

PROBLÈMES MÉTHODOLOGIQUES SOULEVÉS PAR UNE ENQUÊTE INTERNATIONALE – *L'ENQUÊTE INTERNATIONALE SUR L'ALPHABÉTISATION DES ADULTES*

Alain Blum, France Guérin-Pace¹

RÉSUMÉ

L'Enquête internationale sur l'alphabétisation des adultes (EIAA), coordonnée par Statistique Canada, a été menée dans une vingtaine de pays entre 1994 et 2000. Ses résultats ont donné lieu à la publication en 2000, sous l'égide de l'OCDE, d'un rapport de synthèse proposant une analyse comparative des capacités de lecture dans les pays participants. Or, cette enquête soulève des problèmes méthodologiques très importants qui la rendent inutilisable à des fins comparatives.

La présente communication porte sur les faiblesses de l'enquête et, de manière plus générale, sur les écueils que suscite la mise en œuvre d'une mesure universelle des compétences dans des pays de culture et de langue différentes. Une série d'analyses des résultats de l'enquête montre à quel point la traduction et la formulation des questions dans les différentes langues influent sur la difficulté des questions du test. D'un pays participant à l'autre, la hiérarchie de réussite aux différents items confirme ce biais linguistique. De plus, si l'attitude des répondants (motivation, degré d'attention, refus, etc.) est marquée par la culture et étroitement liée à la tradition de chaque pays en matière d'enquêtes, elle constitue aussi un facteur déterminant dans une enquête destinée à construire une mesure des compétences. Seule une codification très précise permettrait de cerner l'attitude des répondants face à une enquête longue et difficile et de distinguer une absence de compétences d'un manque de motivation ou d'attention.

La présente communication est fondée sur une évaluation de l'enquête menée au sein de la Communauté européenne. Ses résultats ont donné lieu à la publication d'une série d'articles et d'un ouvrage sur le sujet.

MOTS-CLÉS : Alphabétisation, erreur de mesure, enquêtes comparatives.

1. PRÉSENTATION ET GENÈSE DE L'EIAA

Lancée en 1994, l'Enquête internationale sur l'alphabétisation des adultes (EIAA) illustre une démarche comparative qui est maintenant souvent adoptée par l'Union européenne ou par des organismes internationaux. L'EIAA avait pour objectif de mesurer les compétences nécessaires pour lire et comprendre des documents utilisés dans la vie courante (mode d'emploi d'un appareil ménager, posologie d'un médicament, articles d'actualité, représentation graphique de statistiques descriptives, indicateur d'un service de transport, etc.).

L'EIAA a été coordonnée par Statistique Canada, qui en a assuré la mise en œuvre à l'échelle internationale en 1992, en collaboration avec l'organisme privé américain ETS (Education Testing Service) et l'OCDE. Huit pays ont pris part au premier cycle d'enquêtes : le Canada, les États-Unis, la France, l'Allemagne, les Pays-Bas, la Pologne, la Suède et la Suisse. Un autre cycle d'enquêtes a eu lieu en 1996 en Grande-Bretagne, en Irlande du Nord, en Nouvelle-Zélande et en Irlande. L'enquête a été menée dans d'autres pays de l'OCDE (OCDE, 2000). Au total, 21 pays ont participé à l'enquête, dont 19 pays de l'OCDE.

Les résultats de l'enquête internationale ont été publiés sous forme de tableaux comparatifs présentant la répartition de la population selon cinq niveaux d'alphabétisation en ordre croissant. Le niveau 1, le plus

¹ INED, 133, boul. Davout, 75020 Paris, France. blum@ined.fr; guerin@ined.fr. La présente communication est une synthèse des articles énumérés dans la bibliographie.

faible, correspond aux personnes presque analphabètes, à peine capables de déchiffrer des phrases ou d'effectuer des calculs simples. Le niveau le plus élevé (niveau 5) correspond à celles qui n'ont aucune difficulté à comprendre des articles et des textes non spécialisés, qui peuvent écrire des lettres, effectuer des calculs, etc.

