

# LES BIAIS CULTURELS DANS LES TESTS INTERNATIONAUX DE FRANÇAIS LANGUE ÉTRANGÈRE : DIAGNOSTIC ET ANALYSE

DESROCHES Franck, *Responsable du TEF*, Chambre de commerce et d'industrie de Paris, Paris, France, [fdesroches@ccip.fr](mailto:fdesroches@ccip.fr)

Chambre de commerce et d'industrie de Paris, Direction des relations internationales de l'enseignement, 28, rue de l'Abbé Grégoire, 75279 PARIS Cedex 06, France.

CRENDAL Alexandra, *Responsable pédagogique*, Chambre de commerce et d'industrie de Paris, Paris, France, [acrendal@ccip.fr](mailto:acrendal@ccip.fr)

RENAUD François, *Responsable pédagogique*, Chambre de commerce et d'industrie de Paris, Paris, France, [frenaud@ccip.fr](mailto:frenaud@ccip.fr)

CASANOVA Dominique, *Responsable pédagogique*, Chambre de commerce et d'industrie de Paris, Paris, France, [dcasanova@ccip.fr](mailto:dcasanova@ccip.fr)

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

ARTUS Frédérique, *Chercheuse*, Institut d'Administration scolaire, Université de Mons-Hainaut, Mons, Belgique, [frederique.artus@umh.ac.be](mailto:frederique.artus@umh.ac.be)

**Mots clés** : évaluation, français langue étrangère, biais culturels

---

## Introduction

Les tests internationaux de français langue étrangère (FLE) sont récents : les premiers (CGM 62, SOL test) sont apparus dans les années 70. Ils se caractérisent par une standardisation des conditions et des consignes de passation, première exigence pour assurer la qualité d'un test, et par un renvoi au référentiel de l'époque, le Français Fondamental (Gougenheim et al., 1956). Leur diffusion à l'étranger aura finalement peu de portée. Les années 80 voient un tournant avec l'avènement de nouvelles tendances en didactique des langues : les approches communicatives ; les tests de langues sont alors relégués au second plan. A cette époque, peu d'études sont menées sur les biais culturels.

Ce n'est qu'à la fin des années 90, en 1998, que le premier Test d'Évaluation de Français (TEF) pour non-francophones est mis au point par une équipe de linguistes et didacticiens de la Chambre de commerce et d'industrie de Paris (CCIP)<sup>1</sup>, rejointe par un Comité Scientifique pluridisciplinaire (CCIP, 2006). Sa conception repose sur une multiréférentialisation (Demeuse, 2004), les ouvrages de référence étant le Cadre européen commun de référence pour les langues (CECR) du Conseil de l'Europe (2000), échelle de niveaux et de compétences s'appuyant sur la perspective actionnelle, et les Niveaux de compétence linguistique canadiens (CIC, 2002)<sup>2</sup>. Le TEF se base sur les approches communicatives. Deux visions, *a priori* opposées, semblent coexister : d'une part une langue normée pour les besoins de standardisation du test et, d'autre part, une langue vivante qui évolue au fil des usages que les natifs en font et qui s'inscrit dans les approches communicatives.

---

<sup>1</sup> Consultable sur <http://www.fda.ccip.fr>

<sup>2</sup> Cette dénomination est récente et remplace désormais celle utilisée de 2002 à 2006 : les Standards linguistiques canadiens. Ces niveaux sont consultables sur [http://www.language.ca/display\\_page.asp?page\\_id=533](http://www.language.ca/display_page.asp?page_id=533)

Face à ce paradoxe, quel choix doit préexister à la création d'un test de FLE destiné à des candidats de langues maternelles différentes dont, selon G. Mounin (1975), la pratique et la maîtrise « déterminent, prédéterminent même, notre vision du monde » ? C'est précisément la question que les auteurs du TEF se sont posée lors du processus d'élaboration du test. D'autres interrogations, secondaires, ont également fait l'objet de discussions : un accent neutre peut-il être retenu pour l'épreuve de compréhension orale ? Quels outils développer pour éviter le "franco-français" et les préjugés culturels ? Comment faire en sorte que les questions liées à la partialité soient rejetées ? Comment rester objectif en expression écrite et orale ? Autant de questions auxquelles les concepteurs du test répondent à travers les spécifications du test. Ces interrogations ne sont pas sans lien avec la notion de biais culturels (CCIP-DRI/E, 2003) qui a fait l'objet d'un contrôle particulier de la qualité des items.

L'étude sur les biais culturels présentée ci-après se proposera de présenter une typologisation des biais culturels rencontrés aux différentes étapes de conception, validations pédagogique et psychométrique, leur fréquence et les outils créés pour prévenir leur apparition dans les épreuves du TEF. Pour la démonstration, les auteurs s'appuieront sur un corpus de données empiriques recueillies en amont et en aval des pré-tests et s'étalant sur la période de janvier 2004 à juin 2006. L'exposé proposera deux approches complémentaires : une approche psychométrique complétée par une lecture pédagogique critériée.

