

Application de la méthodologie IALS aux données de l'enquête IVQ

Fabrice MURAT

INSEE, Division Emploi

En 1994, Statistique Canada et ETS (Educational Testing Service) ont organisé l'enquête internationale (IALS) sur les compétences en littératie¹ des adultes, dont les résultats ont ensuite été largement diffusés par l'OCDE (OCDE 2000). Les résultats français ont suscité de fortes interrogations : le classement international de la France était très mauvais et les journaux ont repris des fuites alléguant que 40 % des français seraient illettrés². Un certain nombre de problèmes méthodologiques sont apparus lors des expertises, justifiant le retrait de la France de cette enquête.

Plusieurs travaux récents ont contribué à renforcer les doutes sur les résultats de 1994 (Dickes et Vrignaud 1995, Blum et Guérin-Pace 1999, Blum et Guérin-Pace 2000, ONS 2000, Vrignaud 2001). L'opération PISA menée par l'OCDE en 2000 donne ainsi une proportion de 13 % au niveau de compétence le plus bas de l'échelle IALS parmi les élèves de 15 ans. La reprise des items et de la méthodologie IALS pour une partie de l'enquête IVQ amène un résultat très proche : 15 % à ce niveau dans l'ensemble de la population, avec l'hypothèse de travail qui semble la plus solide, soit très loin du chiffre de 41 % obtenu en 1994. De plus, au-delà de la question du niveau moyen, il apparaît de fortes incertitudes sur l'étude des déterminants des compétences en lecture : alors que IALS fait apparaître en France et dans les autres pays, une forte corrélation entre les scores et l'âge, les données IVQ montrent un lien moins fort, entièrement « expliqué » par les écarts de qualification scolaire entre générations.

Expertise de l'enquête IALS

Les problèmes apparus lors de l'expertise de IALS ont servi d'enseignement lors de la mise en place d'IVQ pour obtenir des données de bonne qualité. Il convient donc de revoir rapidement les pistes qui ont été explorées pour expliquer la particularité des résultats français.

¹ On désigne par ce néologisme d'origine anglo-saxonne l'aptitude à comprendre et à utiliser l'information écrite dans la vie courante, au travail et dans la collectivité en vue d'atteindre des buts personnels et d'étendre ses connaissances et ses capacités.

² Il importe de signaler que la définition de l'OCDE ne correspond que partiellement à la notion d'illettrisme. Les personnes se trouvant au niveau 1 de l'échelle de littératie sont celles qui sont uniquement capables de répondre à des questions simples. Il existe donc une grande hétérogénéité dans cette population entre les personnes qui réussissent les questions de niveau 1 les plus difficiles (mais pas celles qui ont été classées en niveau 2) et les personnes qui parviennent à peine à réussir les questions les plus simples. Cependant, les formulations de l'OCDE insistant sur les grandes difficultés de ces personnes, leur très faible niveau de compétence ont pu justifier l'usage du terme « illettrés » dans ce qu'il a de plus stigmatisant.

L'échantillonnage

L'échantillonnage était fondé sur la liste des numéros téléphoniques et la méthode des itinéraires pour régler le problème des numéros en liste rouge. L'enquêteur pouvait remplacer le logement en cas d'impossibilité à joindre le ménage de façon persistante. Il n'est pas certain que la procédure ait été toujours parfaitement appliquée. Le taux de réponse de l'enquête est d'ailleurs problématique : 45 % des ménages ont refusé de répondre. De plus, des biais importants dans l'échantillon, en terme de diplôme notamment sont assez mal corrigés par le calage des pondération effectué. Le traitement des personnes ne parlant pas bien français est aussi assez flou et c'est encore plus le cas dans certains autres pays comme l'Allemagne, où toutes les personnes interrogées étaient germanophones. Dans le cas d'IVQ, l'usage de l'échantillon maître de l'Insee garantit une meilleure représentativité de l'échantillon obtenu.