Selon cette méthode, la population de chaque pays participant à l'enquête est répartie sur l'échelle d'alphabetisation, ce qui permet de comparer les proportions des personnes qui se situent aux divers niveaux d'alphabetisation. Largement diffusées, ces données ont fait l'objet d'une couverture abondante dans la presse et auprès des personnalités politiques. Dans le cas de la France, cette attention des médias et des personnalités politiques a été accentuée par le fait que les trois quarts des Français se situaient à des niveaux d'alphabetisation trop faibles pour leur permettre d'accomplir des tâches courantes telles que lire un journal, écrire une lettre, déchiffrer un texte court ou une fiche de paie. Sur les échelles utilisées par les concepteurs de l'EIAA, 75 % des adultes français se situaient à un niveau d'alphabetisation estimé à 1 ou 2 en compréhension de textes suivis, contre 46 % des Américains, 40 % des Néerlandais et 28 % des Suédois. En compréhension de textes schématiques, les proportions d'adultes aux niveaux 1 et 2 étaient respectivement de 63 %, 50 %, 36 % et 25 %, alors qu'en compréhension de textes au contenu quantitatif, les proportions pour les mêmes pays étaient de 57 %, 46 %, 35 % et 25 %. En ce qui concerne le niveau d'alphabetisation le plus élevé (niveau 5), seulement 11 des 3 000 personnes ayant pris part à l'enquête en France possédaient les compétences correspondant à ce niveau, soit beaucoup moins que le nombre de répondants qui avaient fait des études supérieures.

2. LES LIMITES DE LA MÉTHODE COMPARATIVE

La valeur de l'EIAA à titre de méthode comparative et, de façon plus générale, la validité et le caractère significatif de ses résultats ont fait l'objet de plusieurs études. Dans la présente communication, nous nous concentrons sur la validité des comparaisons d'après un examen des données au niveau individuel et des résultats généraux de l'enquête dans les 13 pays susmentionnés.

Nous examinons notamment trois questions importantes :

- Les questions restent-elles de difficulté équivalente après avoir été traduites? Autrement dit, une bonne traduction des items assure-t-elle nécessairement un test équivalent d'une langue à l'autre?
- Peut-on supposer qu'une population présente un comportement uniforme aux niveaux régional et national? Nous nous interrogeons surtout sur la motivation des répondants à l'égard d'un questionnaire qui est long à remplir et qui vise à mesurer la capacité individuelle.
- La mesure synthétique fondée sur les notes est-elle robuste? Plus précisément, une autre mesure fondée sur des documents semblables produit-elle des résultats équivalents?

2.1 Comparaison, niveau de difficulté et traduction

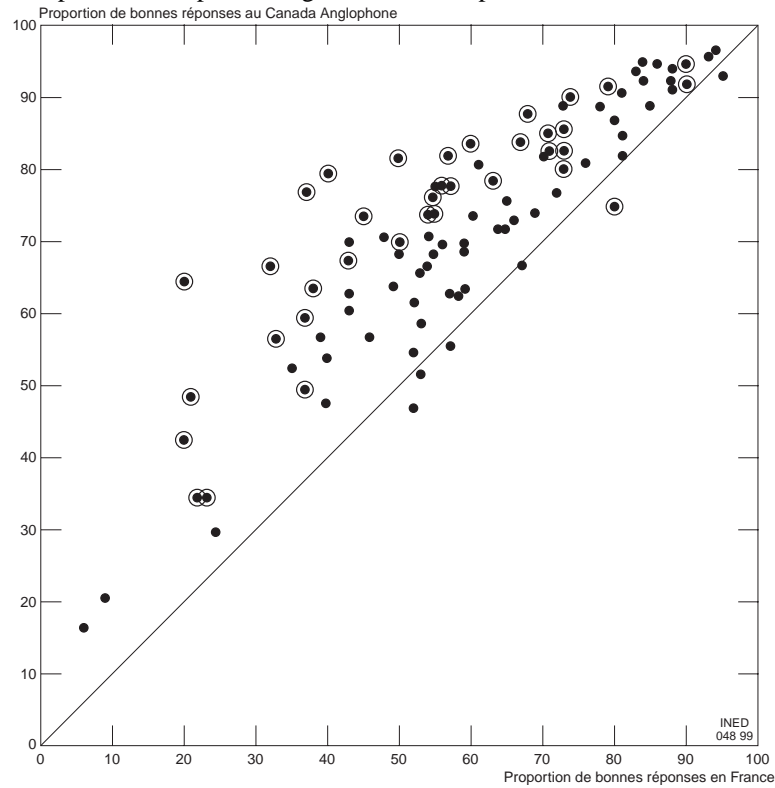
La validité de l'EIAA repose sur l'hypothèse hardie selon laquelle il existe une échelle de difficulté identique pour évaluer les compétences dans toutes les cultures et toutes les langues. Afin de vérifier cette hypothèse, nous avons comparé, question par question, les taux de réussite observés dans deux pays à la fois. Si l'on présume que les questions sont de difficulté équivalente dans les deux pays considérés, la représentation graphique du pourcentage de bonnes réponses à chaque question devrait ressembler à une ligne droite dont la pente est fonction du niveau d'alphabetisation global. Or, l'examen des graphiques montre que les grappes sont de formes extrêmement diverses. En comparant les pourcentages de réussite en France et au Canada anglophone – dont le questionnaire est à l'origine de la version française – nous avons constaté de grandes variations des taux de réussite pour un ensemble de questions et nous avons tenté d'expliquer ces disparités (figure 1).

