fluidité dans l'enchaînement des questions à l'intérieur d'une épreuve est donc garantie pour les différents candidats.

La taille des textes utilisés dans les items du TEF ne dépassant pas une demi-page en général, le nombre de cas où le candidat doit utiliser une barre de défilement est limité. Par ailleurs, la page de consignes précédant le test comporte elle-même le même type de barre de défilement, ce qui permet au candidat de se familiariser avec son utilisation avant le test (le surveillant attire l'attention des candidats sur cette barre de défilement lors de la lecture des consignes).

Ces précautions n'empêchent toutefois pas des différences inévitables dans la modalité de présentation du test et des items. Pour le candidat, l'une des différences essentielles réside dans la gestion du temps rendue beaucoup plus autonome par rapport à la version papier-crayon. 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. Dans la version papier-crayon, 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. En revanche, dans la version électronique, le candidat a la possibilité de transmettre ses réponses aux items des épreuves de compréhension écrite (CE) et de lexique/structure (LS) avant la fin du temps imparti. Pour la compréhension orale (CO), la transmission finale 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.

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'endroit appropriée, 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 constate principalement dans la simplification de la synchronisation du candidat dans l'épreuve de compréhension orale (CO). Alors que, dans la version papier-crayon, 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 réponses, 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, quels que soient le nombre de candidats et leur position dans la salle de test.

En revanche, alors qu'il est aisé pour un candidat de feuilleter les différentes pages d'un livret pour se faire une représentation rapide des différents formats d'item présents dans le test et de les localiser, la navigation dans l'interface logicielle est linéaire et s'effectue par

des boutons « suivant » et « précédent », ainsi qu'un bouton récapitulatif qui permet au candidat de connaître les numéros des questions auxquelles il a répondu ou non et d'accéder d'un clic à l'item correspondant à un numéro de question donné (voir la figure 1 ci-dessous). Par exemple, le repérage par un candidat des items comportant des textes courts en compréhension écrite est donc moins aisé. Cependant, les candidats qui ont étudié le manuel du candidat qui leur a été remis au moment de l'inscription ou qui ont consulté un ouvrage de préparation au TEF (CCIP, 2001 et Pons et Karcher, 2006) connaissent, préalablement à la passation, la structure du test et l'emplacement des différents types d'item.



**Figure 1 - Tableau récapitulatif des réponses aux questions du candidat**

### *Adaptation du format de présentation des items*

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 sur le plan de la présentation et de l'organisation iconique que de certaines modalités de réponse ou de consigne.

En compréhension écrite, par exemple, les supports image ou texte s'affichent au centre de l'écran, dans un cadre aux dimensions fixes. Si le texte est trop long pour tenir dans le cadre, le candidat peut le faire défiler au moyen d'une barre de défilement qui apparaît sur le bord droit du cadre. Les questions se rapportant au support image ou au support texte sont présentées les unes après les autres sous le cadre, lorsque le candidat sélectionne le bouton « suivant ».

La comparaison des figures 2 et 3 ci-dessous met en évidence l'amélioration de l'ergonomie du test dans sa version électronique pour certains types d'items. La figure 2 présente 5 questions de la version papier-crayon et renvoie à la feuille de réponses pour indiquer chacune des solutions choisies. La figure 3 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.

QUESTIONS 146 À 150	
<p>◆ Dans certaines phrases du texte suivant, des parties (A, B, C ou D) ont été soulignées. L'une de ces parties est incorrecte.</p> <p>☒ Cochez sur la fiche optique la réponse correspondant à la partie incorrecte.</p>	
<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 <u>ne date pas</u> d'hier. Mais de l'objet <u>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 par</u> 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 par</u> cuir.</p> <p style="text-align: center;">A B C D</p>
<p><b>QUESTION 148</b></p>	<p>Au XVIII<sup>e</sup> siècle, avec la <u>mode italienne</u>, les chaussures <u>s'affinant</u> et prennent des formes <u>élégantes</u>. 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 industriel</u> 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 avec</u> 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 2 – Présentation d'une unité extraite de l'épreuve de lexique/structure (LS) dans sa forme papier-crayon (TEF)**

**etef** Lexique et structure Valider l'épreuve

Pour chacune des questions suivantes, des parties du texte sont soulignées. Indiquez la partie du texte qui contient une erreur. Exemple :

Plusieurs versions ont été données sur la vie de Christophe Colomb, les débuts de sa carrière, l'origine de son infatigable envie de voyages. Fils d'un tisserand génois, il vint au Portugal en 1476 pour s'y marier et étudier la cartographie.