La passation de l'enquête

L'enquête IALS était une enquête assez longue : le livret proposé à l'enquêté comportait une quinzaine d'exercices. La passation de l'épreuve pouvait prendre jusqu'à deux heures. Il était difficile dans ces conditions de maintenir l'attention et la motivation de l'enquêté tout au long du questionnaire. De plus, la situation de l'enquêteur était assez inconfortable car il se trouvait complètement désœuvré, ce qui accroissait chez la personne interrogée le stress et l'impression qu'elle disposait d'un temps limité pour faire les épreuves, même si les consignes précisaient le contraire. L'usage de CAPI dans IVQ permet des interactions plus naturelles entre enquêteurs et enquêtés. De plus, les exercices étant posés un à un, il est possible d'arrêter l'enquête avant que la fatigue ne rende les réponses de la personne peu significatives. Mais surtout, on a pu collecter une information complète sur chaque exercice, évitant le survol observé dans les données de IALS.

La présentation de l'enquête est apparue aussi très importante. La référence au ministère de l'éducation nationale sur les livrets semble avoir réactivé chez de nombreuses personnes des souvenirs peu agréables et a contribué à donner à l'enquête un tour scolaire peu propre à maintenir la motivation des personnes interrogées. Cela nous a conduits pour notre enquête à définir un protocole d'approche le plus neutre possible, évitant autant que faire se peut, de décrire le contenu exact de l'enquête. Des questions spécifiques de mise en situation ont aussi été introduites dans notre module d'orientation.

La codification des réponses

Une grille de correction avait été élaborée pour les épreuves mais de nombreuses critiques lui ont été adressées. En effet, elle s'avère très grossière : dans le fichier, on ne dispose plus généralement que de trois codes possibles : bonne réponse, mauvaise réponse, non-réponse. Or l'analyse des données et le retour aux questionnaires ont montré qu'il existait un flou assez important dans la codification.

En effet, d'une part, les exercices étaient souvent très ambigus et il est apparu dans de nombreux cas possible de donner une bonne réponse autre que celles qui avaient été prévues dans les consignes. L'alternance de questions simples et de questions difficiles a pu aussi provoquer la sensation de questions-pièges sur les exercices plus évidents, provoquant des réponses plus subtiles que ce que l'on attendait. C'est pourquoi il est apparu important dans le cas d'IVQ de recueillir les réponses de la façon la plus détaillée possible, le recours à CAPI permettant de les retranscrire aussitôt. Une opération de codage a ensuite été entreprise, le retour aux réponses originelles étant toujours possible.

On a aussi constaté une certaine confusion entre mauvaise réponse et non-réponse. Il semble en effet que très souvent les personnes ont sauté des exercices quand il portait sur des thèmes qui ne les intéressaient pas. De façon encore plus fréquente, les personnes interrogées interrompaient

l'enquête avant la fin, du fait de sa durée et de l'investissement demandé. Le codage et le traitement effectués sont alors assez flous : dans certains cas, la non-réponse sera considérée comme un échec à la question, dans d'autres comme une réelle absence d'information, sans que la distinction s'avère toujours pertinente et constante d'un enquêté à l'autre.

La comparabilité internationale

Alain Blum et France Guérin-Pace, ainsi que Pierre Vrignaud³ ont beaucoup travaillé sur la comparabilité internationale de l'enquête. Ils ont mis en évidence de nombreux problèmes de traduction, qui souvent ont pu rendre plus difficiles les questions de la version française. Mais au-delà de la qualité de la traduction, ils mettent en doute la possibilité de construire une mesure unidimensionnelle permettant des comparaisons entre pays. De nombreux facteurs peuvent provoquer ce que l'on qualifiera de biais culturels et rendre telle question plus difficile dans tel pays que dans tel autre, alors que l'inverse s'observera sur un autre exercice. Des techniques statistiques existent pour repérer ces décalages mais le traitement à apporter n'est pas évident : suppression des items problématiques (mais quel seuil de divergence accepte-t-on) ou reconnaissance du caractère multi-dimensionnel du domaine que l'on veut étudier. Ces questions, très importantes dans le cadre d'une enquête internationale, ne seront pas développées ici car IVQ n'a pas pour l'instant vocation à être passée hors de France.