Tout d'abord, une comparaison de la formulation des questions dans les deux langues a révélé des écarts importants dans la traduction. De manière plus générale, nous avons examiné l'ensemble du questionnaire pour repérer les questions dont la formulation n'était pas « équivalente » en français et en anglais. Nous

avons repéré 35 questions (encerclées dans la figure 1) présentant des divergences de traduction que nous avons analysées sous trois rubriques².

Répétition de termes : Dans le texte qui renferme la réponse, la répétition de termes tirés de la question est plus fréquente dans les documents originaux (en anglais) et représente la première source de biais. Elle a pour effet d'attirer plus rapidement l'attention du répondant anglophone sur la phrase qui renferme la réponse.

Figure 1 : Comparaison des pourcentages de bonnes réponses – France et Canada anglophone



Précision supérieure des termes anglais : En général, la formulation des questions est plus précise en anglais. Par exemple, une question qui se lit en anglais « *What is the most important thing to keep in mind?* » [Quelle est la chose la plus importante à garder à l'esprit?] est traduite en français comme suit : « Que doit-on avoir à l'esprit? » Toutefois, la phrase qui renferme la réponse se lit en anglais « *the most important thing* » et, en français, « la chose la plus importante ». Le lien entre la question et la phrase est donc plus facile à établir en anglais.

Erreurs de traduction : Certaines erreurs de traduction, quoique souvent mineures d'un point de vue strictement linguistique, peuvent avoir une incidence importante sur la compréhension. Par exemple, « *If you wanted to more than double your principal within five years, what rate...* » est traduit en français comme suit : « ...doubler votre capital *en moins* de cinq ans... ». La traduction de « *within five years* » par

² On trouvera dans Guérin-Pace et Blum (2000) la formulation de toutes ces questions et le détail des problèmes liés à chacune d'elles.

« en moins de cinq ans » a pour effet d'exclure une durée de cinq ans en français, et il n'y a pas de solution à la question dans le tableau correspondant. Or, on a attribué une mauvaise réponse aux personnes ayant indiqué qu'il n'y avait pas de réponse à cet item.

Autres sources d'erreur : On peut repérer des sources d'erreur qui, bien que moins répandues, révèlent la complexité d'une définition de l'équivalence de difficulté entre des items exprimés dans des langues différentes. L'exemple qui suit est typique; on peut l'attribuer à un « biais culturel ». L'exercice consiste à déterminer, dans un article portant sur quatre films, lesquels sont des comédies. En anglais et en français, l'article mentionne le mot « comédie » à propos de deux films, ce qui simplifie la question. En France, toutefois, on a constaté que bon nombre d'interviewés avaient donné comme réponse un troisième film même si, d'après la description, il ne s'agissait pas d'une comédie. L'explication est la présence, dans ce film, de l'acteur Michel Blanc, bien connu du public français pour ses rôles dans plusieurs comédies, mais peu connu à l'étranger. Dans ce cas, le processus de réponse était dominé par l'association d'idées, au détriment d'une lecture attentive de l'article.