La partie vint au contient une erreur. (La forme correcte est "vint au").

**La chaussure**  
\*\*\*

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

Les peuples de l'Antiquité portaient surtout des sandales a 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 XVIe siècle, avec la mode italienne, les chaussures s'affinant 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

**QUESTION 4**  
Partie incorrecte

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

Précédent Temps restant Suivant

0% 50% 100%

**Figure 3 – Présentation d'une unité extraite de l'épreuve de lexique/structure (LS) dans sa forme électronique (e-tef)**

### *Les épreuves d'expression écrite et d'expression orale*

Comme il a été évoqué ci-dessus, les candidats n'ont pas toujours le choix du format dans lequel ils peuvent passer les épreuves du TEF. Certains centres agréés ne disposent pas d'un équipement informatique approprié pour organiser des sessions en format électronique alors que d'autres, qui ont investi dans des salles multimédia, souhaitent les rentabiliser en organisant exclusivement des sessions électroniques.

Dans ces conditions, imposer une production écrite au format électronique à des personnes qui n'ont qu'un rapport lointain avec l'outil informatique conduirait à les

désavantager fortement. Il a donc semblé plus équitable de conserver le format papier-crayon pour l'épreuve d'expression écrite, en dépit d'études qui montrent que les nouvelles générations de candidats, dans les pays où l'ordinateur individuel est couramment utilisé, obtiennent de meilleurs résultats lorsqu'ils rédigent des productions écrites avec la modalité écran-clavier.

### *Conséquences pour les centres agréés*

La planification des sessions et la délivrance officielle des résultats (tableau récapitulatif et attestations de résultats) suivent les mêmes procédures pour les deux versions du test, la seule différence étant les délais exigés, compte tenu du format du test, pour l'acheminement du matériel correspondant et la récupération des réponses des candidats.

C'est sur le plan de l'organisation que les changements sont les plus importants. Alors que, pour la version papier-crayon, le service de gestion administrative adresse par courrier au centre de passation l'ensemble du matériel nécessaire au déroulement du test, dans le cas de la version électronique, le centre de passation doit télécharger le matériel de test à partir d'un site Internet crypté. Le responsable du centre de passation doit alors procéder à une installation du logiciel de test sur chacun des postes destinés aux candidats et reproduire les sujets des épreuves d'expression écrite et orale qu'il aura téléchargés au format numérique.

Le jour du test, dans la version papier-crayon du TEF, les surveillants distribuent, au début de chaque épreuve, les livrets correspondants, qu'ils récupèrent à l'issue de l'épreuve. Certains candidats peuvent chercher à mettre à profit ces instants au cours desquels il y a du



mouvement, de la circulation de livrets et du bruit pour s'échanger des réponses, malgré la surveillance rapprochée. Par contre, dans la version électronique, lorsqu'il commence 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), l'épreuve de compréhension orale (CO) est lancée automatiquement. La fin du temps règlementaire imparti à cette épreuve entraîne le lancement de l'épreuve lexique/structure (LS). Lors de la fin d'une épreuve (par le candidat ou à l'issue du temps règlementaire imparti), les réponses sont communiquées automatiquement au site Internet e-tef. 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 papier-crayon, à cause de la compréhension orale.

À l'issue des épreuves, dans le cas de la version électronique, le responsable du centre de passation peut adresser les documents de session (feuille d'émargement, feuilles d'anomalies, fiche récapitulative de session) au service de gestion administrative par voie de télécopie, alors que pour la version papier-crayon ces documents, ainsi que les fiches optiques de réponses des candidats, étaient adressés au service de gestion administrative par courrier. De même, les grilles d'évaluation de l'épreuve d'expression orale sont adressées en format électronique au service de gestion administrative et, si la session ne comporte qu'un nombre limité de copies d'expression écrite, ces dernières peuvent être numérisées et adressées par voie de courriel. Plusieurs études montrent en effet que les différences de notation sont très faibles entre les versions manuscrites et numérisées d'une même copie lors de la correction d'épreuves de production écrite (voir notamment Zhang et collab., 2003 et ETS, 1993, cités dans Bennett, 2003). Enfin, le responsable du centre agréé supprime sur les ordinateurs le

logiciel de passation de la session alors que dans la version papier-crayon le matériel de test doit être retourné par courrier.