La mise en doute des résultats de IALS par l'OCDE

L'OCDE a mené une étude qui indirectement peut servir d'expertise des résultats de IALS (OCDE 2002). En 2000, des échantillons représentatifs dans 32 pays de la population des élèves de 15 ans ont passé des épreuves d'évaluation en lecture (enquête PISA). Ces épreuves comportaient un certain nombre d'items repris de l'enquête IALS (les items proposés étaient cependant parfois sensiblement différents de la version proposée en 1994 en France). A l'aide de ces items, les psychométriciens ont pu effectuer un ancrage des données de PISA sur l'échelle de IALS⁴.

Tableau 1 : Répartition des jeunes de 15 ans et de l'ensemble de la population selon le niveau de compétence IALS, en France

Ligne		niveau 1	niveau 2	niveau 3	niveaux 4 et 5
1	les jeunes de 15 ans dans PISA	13	40	38	9
2	l'ensemble de la population IALS	41	34	22	3
3	les <26 ans IALS	27	35	32	6
4	les 26-45 ans IALS	35	37	25	3
5	les plus de 45 ans IALS	59	28	11	2

Sources : OCDE 2002 (ligne 1) National Center for Educational Statistics (ligne 2) Insee (lignes 3 à 5).

Les moyennes par pays obtenues pour les élèves de 15 ans en 2000 ne sont pas très nettement reliées à celles obtenues par les jeunes de moins de 26 ans à l'enquête IALS. En particulier, la Suède premier pays pour l'enquête IALS est plutôt au milieu de classement pour PISA, tandis que l'Allemagne, dans la moyenne pour IALS, a obtenu de mauvais résultats lors de l'évaluation des élèves de 15 ans. Quant à la France, elle apparaît dans la moyenne dans l'enquête PISA et le tableau ci-dessous montre combien les résultats des deux enquêtes sont différents. Alors que 41 % de la population française se trouvaient au plus bas niveau de littératie pour l'enquête IALS (27 % pour les moins de 26 ans : nous reviendrons sur cet écart plus loin), ils ne sont que 13 % lors de l'enquête PISA parmi les élèves de 15 ans. A moins de faire l'hypothèse d'une amélioration formidable de notre système éducatif, en particulier de 1994 à 2000, cet écart amène à s'interroger sur au moins l'un des deux résultats.

³ Qui participent tous trois au comité de pilotage de l'enquête

⁴ Plus précisément, il s'agit de l'échelle Prose de IALS, portant sur des textes suivis (et non des tableaux/graphiques ou des données quantitatives qui faisaient l'objet des deux autres échelles).

Le protocole de comparaison IALS-IVQ

Le présent travail ne portera que sur certains des problèmes évoqués. En particulier, la comparabilité internationale des données ne pourra faire l'objet d'analyse puisque IVQ n'a été menée qu'en France. L'expertise des épreuves et les problèmes d'équivalence des traductions ont par ailleurs déjà été longuement étudiés, notre objectif est donc plutôt de mesurer l'effet des conditions de collecte sur les résultats. Pour ce faire, une partie des personnes interrogées passaient des exercices repris à l'identique de l'enquête IALS. Il est ainsi possible de mesurer l'impact des changements apportés au recueil des données, en supposant fixée la dimension mesurée. Plus précisément, 6 exercices comportant 16 questions ont été repris. Le choix s'est porté sur des exercices qui avaient déjà été retenus pour l'enquête PISA. En revanche, on a tenu à conserver les formulations de 1994, pour améliorer la comparabilité, même quand les analyses ont fortement mis en doute leur qualité. De plus, sur les 16 items, seuls 13 faisaient partie de l'échelle Prose de IALS (les trois autres appartenaient à l'une des deux autres dimensions de la littérature : Documents ou Textes quantitatifs). Il ne serait sans doute pas impossible de construire une échelle commune aux 16 items, les corrélations étant fortes, mais pour appliquer la méthodologie de 1994, nous nous sommes vu contraint d'éliminer provisoirement ces items.