2.2 Classement des pays selon le profil de réussite par item

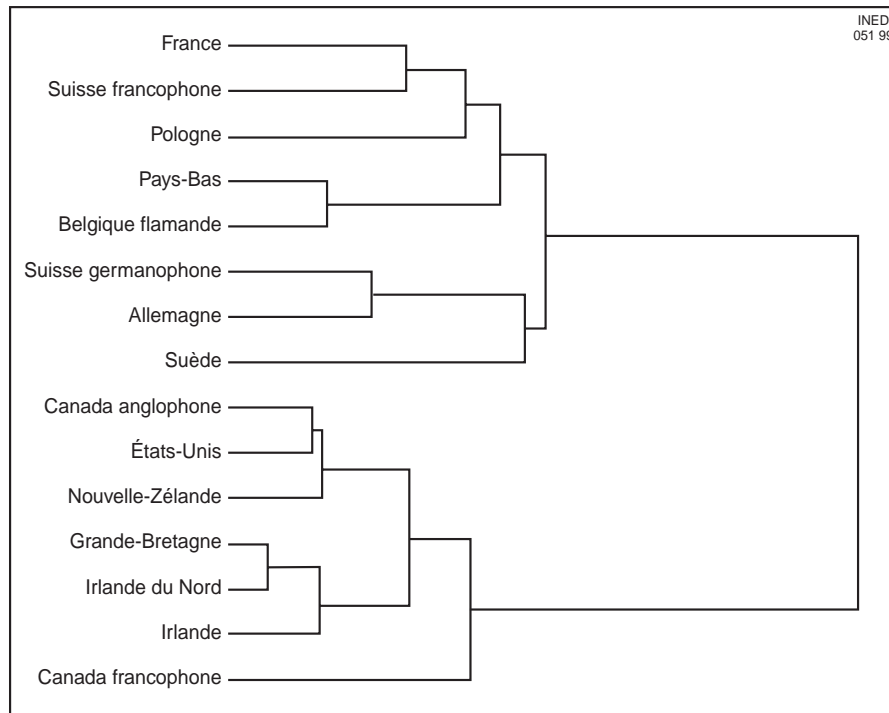
On observe également dans d'autres pays que la France des divergences entre le niveau de difficulté des questions définies à l'avance et le niveau de réussite. Diverses mesures du degré de similitude des hiérarchies de réussite montrent que malgré un haut degré d'association, elles diffèrent considérablement d'un pays à l'autre. Pour mettre l'accent sur les similitudes en fonction de niveau relatif de réussite aux questions sans égard au pourcentage de bonnes réponses à chacun des items, nous établissons, à l'aide d'une analyse typologique, une répartition des pays selon le rang de chacune des questions dans la hiérarchie de réussite. Sans égard à la répartition de la population par niveaux d'alphabétisation, on observe parmi les divers pays une corrélation entre l'ordonnement des questions, la langue de l'interview et le pays dans lequel l'enquête a été menée³.

Le contenu des grappes montre que les parallèles établis sont fondés à la fois sur la proximité géographique et sur la proximité linguistique (figure 2). Tous les pays anglophones sont regroupés en une seule catégorie. Ainsi, les États-Unis sont regroupés d'abord avec le Canada anglophone, puis avec la Nouvelle-Zélande; la Grande-Bretagne, l'Irlande du Nord et l'Irlande du Sud forment une catégorie. Ces derniers pays ont utilisé la même version du questionnaire. Les deux groupes sont combinés et forment ensuite une seule catégorie avec le Canada francophone. Une autre catégorie comprend les pays européens non anglophones, répartis selon la langue du questionnaire. La France et la Suisse francophone forment une seule catégorie, l'Allemagne et la Suisse germanophone en forment une autre et la Belgique flamande et les Pays-Bas, une troisième. La Suède demeure isolée avant d'être ajoutée à la catégorie formée par l'Allemagne et la Suisse germanophone.

Ce résultat global contredit nettement les principes de base du protocole de cette enquête, qui tiennent pour acquis que le rendement est indépendant de la langue du questionnaire.

³ Ici, $r_{i,j}$ est le rang de l'item i (du niveau de réussite le plus élevé au plus faible) dans le pays j ; les analyses typologiques sont faites sur le tableau $R(i,j) = (r_{ij})$

Figure 2 : Classement des pays selon la hiérarchie de réussite, sans égard aux questions abandonnées