#### **4 Mise en équivalence de la version papier-crayon et de la version électronique**

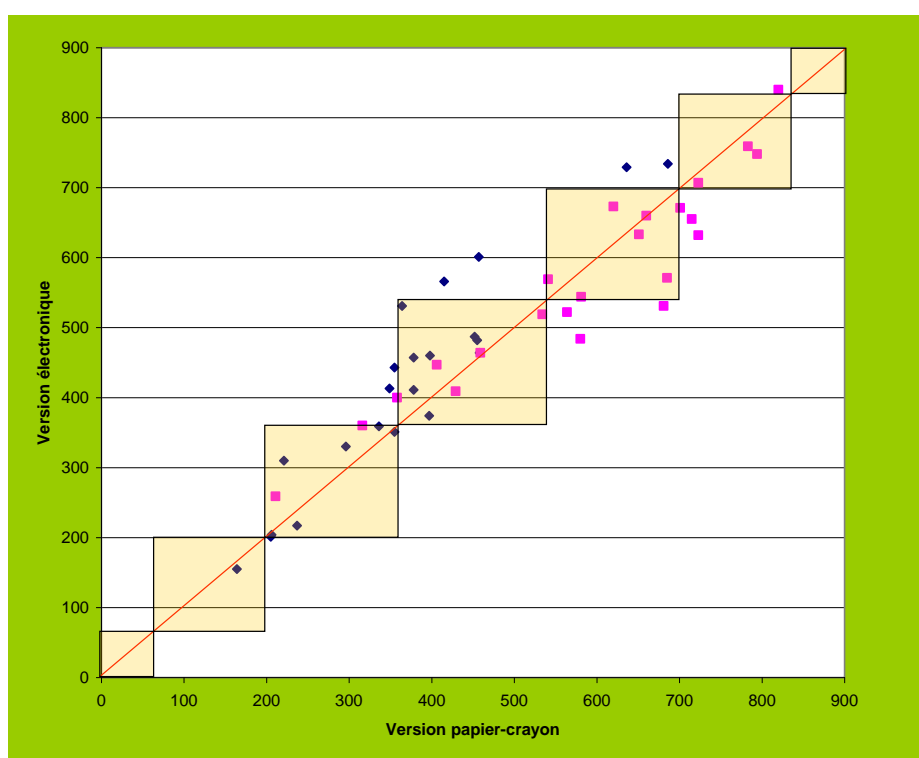
Malgré les précautions prises concernant la présentation des items, des différences existent entre les deux versions du test pour les parties composées de questions à choix multiples. Des études empiriques ont donc été entreprises, pour évaluer l'incidence de ces différences sur les réponses des candidats. Elles avaient pour objectif de vérifier la comparabilité des résultats des candidats entre les versions, d'une part, et d'autre part, de s'assurer que la version électronique ne rendait pas l'épreuve de compréhension écrite plus difficile du fait d'une éventuelle fatigue visuelle liée à la lecture sur écran, ou l'épreuve de compréhension orale plus facile du fait du confort d'écoute au moyen d'un casque audio.

##### *Analyse des résultats de candidats ayant passé successivement les deux versions du test*

La première étude a consisté à comparer des résultats d'échantillons de candidats ayant passé les épreuves obligatoires d'un jeu TEF au format électronique, puis celles d'une forme parallèle au format papier-crayon à quelques jours d'écart (Desroches et collab., 2005). Elle a été complétée par la prise en compte d'un échantillon de candidats de taille équivalente ayant passé, pour des jeux différents, les deux versions du test en ordre inverse.

Le graphique de la figure 4 ci-dessous permet de comparer les résultats obtenus (score global) par les candidats à chacune des deux versions du test. Les losanges bleus représentent les candidats ayant passé en premier la version papier-crayon et les carrés roses les candidats ayant passé en premier la version électronique. Les points situés à l'intérieur des rectangles

représentent les candidats qui ont obtenu le même niveau pour chacune des deux versions (l'échelle des scores du TEF est subdivisée en 7 niveaux qui ont été mis en correspondance avec l'échelle de niveaux du Cadre européen commun de référence pour les langues – Casanova et collab. 2010, à paraître – et avec l'échelle de niveaux des Standards linguistiques canadiens – Demeuse et collab. 2004). Les points situés au dessus du trait rouge représentent les candidats ayant obtenu un score plus élevé avec la version électronique.