Il ne faut cependant pas croire que la reprise à l'identique des épreuves suffit à garantir la comparabilité des résultats. Les problèmes de traduction mis en évidence par l'expertise n'ont pas disparu, laissant subsister un doute sur la comparabilité internationale des données. D'autre part, la modification de la passation peut entraîner un changement dans la nature de la tâche exigée⁵. Par exemple, l'enquête IALS utilisait l'écrit comme moyen de recueil de l'information, alors que dans le cadre d'IVQ, c'est par oral que la personne donne ses réponses. Cette différence n'est pas neutre et selon les questions n'a pas exactement les mêmes conséquences (il y aura sans doute peu d'impact si la réponse à donner est courte ; en revanche, pour une réponse longue, l'usage de l'oral permet de cibler l'évaluation sur la compréhension et élimine une partie de l'effet parasite des problèmes éventuels d'expression, plus sensibles à l'écrit). Il importera donc de s'assurer de la comparabilité des épreuves, ce qui permettra de mieux mesurer l'effet des changements dans le protocole de collecte.

En quoi le protocole de collecte d'IVQ diffère-t-il de celui de IALS ? Plusieurs points ont fait, on va le voir, l'objet d'amélioration, en particulier pour tenir compte de l'impact de la motivation des sujets sur les résultats.

- *La procédure d'orientation* : dans l'enquête IALS, une partie des sujets (en France moins de 5 %) ont arrêté l'épreuve d'évaluation après le livret préliminaire : ils ont réussi au plus une question sur les six proposées, ce qui mettait fin à l'épreuve. Ils sont bien sûr classés au niveau le plus bas de l'échelle de compétence. Dans IVQ, il existe aussi un processus d'orientation, qui oriente la personne soit vers un module adapté aux personnes en difficulté (module ANLCI) soit vers les exercices IALS proprement dit (ou bien l'épreuve originale que nous avons élaborée mais elle ne sera pas étudiée ici). Il faudra faire certaines hypothèses pour introduire les personnes orientées vers le module ANLCI dans l'échelle IALS.
- *Le raccourcissement des épreuves* : au lieu de 15 exercices, seuls 6 ont été proposés aux personnes. Du fait que l'on se restreint à une seule des trois échelles mesurées dans IALS, ce choix a peu d'impact sur la fiabilité des résultats. En revanche, il permet d'alléger la passation et de ne pas trop lasser la personne interrogée.
- *La segmentation de l'épreuve* : chaque exercice est proposé successivement et non avec tous les autres sous forme de cahier. L'attention de la personne interrogée est donc concentrée à chaque fois sur les questions posées et l'on évite le risque que l'enquêté passe d'un exercice à l'autre, pour choisir ceux qui lui plaisent. Au contraire de IALS,

⁵ La forme des exercices est d'ailleurs un élément essentiel à prendre en compte dans l'interprétation des résultats, en particulier dans le cadre des enquêtes internationales : par exemple, contrairement à une idée reçue, les résultats des élèves français sont généralement meilleurs sur les QCM que sur les questions ouvertes, par rapport aux autres pays (DPD 2002).

dans IVQ, les non-réponses partielles sont négligeables. La possibilité d'interrompre l'épreuve a aussi prévue⁶.