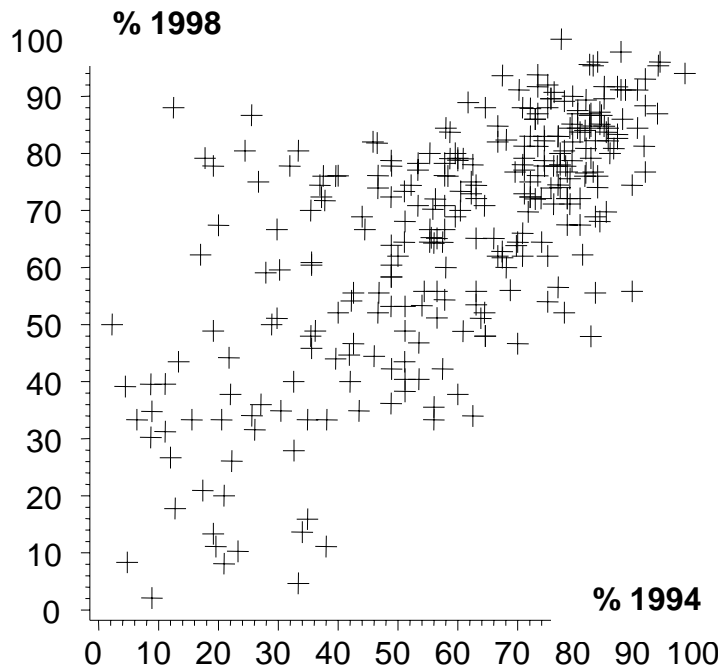
2.3 Effets du deuxième test et de la traduction

On a procédé à un deuxième test (voir Carey, 2000) en interviewant de nouveau, en 1998, un échantillon de Français déjà interviewés en 1994. On a interviewé près de la moitié de l'échantillon (300 répondants) en utilisant le questionnaire original en français, et l'autre moitié (422 répondants) à l'aide de la version suisse du questionnaire⁴. Il s'agissait avant tout d'évaluer l'influence qu'exerçait l'utilisation d'une traduction différente, en sachant qu'une partie des problèmes présents dans le questionnaire français ne se trouvaient pas dans la version suisse.

Tout d'abord, les écarts entre le profil de réponse de 1994 et celui de 1998 révèlent une grande instabilité des résultats (figure 3). Cette variabilité importante remet en question la fiabilité de cette mesure de l'alphabétisation.

⁴ Une première analyse montre que les deux échantillons sont presque semblables sur le plan de la structure socio-démographique. L'utilisation de poids calculés en fonction de chaque échantillon n'a aucune incidence sur le profil de réponse des items.

Figure 3 : Proportion de bonnes réponses par personne en 1994 et 1998
(personnes ayant répondu au questionnaire français en 1998)



Deuxièmement, le résultat révèle un effet important de la traduction, et plus particulièrement son incidence sur la variation des « mauvaises » réponses entre 1994 et 1998. Pour presque tous les items, la proportion de bonnes réponses est beaucoup plus importante lorsqu'on utilise le questionnaire suisse au lieu du questionnaire français (Blum et Guérin, 2000). Le résultat révèle un effet positif important de la traduction sur la proportion de bonnes réponses par personne. Autrement dit, le fait d'utiliser en 1998 le questionnaire suisse, au lieu du questionnaire français, augmente considérablement la probabilité d'obtenir une bonne réponse à un item qui présentait un problème de traduction en 1994.

2.4 Analyse des réponses manquantes

Les conséquences de l'inattention et du manque d'intérêt de la part des interviewés à l'égard d'un long questionnaire exigeant une grande concentration sont très importantes. Néanmoins, les résultats obtenus dans les divers pays européens et nord-américains laissent supposer qu'il s'agit d'une question importante. Le pourcentage de la population qui se situe aux niveaux d'alphabétisation 1 et 2 varie entre 28 % (en Suède) et 77 % (en Pologne) pour les tâches liées aux textes suivis. Ces pourcentages sont considérables, non seulement en France mais aussi dans les autres pays : les tâches de niveau 1 exigent du lecteur qu'il « repère un élément d'information contenu dans le texte qui est identique ou synonymique à l'information donnée dans la directive » et celles de niveau 2, qu'il « repère un ou plusieurs éléments d'information dans le texte; ce dernier peut contenir plusieurs éléments de distraction, ou le lecteur peut devoir faire des déductions de faible niveau » (Statistique Canada, OCDE, 1995). À l'autre extrémité de l'échelle de capacités, le pourcentage de personnes qui se situent au niveau 5 est tellement faible qu'il n'a pas été publié : on a regroupé les niveaux 4 et 5 dans la même catégorie dans toutes les publications et dans la base de données de l'EIAA. Les tâches de niveau 5 « exigent du lecteur qu'il recherche de l'information dans un texte dense qui contient un certain nombre d'éléments de distraction plausibles ».

Ces résultats étonnants nous incitent à analyser de manière plus détaillée le sérieux avec lequel les personnes interviewées ont répondu à l'enquête. Nous avons mené cette étude à l'aide des données

françaises, en examinant les conséquences de ce manque d'intérêt sur les différenciations régionales et en nous posant la question suivante : Peut-on déceler, dans les différentes régions de France, des attitudes différentes à l'égard de l'enquête?