**Figure 4 – Comparaison des résultats obtenus par deux échantillons de 22 et 23 candidats chacun soumis, en ordre inverse, à la version papier-crayon et à la version électronique du TEF**

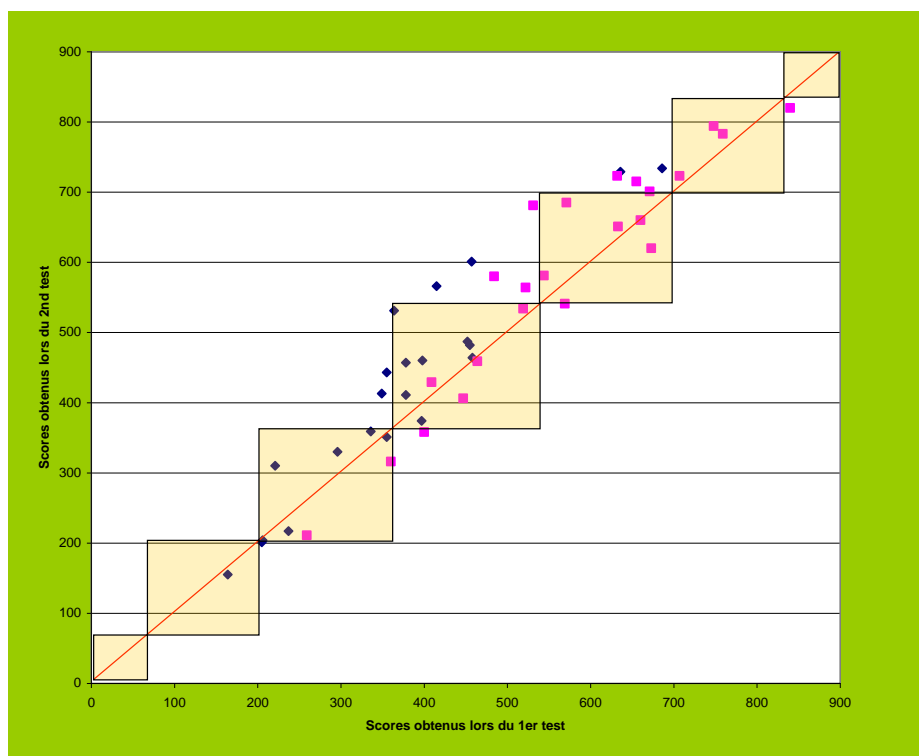
La corrélation entre les résultats de l'échantillon global (n=45) aux deux versions du test, pour des jeux d'épreuves différents, s'élève à 0,93. Les corrélations entre les différentes épreuves pour ces deux versions du test, papier-crayon et électronique, s'élèvent respectivement à 0,85 pour la compréhension écrite, à 0,89 pour la compréhension orale et à

0,80 pour le lexique/structure. Toutes ces corrélations sont statistiquement significatives avec une probabilité associée  $p < 0,0001$ . Par ailleurs, deux tiers des candidats (il s'agit des points de la figure 4 se situant à l'intérieur des carrés colorés) se voient attribuer un niveau identique avec chacune des deux versions.

Pour comparaison, les corrélations entre deux formes parallèles papier-crayon (Demeuse et collab. 2005), soumises aux mêmes candidats, à quelques jours d'écart, calculées sur la base de cinq séances de prétests regroupant chacune de 57 à 81 candidats (avant tout ajustement des formes et suppression/remplacement d'items), étaient comprises entre 0,84 et 0,91 (Chambre de commerce et d'industrie de Paris 2003). Les corrélations constatées sont par ailleurs comparables à celles obtenues entre les résultats aux versions papier-crayon et électronique pour le IELTS dans l'étude de Shaw et collab. (2001) citée dans Green et Maycock (2004), c'est-à-dire 0,83 pour la compréhension orale et 0,90 pour la compréhension écrite.