- *Le codage des réponses* : les réponses données sur le cahier de IALS ont ensuite été codées de façon rudimentaire sous la forme : « Non-réponse » / « Bonne réponse » / « Mauvaise réponse ». Vu les ambiguïtés de certaines questions, un plus grand détail, en particulier pour les mauvaises réponses aurait été préférable. C'est pourquoi dans l'enquête IVQ, on a demandé aux enquêteurs de coder en clair la réponse de l'enquêté (certaines réponses ayant cependant été précodées). Il est ainsi possible de procéder à plusieurs variantes de corrections⁷. Dans la présente note, on retiendra celle qui correspond aux consignes IALS de 1994. Ces consignes ont en général été jugées trop strictes, vu les ambiguïtés relevées pour certaines questions. Une variante plus large sera élaborée par la suite.
- *Les interactions enquêteur/enquêté* : l'usage de CAPI et la segmentation de l'épreuve permettent de donner à son déroulement un caractère un peu moins artificiel. Il est ainsi possible à l'enquêteur de maintenir la concentration et la motivation de la personne interrogée, de rappeler les consignes quand elles ne sont pas bien comprises. Lors des formations, il a cependant été nettement précisé que ces interactions devaient se faire en respectant la neutralité la plus stricte, afin de ne pas influencer directement ou indirectement l'enquêté dans ses réponses.
- *Des indicateurs de motivation* : deux types d'indicateurs sur la passation de l'épreuve ont été recueillis. D'une part, on dispose du temps passé par chaque personne sur chaque exercice, ce qui donnera une indication de l'investissement de la personne. D'autre part, une grille d'observations a aussi été remplie pour chaque exercice par l'enquêteur indiquant comment la personne interrogée avait réagi. Ces indicateurs n'ont pas été utilisés dans la présente note, du fait de la difficulté à combiner leur usage avec la méthodologie de IALS (voir par contre la contribution de P. Zamora).

Les données et les traitements préparatoires

Sur les 2086 personnes ayant répondu à l'enquête IVQ⁸, 682 étaient destinées à passer les exercices de IALS⁹. Sur ces 682, 25 personnes n'ont pas passé les exercices. Pour 6 d'entre elles, des difficultés particulières ont été signalées (ils disent ne pas savoir lire ou ne parlent pas le français) : on va imputer des réponses fausses à l'ensemble des exercices IALS (ce qui correspond à la procédure suivie en 1994) ce qui conduira naturellement à les classer en niveau 1. Pour les 19 autres, la raison évoquée pour éviter les exercices ne semble pas liée à la lecture (problèmes de temps, méfiance vis-à-vis des enquêtes, etc.) : on va provisoirement laisser de côté ces personnes et on leur imputera ensuite un score en fonction de leurs caractéristiques socio-démographiques¹⁰ (là encore, de façon comparable aux procédures de 1994).

Sur les 663 personnes restantes, 63 ont été orientées vers le module ANLCI (et n'ont donc passé aucun item de IALS ; ce groupe inclut aussi les 6 personnes précédemment évoquées ayant déclarées ne pas savoir lire). Parmi les 600 ayant passé les exercices IALS, 53 ont interrompu l'épreuve en cours et, du fait d'un problème informatique (voir note 5), on ne dispose plus que de leurs réponses à l'exercice d'orientation. Finalement, 547 personnes ont passé l'ensemble des épreuves IALS.

⁶ L'interruption d'épreuve provoque donc des valeurs manquantes finales. On résout généralement ce problème en se restreignant aux questions passées et en tenant compte de leur difficulté. Malheureusement, un problème informatique imprévu a provoqué l'effacement de toutes les réponses aux exercices IALS, dès que l'interruption se produisait : environ 8 % des personnes se trouvent dans ce cas. On ne dispose plus pour elles que de leurs réponses à l'exercice d'orientation et de leurs caractéristiques socio-démographiques.

⁷ La codification des réponses aux questions de IALS a été opérée par trois équipes indépendantes : une équipe Insee-Crest (F. Bulot, L.-A. Vallet et D. Verger), une équipe Dares (P. Zamora) et une équipe DEP (T. Rocher), avant la mise en commun et discussion des codages proposés.