## 1. Vers une sociolinguistique du français

La question du choix d'une langue pour la conception d'un test de français langue étrangère oblige les auteurs du test à faire une analyse de besoins, étude qui ne peut être réalisée qu'après avoir déterminé précisément les objectifs et le public ciblés par le test.

### ***De la commande institutionnelle à l'analyse des besoins du public***

Le Test d'Évaluation de Français a été initialement créé pour répondre à deux commandes : l'une des écoles de la Chambre de commerce et d'industrie de Paris, l'autre d'établissements d'enseignement à l'étranger dispensant des cours de français ou en français à un public non francophone. Les besoins langagiers de ces publics ont donc été analysés et ont révélé des situations de communication authentiques variées ; celles communes à l'ensemble de ces publics ont permis à la DRI/E-CCIP d'inscrire les domaines (personnel, éducatif, etc.) et thèmes de communication (l'environnement, les achats, etc.) dans les spécifications du test en précisant leur niveau en regard du CECR.

Outre ces premiers choix, la DRI/E-CCIP a dû s'orienter vers une variété du français, celle la plus adaptée au public et objectifs du test. Parlé dans plusieurs pays, la langue française possède différents statuts selon les régions ou les pays francophones. En outre, elle subit constamment des actualisations rendues nécessaires par l'évolution socio-économique des groupes sociaux dont elle reflète l'identité. Le statut de la langue et le passé historique du pays expliquent aussi les évolutions de cette langue que les candidats au test sont amenés à rencontrer aussi bien en contexte formel, en classe de langue, qu'informel, dans la rue. Le test s'adressant à un public large amené à rencontrer toutes les variétés du français, aussi bien en contexte endolingue qu'exolingue, le choix définitif s'est porté sur le français standard de France, des médias.

Ce choix nécessaire à la standardisation n'est pas sans influencer la passation du TEF dans des pays francophones. Par exemple, en compréhension orale, il arrive que des étudiants soient surpris, en Belgique, d'entendre « quatre-vingt-dix » et non « nonante » ; en lexique / structure, les candidats passant le TEF au Québec peuvent être induits en erreur par le verbe « dîner » qui signifie « déjeuner » en France. À cette variété écrite s'ajoute celle de l'oralité rencontrée dans les épreuves de compréhension et production orales : au-delà du texte support, les éléments segmentaux et suprasegmentaux viennent complexifier le niveau d'évaluation. Afin d'atténuer cette variante, la DRI/E-CCIP porte un intérêt particulier aux transcriptions transmises par les concepteurs et aux normes d'enregistrement de ces supports.

Au-delà du choix des spécifications du test qui affirment la recherche d'une standardisation de la langue, la DRI/E-CCIP reste vigilante quant à la production de batteries de tests qui doivent rester équivalentes au fil du temps pour assurer la fiabilité du test (Demeuse et al., 2005) et ne pas favoriser un groupe particulier. Là intervient la notion de culture, « un ensemble de caractéristiques propres à une société donnée », selon la définition donnée par Robert Galisson. Le TEF cherche constamment à neutraliser les aspects singuliers d'une culture donnée afin d'éviter de biaiser le résultat des candidats au test. Ces spécificités culturelles sont à la source de ce que nous nommerons dans cette étude « biais culturels ».

### **Présentation des spécifications du test**

Le Test d'Évaluation de Français (TEF) est un test de français langue étrangère qui s'adresse à des non francophones. Il se compose de plusieurs épreuves : la compréhension écrite, la compréhension orale, le lexique / structure, l'expression écrite et l'expression orale dont les spécifications générales sont présentées dans le tableau 1.

**Tableau 1** – Les épreuves du TEF

Indexation sur le CECR	Epreuves obligatoires	Durée	Structure	Score sur
Réception	Compréhension écrite	60 min	50 questions	300 points
Réception	Compréhension orale	40 min	60 questions	360 points
Linguistique	Lexique / structure	30 min	40 questions	240 points
	<b>TOTAL</b>	<b>2H10</b>	<b>150 questions</b>	<b>900 points</b>
Indexation sur le CECR	Epreuves facultatives	Durée	Structure	Score sur
Production	Expression écrite	60 min	2 sujets à traiter	450 points
Interaction	Expression orale	35 min	2 sujets à traiter	450 points

Source : D'après la brochure de présentation du Test d'Évaluation de Français, 2006.

Pour les deux dernières épreuves, ouvertes et plus exposées à la subjectivité, des grilles d'évaluation ont été élaborées afin de réduire les écarts entre examinateurs et de rendre objective l'évaluation de ce type d'épreuve. Conçues à partir du CECR, ces deux grilles comptent chacune douze critères d'évaluation et 7 niveaux TEF indexés sur les six niveaux du CECR.