La figure 4 illustre cependant que si deux tiers des candidats se voient attribuer un niveau identique, 16 % obtiennent un niveau plus élevé avec la version papier-crayon et 18 % obtiennent un niveau plus élevé avec la version électronique. Les fortes variations de scores ainsi constatées sont probablement liées à une familiarisation des candidats avec la nature des épreuves lors du premier test (quelle que soit la version utilisée), qui les rend plus performants lors du second test. En effet, lorsqu'on compare les scores obtenus à chacun de ces deux tests (voir la figure 5 ci-dessous), on constate une amélioration sensible des résultats. De nouveau, les losanges bleus représentent les candidats ayant passé en premier la version papier-crayon et les carrés roses les candidats ayant passé en premier la

version électronique. Les points situés dans les carrés correspondent aux candidats ayant obtenu un niveau identique pour chacune des deux versions.



**Figure 5 – Comparaison des résultats obtenus par les candidats selon l'ordre des tests**

Si deux tiers des candidats se sont vu attribuer le même niveau par les deux tests, 27 % des candidats parviennent au niveau supérieur lors de la seconde passation et seuls 7 % se voient attribuer, de justesse, un niveau inférieur lors de la seconde passation.

L'ampleur des variations dépend de l'ordre dans lequel sont passées les deux versions et on constate que l'amélioration moyenne des scores est plus élevée quand les candidats passent la version papier-crayon avant la version électronique que dans le cas inverse. Le

tableau 1 ci-dessous permet de comparer la moyenne et l'écart-type des différences de scores obtenues entre deux passations successives dans trois cas de figure :

- passation de la version papier-crayon puis de la version électronique;
- passation de la version électronique puis de la version papier-crayon;
- passations successives, par 205 candidats, de deux formes parallèles en version papier-crayon dans le cadre d'un prétest.

**Tableau 1 - Moyenne et écart-type des différences de scores obtenues lors de la passation successive de deux formes parallèles du test, dans un format différent ou identique**

	Papier-crayon / électronique	Électronique / papier-crayon	Papier-crayon / Papier-crayon
Moyenne des différences de score (score maximum = 900)	49,1	20,8	32,9
Écart-type des différences de score (score maximum = 900)	55,6	54,9	59,2

On constate que si l'écart-type des variations de score est semblable dans les trois cas, il n'en va pas de même pour la moyenne, ce qui laisse supposer que les candidats obtiennent de meilleurs résultats avec la version électronique du test. Cependant, lorsqu'on mène une analyse de comparaison des moyennes sur les résultats obtenus par les candidats (tableau 2 ci-dessous), indifféremment de l'ordre de passation, à la version électronique et à la version papier-crayon du test (échantillons appariés), on constate que les valeurs de la probabilité associée  $p$  sont supérieures au niveau de signification seuil de 0,05. Ainsi, on peut valider l'hypothèse nulle  $H_0$  selon laquelle la différence entre les moyennes est nulle et conclure que les résultats des candidats aux deux versions du test sont comparables.

**Tableau 2 - Comparaison des moyennes au TEF et à e-tef pour l'échantillon de 45 candidats**

	<b>Différence de moyenne TEF / e-tef</b>	<b>t (Valeur observée)</b>	<b>t (Valeur critique)</b>	<b>Degrés de liberté</b>	<b>Probabilité p (bilatérale)</b>
<b>Total</b>	-13,400	-1,382	2,015	44	0,174
<b>CE</b>	-1,511	-0,294			0,770
<b>CO</b>	-8,733	-1,631			0,110
<b>LS</b>	-3,156	-0,657			0,514

*Comparaison des estimations de difficulté des items de jeux identiques soumis à des échantillons de candidats différents selon le format du test*

La seconde étude, menée à partir des résultats de candidats différents, cherchait à vérifier que les estimations de la difficulté des items d'un même jeu (par application du modèle de Rasch<sup>2</sup>) étaient comparables selon qu'elles étaient calculées à partir des résultats d'un échantillon de candidats soumis au format papier-crayon ou à partir des résultats d'un échantillon indépendant de candidats soumis au format électronique du test.