⁸ Le fichier comporte 4011 ménages. Dans 951 cas, l'enquête n'a pu être réalisée pour des raisons « objectivement neutres » (logement vacant, aucune personne dans le champ de l'enquête, inaptitude, absence de longue durée). Sur les 3060 ménages restants, 974 ont refusé l'enquête ou sont impossibles à joindre (ce qui est souvent la même chose) : le taux de « refus » est donc de 32 %. Un calage sur marges permet de corriger les biais induits par ces refus.

⁹ La règle était la suivante : si le numéro identifiant du logement (aléatoire) + 1 est divisible par 3, la personne interrogée passera les exercices IALS si elle réussit l'exercice d'orientation (sinon, elle passe le module ANLCI).

¹⁰ Le niveau d'études et le diplôme ont été utilisés, prédisant 29 % de la variance du score IALS pour les répondants.

Les résultats que nous allons présenter vont être sensibles aux hypothèses faites pour les personnes n'ayant pas passé les items IALS. Les classer directement au niveau 1 (ou leur imputer des réponses fausses partout) est un choix qui peut se justifier pour les personnes orientées vers le module ANCLI. Ce choix peut paraître sévère pour ceux qui ont arrêté l'épreuve en cours : ils sont un peu moins performants à l'orientation que ceux qui vont au bout des exercices (ils ont une moyenne de 16,7 sur 19 à l'exercice de compréhension contre 17,7) mais se situent nettement au-dessus des personnes orientées vers le module ANLCI (en moyenne, 8,7 sur 19).

Une variante a donc été produite en leur imputant des réponses aux exercices de IALS. On a modélisé (par des régressions logistiques) sur les 547 répondants à IALS, la réussite à chacun des items en fonction du score de compréhension de l'orientation. Ce modèle a ensuite servi à imputer des réponses au module IALS, pour ceux qui ne l'ont pas passé, c'est-à-dire à la fois pour ceux qui l'ont abandonné en cours et pour ceux qui ont passé le module ANLCI (pour ces derniers bien sûr l'imputation est très fragile et l'on préférera aussi l'hypothèse qui les place directement en niveau 1 même si elle est un peu sévère car ces personnes auraient pu réussir certains items). Il importe de souligner que pour les personnes ayant arrêté l'épreuve IALS en cours, les imputations relèvent deux logiques : pour les questions que les personnes ont effectivement passées mais pour lesquels leurs réponses ont « disparu » du fait d'un problème informatique, l'imputation est une simple réparation ; pour les questions situées après l'interruption, l'imputation est un peu moins justifiée et on pourrait retenir une autre hypothèse comme par exemple, l'échec à toutes ces questions, si l'enquête invoque la trop grande difficulté des exercices pour abandonner.

Comparaison des réussites aux items IALS-IVQ

Dans un premier temps, avant de procéder à l'estimation des compétences individuelles, un travail de vérification et d'ajustement est à effectuer sur les items, afin de s'assurer que les dimensions mesurées en 1994 et 2002 sont bien comparables. On a d'abord cherché à comparer l'ordre de réussite des items en 2002 avec ce qui a été observé en France et à l'étranger en 1994. le tableau suivant donne l'ensemble des données disponibles (NCES 1998).

Tableau 2 : réussites aux items repris de IALS dans IVQ et dans IALS

	IVQ (1)	IVQ (2)	Paramètre IRT	IALS 94 France	IALS 94 Autres pays
m1	56,1%	49,6%	275	55%	77%
m2	48,6%	44,3%	310	38%	62%
m4	74,7%	67,6%	281	57%	77%
m5	59,3%	53,6%	318	54%	65%
m6	82,7%	78,5%	314	43%	69%
m7	69,0%	66,1%	306	43%	60%
m9	68,9%	63,3%	306	54%	60%
m10	90,8%	87,0%	338	40%	53%
m11	86,0%	78,6%	287	40%	73%
m12	63,0%	58,8%	329	20%	55%
m13	42,9%	38,1%	374	22%	27%
m14	74,2%	67,7%	297	55%	70%
m16	80,5%	75,1%	246	71%	87%