Basé sur les approches communicatives, le TEF s'appuie sur des documents authentiques et présente, pour ses épreuves de production écrite et orale, des sujets, des situations que les candidats au test peuvent rencontrer dans la vie quotidienne. Attrayantes et motivantes pour le candidat, ces situations de communication authentiques ne sont pas sans complexifier le travail des concepteurs. Elles requièrent toute l'attention du Comité d'Évaluation et de Validation, chargé de « traquer » toute imperfection au niveau de l'item ou du support, et notamment les biais culturels.

Les spécifications du TEF et leurs mises en œuvre s'appuient sur des critères de scientificité légués par la recherche dans les domaines de la linguistique et de la psychométrie. Ainsi, parmi les réflexions et recherches sur lesquelles reposent la création du test, président les travaux en sociolinguistique de Louis-Jean Calvet (1987) et les références données par Gilbert De Landsheere sur la définition scientifique d'un test (De Landsheere, 1992).

## 2. Définition et essai de typologisation des biais

Précisons la définition de biais culturel. De Landsheere (1992) indique qu'« une mesure est *biaisée* si elle conduit systématiquement à des résultats inexacts dus à des déficiences inhérentes à l'instrument utilisé, à la façon d'utiliser un instrument ou à la façon d'interpréter les résultats ». Pour ce qui concerne le Test d'Évaluation de Français (TEF), la question du biais culturel se pose de multiples façons, à toutes les phases critiques de l'acte d'évaluer un candidat avec le TEF : depuis la conception des items de test et des sujets d'épreuves de production, jusqu'à l'évaluation proprement dite de la performance du candidat par l'examineur d'expression orale.

Un biais culturel surgissant lors de la phase de conception du test renvoie, dans la définition de De Landsheere, à « *des déficiences inhérentes à l'instrument utilisé* ». En regard des activités du test, on pourrait reformuler la définition comme suit : **Erreur de conception touchant un ou plusieurs éléments constitutifs d'une unité d'évaluation** (consigne, document support, amorce, clé, distracteur) **susceptible d'influencer la réponse** du candidat (vers une réponse fautive, vers une réponse vraie, vers l'abstention) **pour des raisons d'appartenance** à un « groupe » culturel (genre, orientation politique, religieuse, groupe social, groupe national, âge, etc.) **ou de connaissances** culturelles (culture générale non universellement partagée sur des faits culturels ou des caractéristiques historiques, géographiques, sociales, économiques, religieuses, etc. auxquels peuvent renvoyer un ou plusieurs éléments constitutifs d'une unité d'évaluation) **indépendamment de sa compétence réelle à communiquer** (et qui est, normalement, la seule compétence évaluée).

La typologie d'erreurs de conception TEF, dont un extrait est présenté dans le tableau 2, développée à l'intention du Comité d'Évaluation et de Validation (chargé de valider pédagogiquement les items proposés par les concepteurs) et des concepteurs eux-mêmes, prévoit deux catégories d'erreurs sous la rubrique « Culture ».

**Tableau 2** – Extrait de la typologie d'erreurs de conception

Typologie	Définition	Remédiation
Culture	Thématique culturellement / politiquement sensible pour certains candidats.	<i>Neutraliser</i>
	Référence à un particularisme culturel / politique / social / économique non explicité.	<i>Rendre universel</i>

Source : Document extrait de la typologie d'erreur (CTEF-LIS-15)

Ces deux catégories renvoient, d'une part, à l'appartenance du candidat à un groupe culturel donné et, d'autre part, aux connaissances culturelles que le candidat possède. Naturellement, appartenance culturelle et connaissances culturelles varient considérablement d'un candidat à l'autre. Cette catégorisation binaire des biais culturels a un caractère général. Chaque catégorie renvoie en fait à des réalités multiples et complexes. Ainsi, lorsque l'on parle de groupe culturel, on peut définir ce groupe de différentes façons. Mais ce sont les mêmes mécanismes qui sont à l'œuvre pour constituer le biais culturel dans l'outil d'évaluation conçu. Lorsqu'une unité de test peut donner lieu, d'un candidat à l'autre, à une différence de traitement (traitement de la tâche par le candidat ou traitement/évaluation du candidat par le correcteur/examineur) en raison de leur appartenance à tel ou tel groupe culturel, le biais culturel est établi.

Dans un premier temps, on peut définir le « groupe culturel » comme un ensemble d'individus liés par leur appartenance nationale (ou régionale), par leurs convictions idéologiques (politiques ou religieuses), par leur statut social ou professionnel, par leur culture éducative, etc. Ainsi, une unité de test peut être porteuse d'un biais culturel lié à l'appartenance du candidat à un groupe de conviction politique ou idéologique car elle impose un point de vue distinct ou opposé à celui du groupe. Par exemple, la figure 1 présente un exemple d'unité qui a été rejetée par le Comité d'Évaluation et de Validation. La situation proposée dans cette unité est très marquée idéologiquement : elle impose un point de vue qui n'est pas forcément celui du candidat. Il éprouvera plus de difficultés à défendre un point de vue contraire à ses convictions qu'un candidat partageant la prise de position imposée par la consigne.