L'échantillon de candidats soumis à la version électronique comportait 41 candidats non francophones, ayant pour la plupart entre 18 et 30 ans et suivant des cours de français dans un centre de langue en France, et l'échantillon soumis à la version papier-crayon rassemblait 77 candidats de profil similaire. Pour chacune des épreuves, la corrélation entre les estimations de difficultés des items, calculées pour chacun deux formats, est élevée : 0,93 en compréhension écrite, 0,92 en compréhension orale et 0,95 en lexique/tructure. L'estimation des difficultés s'effectuant sous la contrainte que leur moyenne est nulle et les écart-types de ces estimations étant très proches, les difficultés sont comparables d'une version à l'autre. En reportant sur une même échelle les estimations des difficultés de

---

2. Le lecteur intéressé par les modèles de mesure pourra se reporter à l'ouvrage de Bertrand et Blais (2004).

l'ensemble des items du prétest pour le format électronique, d'une part, et pour le format papier-crayon, d'autre part, on peut comparer la moyenne et l'écart-type des difficultés par épreuve. Le tableau 3 ci-dessous présente ainsi la moyenne et l'écart-type des difficultés des items pour chaque épreuve.

**Tableau 3 - Moyenne et écart-type des estimations de difficulté des items d'un même prétest, pour chacun des deux formats utilisés**

	<b>TEF (N=77 candidats)</b>		<b>e-tef (N=41 candidats)</b>	
	Moyenne	Écart-type	Moyenne	Écart-type
<b>CE</b>	0,08	1,51	0,08	1,41
<b>CO</b>	-0,29	1,51	-0,21	1,47
<b>LS</b>	0,33	1,61	0,21	1,67

On constate que, pour ce prétest (avant donc, tout ajustement de forme), quel que soit le format de test utilisé, les items de compréhension orale semblent en moyenne un peu moins difficiles et que les items de lexique/structure semblent un peu plus difficiles, les items de compréhension écrite étant de difficulté moyenne proche de 0. Les valeurs obtenues tant pour la moyenne que pour l'écart-type sont par ailleurs très proches et la version électronique ne semble donc pas rendre plus difficile ou plus facile une des épreuves relativement aux autres. Il n'y a donc pas d'effet notable du format de test sur les estimations de difficulté des items. Les corrélations obtenues entre les estimations à partir des résultats à chacune des deux versions sont par ailleurs proches de celles que l'on constate pour un prétest au format papier-crayon, quand on considère l'estimation de la difficulté des items à partir de deux échantillons de candidats semblables (0,92 en compréhension écrite, 0,88 en compréhension orale et 0,95 en lexique/structure).

Ces corrélations sont d'autant plus encourageantes qu'elles sont établies à partir de résultats à des prétests, avant tout ajustement des formes et suppression/remplacement



d'items, et sans répartition aléatoire (les candidats d'une même session se retrouvaient dans un même échantillon). On peut donc s'attendre à des corrélations plus élevées lorsque les estimations sont effectuées sur des jeux validés. Ainsi, lorsqu'on constitue deux échantillons de 100 personnes par répartition aléatoire d'une partie des candidats à un jeu TEF validé, on obtient des corrélations sensiblement plus élevées (0,96 en compréhension écrite, 0,97 en compréhension orale et 0,97 en lexique/structure).

## **Conclusion**

La conception d'une version électronique d'un test équivalente à une version papier-crayon originale est une entreprise délicate, qui requiert un soin particulier tant sur le plan de la présentation des items que de la standardisation de la passation. Cela est d'autant plus vrai lorsque les deux formes subsistent en parallèle, ce qui contraint souvent à conserver la structure initiale du test et empêche le recours à une version adaptative.

Pour s'assurer de l'efficacité des dispositions prises pour rendre les deux versions équivalentes, il est nécessaire de mener des études empiriques de comparaison des résultats et des difficultés relatives des différentes parties du test. Les expérimentations mises en œuvre pour vérifier cette équivalence dans le cas du TEF montrent que les choix effectués pour la mise au point de la version électronique permettent d'obtenir des résultats comparables à ceux que l'on constate pour la version papier-crayon.

Au-delà de la réponse qu'elle apporte au besoin exprimé d'une évaluation plus rapide du niveau de compétence en langue française des candidats, la version électronique présente un ensemble d'avantages (réduction des délais d'organisation, délivrance immédiate des

scores, renforcement de la sécurité par l'affectation individuelle des épreuves, dématérialisation des échanges) et rend la passation plus confortable pour les candidats. Cependant, compte tenu des contraintes techniques qu'elle fait peser sur certains centres agréés, en termes d'équipement informatique et de liaison Internet, la version électronique ne peut pas se substituer complètement à la version papier-crayon. Les contraintes techniques, combinées à des considérations d'équité vis-à-vis des candidats, constituent également un frein à la mise en œuvre d'une version électronique des épreuves d'expression écrite et d'expression orale du TEF.