Note : 13 items ont été retenus comme faisant partie de l'échelle Prose. La deuxième colonne donne le taux de réussite pour les 547 personnes ayant passé les items IALS dans IVQ. La troisième élargit à l'ensemble des 663 personnes ayant passé les épreuves avec imputation à partir des résultats à l'exercice d'orientation. La quatrième colonne présente les paramètres IRT attribués par les psychométriciens en 1994. La cinquième colonne donne le taux de réussite en France en 1994, tandis que la dernière donne le taux de réussite moyen en 1994 dans 7 pays/régions linguistiques (Allemagne, Canada anglais, Canada français, Etats-Unis, Suède, Suisse allemande, Suisse Française).

Si l'on prend l'ensemble des items, les hiérarchies ne se correspondent pas très bien : La corrélation linéaire entre la série de taux de réussite IVQ et la série des taux de réussite dans les autres pays en 1994 n'est que de 0,48 ; la corrélation est de 0,35 avec celle concernant la France. L'item m10 paraît particulièrement problématique : il est réussi à 90,8 % en 2000, alors que le taux de réussite est de 40 % en France en 1994 et 53 % dans les autres pays, des écarts bien supérieurs à ce que l'on observe sur les autres items¹¹. Deux autres items nous semblent poser problème : ceux de l'exercice 1. En effet, alors que sur les autres items, on observe une progression sensible entre les résultats français à IALS et ceux à IVQ, rien de tel n'apparaît sur ces items-là¹². Si l'on exclut ces trois items de la comparaison, la cohérence est beaucoup plus forte : la corrélation est de 0,84 entre la série IVQ et celles des taux de réussite étrangers ; elle reste peu satisfaisante avec les taux français (0,5).

Cette confrontation indique donc une cohérence imparfaite entre les données IVQ et les données IALS dans la hiérarchie des items. Il semble cependant qu'en excluant les trois items qui posent le plus de problème, on puisse faire l'hypothèse que les échelles sont équivalentes. En attendant une expertise plus approfondie, ce sont donc les dix items restants qui ont été utilisés pour construire un score comparable à celui de IALS. On a en effet appliqué un modèle MRI à deux paramètres¹³ sur les données, en reprenant pour les 10 items les paramètres estimés en 1994, ce qui permet d'estimer pour chaque personne sachant sa réussite ou son échec à ses items, un score exactement comparable avec ceux qui ont été calculés en 1994 pour l'enquête IALS.

Quelques résultats

Il est alors possible de donner la distribution selon l'enquête IVQ de la population française dans les niveaux de compétences IALS. Les résultats sont présentés dans le tableau 3 en fonction de l'hypothèse retenue pour les personnes n'ayant pas passé les items.

Comme on l'a indiqué, l'hypothèse qui semble pour l'instant la plus vraisemblable est la deuxième consistant à imputer des réponses pour les personnes ayant abandonné l'épreuve IALS en cours (et dont nous avons perdu les réponses pour les questions qu'ils ont quand même passé) et à classer au niveau 1 les personnes en difficulté au module d'orientation et orientées vers le module ANLCI (de fait, avec la première hypothèse, 55 % sont déjà classées au niveau 1, la plupart des autres se trouvant au niveau 2).

Tableau 3 : répartition de la population française dans les niveaux de compétences IALS selon IVQ

	niveau 1	niveau 2	niveau 3	niveau 4 ou 5
Imputation de réponses pour tous ceux qui n'ont pas passé l'épreuve IALS	10,0	29,6	44,7	15,6
Imputation pour les personnes abandonnant l'épreuve IALS + mise au niveau 1 pour ceux orientés vers le module ANLCI	15,4	25,6	43,6	15,4

¹¹ Deux raisons peuvent expliquer cette différence : d'une part on peut évoquer une certaine confusion dans les consignes de correction de 1994 (en tout cas dans leur version française) qui ont pu exclure des bonnes réponses ; d'autre part, cet item est un bon exemple des difficultés de passage de l'écrit à l'oral. En effet, la question est « En utilisant les renseignements de la brochure, formulez à votre manière une différence entre l'entretien par jury et l'entretien en groupe ». Cet item demande donc explicitement un effort d'expression, qui n'a pas la même difficulté selon qu'elle se fait par écrit ou à l'oral.