Nous avons examiné la théorie de la motivation inégale des interviewés par région, dont témoigne le comportement différent dans la façon de répondre aux questions. Par exemple, face à une question qu'ils ne comprennent pas ou qu'ils trouvent difficile, certains interviewés tentent de donner une réponse, même si elle est mauvaise, alors que d'autres sautent la question et passent à la suivante. Si l'interviewé omet une question à cause de sa difficulté, il convient de considérer l'omission comme une mauvaise réponse, car la raison de l'omission est certainement la crainte d'échouer. Toutefois, on ne peut pas interpréter ainsi toutes les omissions, car certaines personnes interviewées choisissent librement de ne pas répondre par manque d'intérêt ou par paresse, mais pas parce qu'elles ne comprennent pas.

Les concepteurs de l'EIAA ont pris deux mesures pour tenter de surmonter cet écueil. Les directives remises aux intervieweurs mentionnent que ces derniers doivent informer les répondants, dès qu'ils lisent une question, que s'ils ne connaissent pas la réponse, ils doivent apposer une marque quelconque sur la question (tracer une croix, raturer les lignes prévues pour la réponse, etc.). On considère alors qu'il s'agit d'une mauvaise réponse. Il ne doit pas y avoir de marque si la question est réellement omise, c.-à-d. si l'interviewé ne lit pas le document parce qu'il a abandonné.

Nous avons ensuite tenté de cerner les omissions ayant pu être interprétées comme des mauvaises réponses. Comme il a été constaté plus haut, un seul document de base sert à poser plusieurs questions. Le document et les questions qui s'y rapportent sont disposés côte à côte. Souvent, en examinant les questionnaires remplis par les interviewés français, on découvre des pages entières raturées, même dans les questionnaires remplis avec succès ailleurs. Il est vraisemblable que certaines réponses codées « mauvaises » révèlent en fait un manque d'intérêt et de motivation plutôt qu'un manque de capacités.

Nous avons montré ailleurs (Guérin-Pace et Blum, 2000) qu'en France, la carte régionale des omissions (omissions partielles ou abandon avant la fin du test) et celle des « refus » étaient complémentaires, ce que rien n'aurait pu nous laisser supposer. Les régions présentant un grand nombre de réponses « raturées » enregistrent un petit nombre de non-réponses, et inversement. Cette constatation laisse supposer que les réponses « raturées » correspondent en réalité à une forme de non-réponse.

Les déterminants des divers types de réponse, calculés à l'aide de différents modèles de régression logistique, confirment ce résultat. Chacune des trois variables – nombre de « mauvaises réponses » par interviewé, nombre de « refus » et nombre d'omissions – était expliquée en fonction des caractéristiques individuelles suivantes : région de résidence, âge, compétences, note obtenue par question et temps nécessaire pour remplir le questionnaire. Les déterminants des mauvaises réponses ont plutôt une origine socio-démographique, les compétences ayant (avec l'âge) une influence importante. La variable « refus » est plus proche de la variable « omission » que de la variable « mauvaise réponse » (Guérin-Pace, Blum, 2000).

Cette analyse a donc montré que rien ne permettait de supposer l'existence d'une attitude homogène à l'intérieur d'un même pays, et encore moins entre les pays. Nous n'avons pas fait d'évaluations semblables pour les autres pays, mais nous croyons que ces différences existent pour les mêmes raisons, ce qui remet encore en question la nature comparative de l'enquête. Les variations de la motivation d'une personne à l'autre deviennent elles-mêmes un élément de biais important qui modifie notre estimation du niveau d'alphabétisation de chaque personne. Bon nombre d'échecs résultent probablement d'un manque de sérieux ou d'intérêt face à une longue enquête. De plus, on ne peut pas vraiment effectuer une analyse comparative puisque ce manque d'intérêt peut être différent d'un pays à un autre et qu'il peut être exprimé différemment sur les questionnaires.

3. AUTRES MÉTHODES D'ÉCHELONNEMENT

La supposition qui sous-tend l'échelonnement de la réponse par item dans le cas de l'EIAA n'est pas la seule qu'on puisse utiliser. Pour être robuste, une mesure doit être confirmée par d'autres mesures, fondées sur les mêmes principes et les mêmes documents. Dans l'EIAA, un niveau individuel est fondé sur une moyenne pondérée des réponses à tous les items d'un domaine. Il y a toutefois bien des façons différentes d'attribuer des niveaux en fonction de la *tendance* des réponses. Une solution très simple consiste à définir le niveau en fonction de l'item le plus difficile auquel on a répondu correctement. Elle est liée à une simple échelle de Guttman où l'on suppose qu'une personne ayant répondu correctement à une question d'un niveau donné est capable de répondre à une autre question de niveau inférieur.