ES511005 - CEB « Rébellion-Energétique. L'EPR contre-attaque »

Consigne : « Vous êtes sûr(e) que l'énergie éolienne est une bonne énergie et que le nucléaire n'est pas une bonne énergie »

*Figure 1 – Item présentant un risque de biais culturel lié aux valeurs*

De la même manière, une unité de test contenant des assertions négatives sur une culture nationale ou régionale, peut entraîner une gêne déstabilisante chez un candidat ressortissant de cette culture. Par exemple, l'unité présentée en figure 2 a été corrigée par le Comité d'Évaluation et de Validation pour cette raison. En effet, le document fait état d'un jugement de valeur sur un fait culturel (ici, américain) en l'opposant à une autre culture (ici, française). Ce jugement de valeur négatif est susceptible de gêner un candidat américain et de perturber sa concentration pour la suite du test.

SP0205-04 CEB « Histoire d'une rencontre »

Document : « J'adore le pain français. Je souligne français, parce que je suis américain, et que le pain dans mon pays est généralement immonde. »

*Figure 2 – Item présentant un risque de biais culturel lié à l'image*

Si l'on intègre à la définition de « groupe culturel » des caractéristiques comme le sexe ou l'âge (dans leur dimension culturelle), le même type de biais peut survenir. Ces caractéristiques peuvent représenter une source de biais en elles-mêmes (lorsque la situation d'évaluation exige, par exemple, qu'un candidat masculin joue le rôle d'une femme), ou bien agir comme un renforcement de caractéristiques autres (sociales, religieuses, etc.) et donc renforcer le biais induit par l'unité porteuse de l'erreur de conception.

De la même manière, si l'on intègre à la notion de « groupe culturel » des caractéristiques plus individuelles, relevant peut-être davantage de la psychologie individuelle ou de l'histoire personnelle que de l'imprégnation culturelle, de nouvelles sources de biais apparaissent, qui établiraient une inégalité face à la tâche en jouant sur les différences de sensibilité des candidats face à une situation morbide ou macabre. Tel est le cas de l'unité présentée dans la figure 3. Cette unité a été rejetée par le Comité d'Évaluation et de Validation pour cette raison. Pour peu que le sujet développé dans le document fasse écho chez le candidat à une angoisse ou à un traumatisme personnel, sa capacité à répondre risque d'être perturbée.

2TF0104-12 CEC « Disparitions »

Document : « Il y a des histoires qui fascinent parce qu'elles naviguent dans ces **territoires troubles** où l'espérance et la **peur** n'en finissent pas de se confondre : ce sont les affaires de **disparitions**. »

**Figure 3** – Item présentant un risque de biais culturel lié à la sensibilité personnelle

Lorsque l'on parle de connaissances culturelles, il peut s'agir d'une compétence culturelle générale (culture générale), forcément variable d'un candidat à l'autre, et qui peut servir de pierre d'achoppement à un biais culturel lorsque l'unité en cause propose un item auquel il est plus facile de répondre si l'on possède tel ou tel élément de culture générale. La figure 4 en est une bonne illustration. Les options A et B renvoient à la distinction entre un nom et un prénom japonais. Cette distinction n'est pas évidente, voire impossible pour un candidat qui n'a aucune connaissance particulière de la culture japonaise. En l'occurrence, il ne s'agit pas seulement de savoir à quoi renvoient les termes « nom » et « prénom » (compétence linguistique), mais également si Akiko est un nom ou un prénom (connaissance culturelle).

ES230504-LSA « Mon \*\*\* est Akiko »

Item : 1) Mon..... est Akiko.

A) **prénom** B) **nom**

**Figure 4** – Item présentant un risque de biais culturel lié à des connaissances culturelles

Cela vaut quelles que soient les langues ou les cultures concernées. A fortiori lorsqu'il s'agit d'une culture française (de France). Dans ce cas, les candidats appartenant à des cultures « proches » (occidentales, européennes) sont avantagés par rapport à des univers culturels plus éloignés. Un exemple symptomatique de biais culturel de ce type est présenté dans la figure 5. Par exemple, l'enseigne d'une pharmacie n'est pas symbolisée dans tous les pays par une croix. Un dessin particulièrement signifiant pour les uns ne le sera donc pas forcément pour les autres.

SP0106-10 COA « Itinéraire en ville »

Dessin A : Une pharmacie qu'on peut identifier d'après son enseigne : une **croix**

**Figure 5** – Item présentant un risque de biais culturel lié à des connaissances culturelles

De la même manière, l'erreur de conception dans l'unité peut renvoyer à une réalité quotidienne (probablement celle du concepteur) mais très localisée. Il est très possible qu'un candidat n'ait aucune expérience de ce dont on parle. Si, comme c'est le cas dans l'unité présentée dans la figure 6, le candidat est censé identifier le lieu où ce type de message est diffusé, le fait qu'un système de métro existe dans sa ville ou dans son pays représente un aspect facilitateur de la compréhension du message par rapport à un candidat qui n'a jamais eu l'expérience de ce type de transport.