## Références

Bennett, R. E. (2003), *Online Assessment and the Comparability of Score Meaning*, Educational Testing Service Research Memorandum RM-03-05, [en ligne] [[www.ets.org/Media/Research/pdf/RM-03-05-Bennett.pdf](http://www.ets.org/Media/Research/pdf/RM-03-05-Bennett.pdf)]

Bertrand, R. et J.-G. Blais (2004). *Modèles de mesure : l'apport de la théorie des réponses aux items*. Montréal, Canada, Presses de l'Université du Québec.

Blais, J.-G. (dir.) (2009). *Évaluation des apprentissages et technologies de l'information et de la communication: Enjeux, applications et modèles de mesure*. Québec, Canada, Les Presses de l'Université Laval.

Bridgeman, B., M. L. Lennon et A. Jackenthal (2003). « Effects of screen size, screen resolution, and display rate on computer-based test performance ». *Applied Measurement in Education*, 16, p. 191-205.

J.-G. Blais et J.-L. Gilles (2011), *Évaluation des apprentissages et technologie de l'information et de la communication vol.2*. Québec, Canada, Les Presses de l'Université Laval - chapitre 12

Burton R., M. Reichert et R. Martin (2009). « L'évaluation des compétences langagières », dans J. G. Blais (dir.), *Évaluation des apprentissages et technologies de l'information et de la communication. Enjeux, applications et modèles de mesure*, p. 57-76. Québec, Canada, Les Presses de l'Université Laval.

Centre pour les niveaux de compétence linguistique canadiens (2006). *Niveaux de compétence linguistique canadiens. Citoyenneté et Immigration Canada*, [en ligne] [[www.language.ca/display\\_page.asp?page\\_id=570](http://www.language.ca/display_page.asp?page_id=570)].

Casanova, D., A. Crendal, M. Demeuse, F. Desroches et A. Holle (2010, à paraître). *Validation empirique d'un test de français langue étrangère en regard du Cadre européen commun de référence pour les langues*, dans Les Actes du XXI<sup>e</sup> colloque international de l'Association pour le développement des méthodologies d'évaluation en éducation (ADMEE-Europe). Braga, Portugal, 14-16 janvier 2010.

Chambre de commerce et d'industrie de Paris (2001). *Test d'évaluation de Français*. Paris, France, Hachette.

Chambre de commerce et d'industrie de Paris (2003). *Adéquation du Test d'évaluation de français aux critères d'évaluation de Citoyenneté et Immigration Canada*. Paris, France, CCIP, Direction des relations internationales de l'enseignement.

J.-G. Blais et J.-L. Gilles (2011), *Évaluation des apprentissages et technologie de l'information et de la communication vol.2*. Québec, Canada, Les Presses de l'Université Laval - chapitre 12

Chambre de commerce et d'industrie de Paris (2004). *Présentation d'éléments de correspondance entre les questions du TEF et les Standards linguistiques canadiens*. Paris, France, CCIP, Direction des relations internationales de l'enseignement.

Chambre de commerce et d'industrie de Paris (2007). *Le Test d'Évaluation de Français de la Chambre de commerce et d'industrie de Paris*. Paris, France, 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, France, Didier.

Demeuse, M., F. Artus, A. Crendal, D. Casanova et F. Renaud (2008). *Évaluer les productions orales en français langue étrangère (FLE) en situation de test. Étude de la fidélité inter-juge de l'épreuve d'expression orale du Test d'évaluation de français (TEF) de la Chambre de commerce et d'industrie de Paris (CCIP)*, dans Les Actes du XX<sup>e</sup> colloque international de l'Association pour le développement des méthodologies d'évaluation en éducation (ADMEE-Europe). Genève, Suisse, 9-11 janvier 2008.

Demeuse, M., A. Crendal, F. Desroches, F. Renaud, D. Casanova et F. Artus, F. (2005). *La fiabilité de l'évaluation des compétences linguistiques pour des adultes non francophones: présentation d'un protocole d'évaluation*, dans Les Actes du XVIII<sup>e</sup> colloque international de l'Association pour le développement des méthodologies d'évaluation en éducation (ADMEE-Europe), Reims, France, 24-26 octobre 2005.