C'est de cette manière que nous avons estimé de nouveau les notes individuelles des interviewés de l'EIAA qui n'avaient pas été imputées. Nous définissons ainsi le « profil d'alphabétisation » suivant : il s'agit d'un ensemble de cinq chiffres d_1 à d_5 ; d_i est égal à 1 si l'on a répondu correctement à au moins une question de niveau i (aux termes de l'EIAA); sinon, il est égal à 0. Puis, un interviewé se situe au niveau i si d_i est égal à 1, d_{i+1} à d_5 est égal à 0 et d_1 à d_{i-1} est égal à 0 ou à 1. Les ensembles $(d_i)_{i=1,5}$ sont appelés profils d'alphabétisation. On dit que ces profils sont *cohérents* si d_1 à d_{i-1} sont tous égaux à 1, et *incohérents* si au moins un des chiffres $i-1$ d_1 à d_{i-1} est égal à 0. Dans ce dernier cas, cela signifierait qu'une personne est considérée comme se situant au niveau i , mais qu'elle a échoué à une tâche de niveau inférieur. Parmi ces profils incohérents, un profil est dit « faiblement incohérent » si le seul niveau d'échec se situe juste au-dessous du niveau d'alphabétisation. La cohérence est élevée : 91 % des profils sont cohérents en France, 94 % en Grande-Bretagne. De plus, lorsque les profils sont incohérents, il s'agit souvent d'une incohérence faible.

Lorsqu'on utilise cette mesure, la répartition du niveau d'alphabétisation est complètement différente de celle de l'EIAA. Selon la mesure de l'EIAA, 65 % des interviewés français dont les notes n'ont pas été imputées se situent au niveau 1 ou 2 en compréhension de textes suivis; selon la mesure fondée sur un « niveau supérieur » de réussite, la proportion tombe à 5 %. Pour la Grande-Bretagne, les proportions sont respectivement de 48 % au niveau 1 ou 2 selon la mesure de l'EIAA et de 3 % au même niveau selon la mesure de « niveau supérieur ».

En France, parmi les personnes qui se situent au niveau 1 selon l'EIAA, 8 % restent à ce niveau selon la mesure « supérieure », 9 % accèdent au niveau 2, 56 % au niveau 3 et environ 18 % au niveau 4 ou 5. Ces variations montrent qu'on arrive à des conclusions entièrement différentes lorsqu'on utilise un méthode d'échelonnement différente.

En présentant cette méthode d'échelonnement très simple, nous ne voulons pas proposer une autre mesure de l'alphabétisation, mais montrer comment, en utilisant les pourcentages réels d'items répondus correctement tels qu'ils sont regroupés en niveaux par l'EIAA, une autre méthode d'échelonnement des personnes produit des résultats différents.

4. CONCLUSION

Il existe donc des doutes sérieux quant à la valeur de l'EIAA comme outil de comparaison internationale. L'étude de ces problèmes porte à formuler des conclusions importantes concernant des enquêtes comparatives plus générales, d'un intérêt qui va au-delà de la seule EIAA. La traduction et l'adaptation en différentes langues, ainsi que l'attitude de différentes populations nationales à l'égard de ce type d'enquête, entraînent des disparités dans les mesures qu'on ne peut interpréter en fonction de l'alphabétisation. La production de plusieurs mesures synthétiques éclipsent la complexité des phénomènes à l'étude et fait oublier le processus qui mène à la construction de la mesure.

Dans le cas de l'EIAA, il semble qu'on ait étendu précipitamment le modèle américain à un contexte international. À elle seule, la collaboration intervenue entre les organismes internationaux ne justifiait pas la dimension internationale qu'on a donnée à l'enquête. La conception et le contenu du questionnaire étaient

presque exclusivement empruntés à ce qui était à l'origine une enquête nord-américaine. Il ne s'agissait donc pas du produit d'une comparaison entre les réalités nationales différentes à l'égard de ces questions. Pire encore, lorsque les méthodes statistiques employées sont mises à l'essai dans un contexte national, elles s'avèrent incapables de valider la valeur comparative de l'enquête.