ES480905-COB « Nous vous informons qu'en raison d'un incident technique »

Document : « Nous vous informons qu'en raison d'un incident technique, la circulation des **rames** sur les **lignes 2 et 9** est fortement perturbée ».

**Figure 6** – Item présentant un risque de biais culturel lié à des connaissances culturelles

La *Typologie d'erreurs* évoquée précédemment (cf. tableau 2) s'est appuyée sur le recensement des erreurs de conception les plus fréquemment relevées par le Comité d'Évaluation et de Validation. Pour les besoins de l'étude présentée ici, un nouveau recensement des erreurs de conception relevant du type d'erreur « Culture » a été effectué. Les deux catégories prévues dans la typologie ont été développées en sous catégories permettant d'identifier plus précisément les erreurs de conception relatives aux biais culturels. Elles sont présentées dans le tableau 3.

**Tableau 3** – Catégories d'erreurs de conception risquant d'entraîner un biais culturel

Appartenance à un groupe culturel		Connaissances culturelles	
A1	Atteinte aux valeurs du groupe	C1	Renvoi à des faits culturels localisés (non universels)
A2	Atteinte à l'image du groupe		
A3	Atteinte à la sensibilité personnelle	C2	Appel à des savoirs ou savoir-faire spécialisés

Les catégories « Appartenance... » et « Connaissances... » renvoient au profil culturel du candidat. Les sous-catégories d'erreurs numérotées A1, A2, A3 ou C1, C2, C3 sont la résultante de l'interaction « biaisante » entre le profil culturel du candidat et le profil culturel attendu par l'unité de test erronée.

### **3. Les outils de prévention pour éviter les biais culturels**

Pour prévenir les risques d'apparition de biais culturels dans la production des concepteurs, plusieurs outils ont été élaborés : ils sont utilisés aux différents moments de la phase de conception des unités de test.

En amont, dès la phase de recrutement des concepteurs d'items de test et des correcteurs d'épreuves d'expression écrite, la procédure prévoit de déterminer si le candidat est sensibilisé à la question des biais culturels en conception et en correction. Ainsi, dans les tests de recrutement pour la conception, un exercice consiste à identifier les éléments erronés dans une unité de test. Les biais culturels font partie des erreurs de conception à identifier (type C2). Et lors de la commande d'essai (2<sup>e</sup> phase du recrutement), les annotations portées sur la commande à corriger après la 1<sup>re</sup> validation pédagogique, pointent, le cas échéant, les erreurs de conception liées à un biais culturel et propose des pistes de correction. Dans la procédure de recrutement pour la correction des épreuves écrites du TEF, le candidat est interrogé sur sa connaissance des critères parasites de l'évaluation dont les biais culturels font partie. Lorsqu'un candidat à la conception a subi avec succès les tests de recrutement et qu'il intègre le vivier des concepteurs d'épreuves, le Guide du concepteur (cf. tableau 4) qui lui est remis identifie les différents risques de biais auxquels il doit être attentif et indique, pour chaque phase de son travail de conception (choix des documents, rédaction des consignes, etc.) comment les réduire.



En aval, lors des phases de validation pédagogique des commandes, les membres du Comité d'Évaluation et de Validation peuvent s'appuyer, pour mémoire, sur la typologie d'erreurs (cf. tableau 2) qui contient la rubrique « Culture » déjà citée. L'expérience interculturelle des membres du Comité (plusieurs années d'expérience de vie et d'enseignement à l'étranger, dans des contextes culturels variés) vient renforcer leur capacité à identifier des biais culturels dans la production des concepteurs. Lorsque le biais culturel est établi, le Comité d'Évaluation et de Validation propose soit la correction de l'erreur lorsque cela est possible, soit le rejet de l'unité en cause.

Enfin, suite à l'analyse psychométrique des items pré-testés, la DRI/E-CCIP identifie et, lorsque cela est possible, remédie aux items comportant une erreur de conception entraînant un biais culturel.