J.-G. Blais et J.-L. Gilles (2011), *Évaluation des apprentissages et technologie de l'information et de la communication vol.2*. Québec, Canada, Les Presses de l'Université Laval - chapitre 12

Demeuse, M., F. Desroches, A. Crendal, F. Renaud, P. Oster et X Leroux (2004). *L'évaluation des compétences linguistiques des adultes en français langue étrangère dans une perspective de multiréférentialisation*, dans Les Actes du XIX<sup>e</sup> colloque international de l'Association pour le développement des méthodologies d'évaluation en éducation (ADMEE-Europe). Lisbonne, Portugal, 18-20 novembre 2004.

Desroches, F., D. Casanova, A. Crendal, F. Renaud et M. Demeuse (2005). *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'auto-évaluation*, dans Les Actes du XVIII<sup>e</sup> colloque international de l'Association pour le développement des méthodologies d'évaluation en éducation (ADMEE-Europe). Reims, France, 24-26 octobre 2005.

Educational Testing Service (1993). *The results of the NAEP 1993 field test for the 1994 National Assessment of Educational Progress*. Princeton, Etats-Unis, Auteur.

Godin, C., J.-G. Blais et M. Laurier (2011). « Textes manuscrits ou textes dactylographiés : le mode de présentation a-t-il un impact sur la correction? » Dans J.-G. Blais et J.-L. Gilles, *Évaluation des apprentissages et technologie de l'information et de la communication vol.2*. Québec, Canada, Les Presses de l'Université Laval.

Green, T. et L. Maycock (2004). *Computer-based IELTS and paper-based versions of IELTS*, Research Notes, issue 18. Cambridge, Grande-Bretagne, Cambridge ESOL.

Houssemand, C., R. Meyers et R. Martin (2009). « L'évaluation du profil psychosocial de demandeurs d'emploi, une population peu familiarisée à la technologie informatique », dans

J.-G. Blais et J.-L. Gilles (2011), *Évaluation des apprentissages et technologie de l'information et de la communication vol.2*. Québec, Canada, Les Presses de l'Université Laval - chapitre 12

J.-G. Blais (dir.), *Évaluation des apprentissages et technologies de l'information et de la communication. Enjeux, applications et modèles de mesure*, p.137-158. Québec, Canada, Les Presses de l'Université Laval.

Laurier, M. D. et L. Diarra (2009). « L'apport des technologies dans l'évaluation de la compétence à écrire », dans J.-G. Blais (dir.), *Évaluation des apprentissages et technologies de l'information et de la communication. Enjeux, applications et modèles de mesure*, p. 77-104. Québec, Canada, Les Presses de l'Université Laval.

Mead, A., D. et F. Drasgow (1993). « Equivalence of computerized and paper-and-pencil cognitive ability tests: a meta-analysis ». *Psychological Bulletin*, 114(3), p. 449-458.

Noël-Jothy, F. et B. Sampsonis (2006). *Certifications et outils d'évaluations en FLE*. Paris, France, Hachette.

Pons, S. et G. Karcher (2006). *Test d'évaluation de français, 250 activités*. Paris, France, Cle international.

Russell, M. et W. Tao (2004). « Effects of handwriting and computer-print on composition scores: a follow-up to Powers, Fowles, Farnum and Ramsey ». *Practical Assessment, Research & Evaluation*, 9(1), [en ligne] [<http://PAREonline.net/getvn.asp?v=9&n=1>].

Shaw, S., N. Jones et T. Flux (2001). *CBIELTS – A comparison of computer-based and paper versions*. Cambridge, Grande-Bretagne, Cambridge ESOL internal Validation report 216.

J.-G. Blais et J.-L. Gilles (2011), *Évaluation des apprentissages et technologie de l'information et de la communication vol.2*. Québec, Canada, Les Presses de l'Université Laval - chapitre 12

Taylor, C., J. Jamieson, D. Eignor et I. Kirsch 1998). *The Relationship Between Computer Familiarity and Performance on Computer-Based TOEFL Test Tasks*, (TOEFL Research Report n°61). Princeton, États-Unis, Education Testing Service.

Zhang, Y.L., D. E. Powers, W. Wright et R. Morgan (2003). *Applying the online scoring network (OSN) to Advanced Placement Program® (AP®) tests* (ETS RR-03-12). Princeton, États-Unis, Education Testing Service.