De plus, les concepteurs de l'enquête ont sous-estimé l'importance du biais culturel lié aux différences d'attitude des répondants de différents pays.

Enfin, les contrastes que nous avons relevés sont révélateurs de certains mécanismes sous-jacents qui déterminent la réussite ou l'échec à l'égard d'un item et qui peuvent, encore une fois, être liés à des méthodes d'enquête concrètes. Les interviews menées auprès d'un certain nombre de répondants montrent que ces derniers préfèrent souvent répondre rapidement, quitte à ne pas comprendre le document tout à fait. Dans ces conditions, il est raisonnable de se demander ce qu'on mesure exactement : une capacité de perception visuelle ou une réelle capacité de comprendre et d'interpréter les renseignements concernant la tâche à accomplir? Autrement dit, on peut vraiment se demander si cet instrument d'évaluation convient à la mesure des compétences nécessaires pour « se débrouiller » dans la vie courante.

Nous avons montré que les données produites par l'EIAA étaient loin de satisfaire les ambitions initiales. Le biais observé se manifeste à divers degrés parmi les pays participants, et il est difficile d'en mesurer les effets respectifs. En ce qui concerne la France, par exemple, nous avons montré que la traduction du questionnaire avait pour effet d'accroître la difficulté de plus du tiers des questions posées.

Cette analyse montre la nécessité d'apporter le plus grand soin à chaque étape de l'élaboration d'un instrument d'enquête. Cette nécessité s'impose encore plus lorsque (comme c'est le cas ici), une fois les renseignements recueillis et les estimations établies, l'enquête quitte le champ de la recherche pour entrer dans le domaine public, acquérant ainsi un caractère irréfutable, malgré les réserves exprimées par les spécialistes.

BIBLIOGRAPHIE

- Blum, A., Guérin-Pace, F. et Goldstein, H. (2001), "An analysis of international comparisons of adult literacy", *Educational Assessment*, juillet.
- Blum, A. et Guérin-Pace, F. (2000), *Des lettres et des chiffres – Des tests d'intelligence à l'évaluation du "savoir lire", un siècle de polémiques*, Paris, Fayard, 191 p.
- Carey, S. (éd.) (2000). *Measuring Adult Literacy – The International Adult Literacy Survey in the European Context*. London, Office for National Statistics.
- Darcovich, N., M. Binkley, J. Cohen, M. Myrberg et S. Persson (1998), « Non-response Bias » dans National Center for Education Statistics, op. cit.
- Développement des ressources humaines Canada, OCDE (1997), *Literacy Skills for the Knowledge society, Further Results from the International Adult Literacy Survey*, Paris, 195 p.
- Goldstein H. et R. Wood (1989): « Five decades of item response modelling », *British Journal of Mathematical and Statistical Psychology*, vol. 42, pp. 139-167.
- Guérin-Pace, F. et Blum, A. (1999), L'illusion comparative. Les logiques d'élaboration et d'utilisation d'une enquête internationale sur l'illettrisme. *Population* 54:271-302.
- Guérin-Pace, F. et Blum, A. (2000), The comparative illusion: the International Adult Literacy Survey; *Population/ An English Selection*, 12:215-246.

Kalton G., Lyberg L. et J.-M. Rempp (1998), Review of methodology in Adult Literacy in OECD Countries in NCES, op. Cit. Kirsch I, A. Jungeblut et B. Mosenthal (1998), « The Measurement of Adult Literacy », dans National Center for Education Statistics, op. cit., ch. 7.

Kirsch, I., Murray, S. (1998). Introduction. dans: Murray, T. S., Kirsch, I. S. et Jenkins, L. B. (1998). *Adult literacy in OECD countries*. Washington, DC, National Center for Education Statistics

Murray, T. S., Kirsch, I. S. et Jenkins, L. B. (1998). *Adult literacy in OECD countries*. Washington, DC, National Center for Education Statistics.

National Center for Education Statistics (1998), *Adult Literacy in OECD Countries, Technical Report on the First International Adult Literacy Survey*, U.S. Department of Education, Office of Educational Research and Improvement, Washington, p. 215 et annexes.

OECD (1997). *Literacy skills for the knowledge society*. Paris, OECD.

Statistique Canada, OCDE (1995). *Littératie, économie et société, Résultats de la première Enquête internationale sur l'alphabétisation des adultes*, Paris, 218 p.

OECD (2000), *Literacy at the Information Age – Final report on the International Adult Literacy Survey*, Paris.