**Tableau 4 – Extrait du Guide du concepteur sur les biais culturels donnés (CTEF-GUID-01)**

p.	Chapitre	Citation	Type	Commentaires
26	Types de textes en CE	<ul style="list-style-type: none"> <li>documents assez courts tirés de la presse francophone généraliste traitant de sujets variés</li> </ul>	C2 C3	<ul style="list-style-type: none"> <li>francophone : pour éviter le franco-français</li> <li>généraliste : pour éviter la langue de spécialité</li> </ul>
72	LS Section C	<ul style="list-style-type: none"> <li>Éviter les tournures familières : préférez un registre standard ou soutenu (pour les niveaux avancés).</li> </ul>	C2	<ul style="list-style-type: none"> <li>tournures familières : souvent l'expression de particularismes linguistiques ou culturels régionaux</li> <li>tournures familières : langue non stabilisée</li> </ul>
76	EE Section A	<ul style="list-style-type: none"> <li>Éviter cependant les thèmes macabres ou trop sensationnels</li> <li>Éviter les thématiques qui pourraient gêner des candidats pour des raisons culturelles, politiques ou religieuses</li> </ul>	A3 A1/A2	<ul style="list-style-type: none"> <li>macabres : risque d'atteinte à la sensibilité individuelle</li> <li>gêner : risque de préjugés, d'atteinte à l'image, aux valeurs</li> </ul>
78	EE Section B	<ul style="list-style-type: none"> <li>(Permettre) le débat sans choquer ni exalter les candidats.</li> <li>L'opinion du candidat n'a aucune importance : seul compte la façon dont il l'exprime et l'articule</li> <li>Préférer un sujet peu original mais fécond en arguments à un sujet si inattendu qu'il laissera le candidat coi !</li> </ul>	A1 C1	<ul style="list-style-type: none"> <li>risque d'atteinte à l'image, aux valeurs</li> <li>risque de disparités liées à la culture générale</li> </ul>
86	Sélection des documents (écrits et oraux)	<p><u>Variété des sources</u></p> <ul style="list-style-type: none"> <li>non seulement français, mais aussi plus largement francophones</li> </ul> <p><u>Authenticité des documents</u></p> <ul style="list-style-type: none"> <li>masquer les marques commerciales et les noms de « célébrités » cités dans les documents originels par des pseudonymes plausibles</li> </ul> <p><u>Thématique</u></p> <ul style="list-style-type: none"> <li>renvoyant à des réalités francophones pouvant intéresser tous les candidats, pas ou peu spécialisés</li> <li>rejeter des documents faisant référence, de manière explicite ou implicite, à des réalités culturelles, sociales ou politiques nationales, françaises ou parisiennes</li> <li>proposer des documents « politiquement corrects », c'est-à-dire ne portant pas atteinte à la sensibilité nationale, politique ou religieuse des candidats</li> <li>veiller à la neutralité ou l'universalité des thèmes exploités</li> </ul> <p><u>Pérennité</u></p> <ul style="list-style-type: none"> <li>thème choisi ne renvoie pas à un événement trop ponctuel et [...] reste d'actualité à un horizon de trois ans</li> </ul>	C2 C2 C3 C2 A1/A2 A/C C2	<ul style="list-style-type: none"> <li>francophone : pour éviter le franco-français</li> <li>une marque, une célébrité peut n'être connue que localement</li> <li>thématique spécialisée : savoirs et savoir-faire spécialisés</li> <li>éviter des références franco-françaises probablement inconnues des candidats</li> <li>ne risquant pas, pour une raison quelconque, de les perturber ou de les blesser</li> <li>L'actualité « brûlante » est souvent liée au contexte culturel</li> </ul>
88	L'exploitation d'un document long écrit ou oral (Compréhension)	<ul style="list-style-type: none"> <li>[Les questions] ne doivent pas non plus porter sur des implicites culturels ni même des connaissances culturelles si celles-ci ne sont pas explicitées dans le document</li> <li>Souvenez-vous que la règle d'or est que le candidat ne doit se déterminer qu'en fonction de la seule compréhension qu'il a du document et de l'item</li> </ul>	C1/C2 C	<ul style="list-style-type: none"> <li>L'explicitation est une remédiation à l'implicite culturel</li> <li>La tâche ne doit pas être réalisable en faisant appel à d'autres compétences que celle visée par l'objectif</li> </ul>
90	Principes de rédaction d'items	<ul style="list-style-type: none"> <li>Lorsque les items s'appuient sur un texte, il ne doit pas être possible d'y apporter une réponse juste seulement fondée sur la culture générale sans avoir lu et compris le texte</li> </ul>	C1	<ul style="list-style-type: none"> <li>risque de disparités liées à la culture générale</li> </ul>
94	Critères de validation	<p>Thématique</p> <ul style="list-style-type: none"> <li>Le document proposé développe une thématique acceptable</li> </ul>	A	<ul style="list-style-type: none"> <li>risque de préjugés, d'atteinte à l'image, aux valeurs</li> </ul>
95	Typologie d'erreurs	<p>Culture</p> <ul style="list-style-type: none"> <li>Thématique culturellement / politiquement sensible pour certains candidats.</li> <li>Référence à un particularisme culturel / politique / social / économique non explicité.</li> </ul>	A C	

## **4. De l'évaluation quantitative des biais à leur correction**

### ***Au niveau de la conception et de la validation pédagogiques***

Les données statistiques recueillies pour cette étude résultent du relevé des erreurs de conception de la catégorie « Culture » commentées par le Comité d'Évaluation et de Validation lors des phases de validation pédagogique des commandes. Elles concernent, d'une part, les « commandes d'essai » proposées aux candidats à la conception d'items et, d'autre part, les commandes TEF « normales » passées auprès des concepteurs du TEF. Il a semblé instructif de relever, au delà du nombre d'unités et d'items contenant des erreurs, la localisation de ces erreurs, soit sur le document support (lorsque l'unité en comporte un), soit sur l'item (une unité peut en compter plusieurs). Ces données apportent un éclairage sur l'efficacité des outils de prévention à disposition.

On constate, par exemple, que le taux d'erreurs de conception va en s'amenuisant, au fur et à mesure que les concepteurs acquièrent de l'expérience. Ainsi pour les commandes d'essai réalisées par des concepteurs « novices », les taux d'erreurs sont deux à trois fois plus élevés que ceux des commandes TEF, réalisées par des concepteurs réguliers. Cette réduction du taux d'erreurs entre commandes d'essai et commandes TEF s'explique également par le fait que les outils de formation tels que le Guide du concepteur n'est mis à la disposition des concepteurs que lorsque ces derniers sont effectivement engagés et donc tenus par leur engagement de confidentialité. On fait également le constat que le taux d'erreurs diminue entre la première et la seconde validation pédagogique des commandes (commandes d'essai et commandes TEF), ce qui s'explique par le fait que les erreurs relevées par le Comité d'Évaluation et de Validation en première lecture sont généralement corrigées par le concepteur avant la seconde lecture.

Les données collectées permettent également d'établir des taux d'occurrence d'erreurs de conception unité-type (toute unité-type renvoie à un niveau, un objectif, une tâche) par unité-type et concepteur par concepteur. Il est ainsi possible de déterminer précisément quelles unités présentent le plus de risques d'erreurs et pour quel concepteur. Ces informations viennent donc réalimenter les outils de suivi et de formation mis en place pour le contrôle de la qualité. A titre d'exemple, on peut se pencher sur les épreuves d'expression écrite et orale. Pour l'épreuve d'expression écrite, l'unité-type de la section A, qui correspond à l'objectif « rédiger un récit » présente un risque d'erreur sensiblement plus élevé que celle de la section B (argumenter un point de vue). Pour l'épreuve d'expression orale, c'est l'inverse : les unités les plus « risquées » sont celles de la section B dont l'objectif est de présenter un document et de soutenir son point de vue.

D'une manière générale, on peut émettre l'hypothèse que les unités faisant appel à un document-support à caractère authentique sont plus porteuses de risques que les unités dépourvues de document-support ou s'appuyant sur des documents-supports didactisés.

## Au niveau des traitements psychométriques

Lors de l'analyse des traitements psychométriques des résultats aux pré-tests, on constate que le manque de discrimination ou le pourcentage de réussite de certains items peuvent s'expliquer par la présence de biais culturels. Le tableau 5 propose un recensement des biais culturels décelés à l'issue des pré-tests pour les 12 derniers jeux.

**Tableau 5 – Relevé des erreurs de conception de type culturel**  
Pré-tests - Période de janvier 2004 à juin 2006

Nombre de pré-tests	12
Biais culturels en compréhension écrite	4
Biais culturels en compréhension orale	4
Biais culturels en lexique et structure	3
Nombre total de biais culturels	11
Nombre total d'items	1386
Fréquence des biais	0,8%

Il permet d'estimer à 1% (soit en moyenne un item par jeu de pré-test) la part des items pré-testés présentant des biais culturels. Contrairement à la phase de conception, ces items se répartissent de façon générale entre les trois épreuves obligatoires. Sur ces onze items, huit avaient un indice de discrimination (R.bis) compris entre -0,22 et 0,19. Les trois autres items avaient un indice de discrimination compris entre 0,31 et 0,37, mais le taux de bonnes réponses était nettement supérieur ou nettement inférieur au taux escompté compte tenu du niveau présumé de l'item. Voici un exemple d'item de compréhension orale (cf. figure 7) qui présente un biais culturel décelé lors de l'analyse psychométrique : dans ce cas précis, il semble qu'acheter un billet d'avion dans un grand magasin ne soit pas une valeur partagée par tous.

2TF0405 – COB  
Script :

*Partez en vacances avec l'agence de voyages d'Uniprix ! Préparez vos vacances d'été ! Vous désirez partir au Sahara ? En Espagne ? Pas de problème ! Cette semaine, nous offrons jusqu'à 75 euros de réduction sur les billets Air France. **Au premier étage, près de l'espace librairie, nos hôtesses vous accueillent pour répondre à vos questions.** Alors, ne dépensez plus des fortunes pour partir !*

Ce message est diffusé

**A.** dans un aéroport  
**B.** dans un grand magasin  
**C.** au guichet d'Air France  
**D.** dans une agence de voyage

**Figure 7 – Item de compréhension orale présentant un biais culturel décelé à l'analyse psychométrique**

### ***Au niveau du retour d'information des centres agréés***

Un troisième contrôle de la qualité est également opéré sur les unités du TEF. Ce contrôle est réalisé grâce à des analyses a posteriori sur les épreuves obligatoires et facultatives (calcul de la corrélation inter-juges pour les correcteurs et les examinateurs). Le retour d'informations des centres agréés faisant passer le TEF est également un moyen efficace de remédier à ces biais culturels et de rétroagir. Ce retour d'informations se fait sur des feuilles d'anomalies que les centres agréés TEF retournent à l'issue de la session à la DRI/E-CCIP.

Les retours d'information les plus fréquents concernent l'épreuve d'expression orale du TEF. L'épreuve étant passée dans le centre agréé TEF et animée par des examinateurs locaux, les retours d'informations sur cette épreuve sont donc plus nombreux. Par exemple, l'unité que constitue la section B de l'épreuve (2<sup>e</sup> sujet de l'épreuve) peut poser quelques difficultés d'ordre culturel pour des candidats peu habitués à défendre leur point de vue. Dans ce cas précis, les habitudes éducatives peuvent influencer la capacité du candidat à s'exprimer. Les biais culturels sont également décelables dans l'échange oral guidé par l'examineur et dans son interprétation des critères inscrits dans la grille d'évaluation. Par exemple, on constate que les écarts entre deux examinateurs sont plus importants sur les critères communicatifs que sur les critères linguistiques.

## **Conclusion**

Les résultats de cette étude sur les biais culturels révèlent que les biais peuvent être de plusieurs natures : ils peuvent être liés aux valeurs, à l'image, à la sensibilité personnelle et aux connaissances culturelles du candidat, du correcteur et de l'examineur. Les conclusions établies par les auteurs quant à la catégorisation des biais culturels à partir d'une base de données d'items rejoignent celles proposées par Bachman (1995).

Ce premier regard posé sur les biais culturels dans les évaluations internationales de français langue étrangère a permis de mettre en évidence les nombreux paramètres qui pourraient influencer l'évaluation du candidat, du correcteur et de l'examineur. Des études complémentaires sont actuellement en cours : il s'agit à l'aide de la méthode DIF (differential item functioning) de poursuivre ces travaux sur les biais culturels en croisant le regard du linguiste avec celui du psychométricien.

L'enjeu de la démonstration est de taille : au delà de la nécessaire standardisation, il s'agit bien de proposer un test équitable pour tous les candidats. Cette notion d'équité est essentielle pour les politiques d'accueil et de suivi des flux de populations allophones dans un contexte de plus en plus marqué par l'internationalisation des cursus d'études supérieures et des parcours professionnels.

## Bibliographie

BACHMAN, L. (1995). *Fundamental Considerations in Language Testing* (3<sup>e</sup> édition). Oxford: Oxford University Press.

CALVET, L.-J. (1987). *La guerre des langues et les politiques linguistiques*. Collection Pluriel. Paris : Hachette Littératures.

CENTRE DES NIVEAUX DE COMPÉTENCE LINGUISTIQUE CANADIENS (2002). *Standards linguistiques canadiens 2002*. Ottawa : Citoyenneté et Immigration Canada.

CENTRE DES NIVEAUX DE COMPÉTENCE LINGUISTIQUE CANADIENS (2006). *Niveaux de compétence linguistique canadiens 2006 : Français langue seconde pour adultes*. Ottawa : Citoyenneté et Immigration Canada.

CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS (2006) : *Le Test d'Évaluation 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.

CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS – DIRECTION DES RELATIONS INTERNATIONALES DE L'ENSEIGNEMENT (2003) : *Adéquation du Test d'Évaluation de Français aux critères d'évaluation de Citoyenneté et Immigration Canada*. 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.

DE LANDSHEERE, G. (1992) : *Dictionnaire de l'évaluation et de la recherche en éducation* (2<sup>e</sup> édition). Paris : Presses Universitaires de France.

DEMEUSE, M., CRENDAL, A., DESROCHES, F., OSTER, P., RENAUD, F., LEROUX, X. (2004). *L'évaluation des compétences linguistiques des adultes en français langue étrangère dans une perspective de multi-référentialisation. L'exemple du Test d'Évaluation 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'Évaluation en Education (ADMEE-Europe). Lisbonne, 18-20 novembre 2004

DEMEUSE, M., DESROCHES, F., CRENDAL, 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'Évaluation en Education (ADMEE-Europe), Reims, 24-26 octobre 2005.

DESROCHES, F., CASANOVA, D., CRENDAL, A., RENAUD, F., MAHY, C., DEMEUSE, M. (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'autoévaluation*, In Actes du 18<sup>e</sup> colloque international de l'Association pour le Développement des Méthodologies d'Évaluation en Education (ADMEE-Europe), Reims, 24-26 octobre 2005

GOUGENHEIM, G., MICHEA, R., RIVENC, P., SAUVAGEOT, A. (1956). *L'élaboration du français élémentaire : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris : Didier.

MOUNIN, G. (1975). *Clefs pour la linguistique*, Editions Seghers.