¹² Là encore, l'examen détaillé des items confirme le doute. Ces deux items portent sur des critiques de films et l'on observe le même phénomène qu'en 1994 : les réponses données sont perturbées par les connaissances des personnes interrogées. Pour le premier, il faut citer des comédies et beaucoup en 2000 encore indiquent *Monsieur Hire*, repérant le nom de Michel Blanc, acteur connu pour ses rôles comiques. Ce distracteur joue évidemment beaucoup moins à l'étranger.

¹³ Le modèle MRI synthétise une matrice d'indicatrices de réussite de 10 colonnes et 547 lignes, en posant que la

probabilité qu'un individu j réussisse l'item i est donnée par $\Pr(x_{ij} = 1 | \theta_j, a_i, b_i) = \frac{1}{1 + e^{-D \times a_i (\theta_j - b_i)}}$, où a_i est le

coefficient de discrimination de la question i , D un facteur d'échelonnement permettant le passage à la fonction de lien ogive normale (constante égale à 1,7), θ_j la compétence du sujet et b_i la difficulté de l'item. Normalement, on estime l'ensemble des paramètres (maximum de vraisemblance). Ici, les paramètres a_i et b_i seront fixés à leurs valeurs de 1994 et seuls les θ_j seront estimés. Voir (D'Haultfoeuille et alii 2002) pour plus de précisions sur les modèles MRI.

Le résultat obtenu frappe par l'écart avec les résultats IALS présentés dans le tableau 1 : même avec l'hypothèse la plus sévère, on est loin de 41 % de niveau 1. La distribution se rapproche plus de ce que l'on observe dans l'enquête PISA, mais elle apparaît cependant plus dispersée dans IVQ : les proportions aux niveaux extrêmes sont plus élevées. La contribution de S. Lollivier et D. Verger donne des résultats similaires avec une méthodologie un peu différente.

D'autre part, l'un des résultats importants de l'enquête IALS, qui apparaît sur le tableau 1, est la forte corrélation entre le niveau de compétences et l'âge de la personne : 59 % des personnes âgées de plus de 45 ans sont classées au niveau 1 contre 27 % des moins de 26 ans¹⁴. La corrélation linéaire entre l'âge et le score à IALS est de -0,32.

Cette corrélation est beaucoup moins nette pour l'enquête IVQ. Certes, la proportion de personnes au niveau 1 est plus faible pour les moins de 26 ans que pour les plus de 45 ans (8,3 % contre 20,3 %) mais le lien est tout de même moins fort : la corrélation n'est que de -0,24 entre le score et l'âge. Cette corrélation est d'ailleurs aussi sensible à l'épreuve utilisée car pour les personnes qui n'ont pas passé IALS mais l'épreuve originale construite pour IVQ, elle n'est que de -0,13. De plus, avec les deux épreuves, cette corrélation n'est plus significative quand on introduit le diplôme et le niveau d'études dans l'analyse¹⁵. Sur les données IALS, au contraire, elle résiste à l'introduction de ces variables même si le coefficient multiplicatif associé à l'âge est tout de même divisé par trois (il passe de -1,42 par année supplémentaire à -0,47).

Tableau 4 : répartition de la population française dans les niveaux de compétences IALS dans IVQ selon l'âge

	niveau 1	niveau 2	niveau 3	niveau 4 ou 5
Moins de 26 ans	8,3	11,5	56,9	23,3
Entre 26 et 45 ans	14,2	23,9	45,2	16,7
Plus de 45 ans	20,3	34,7	35,1	10,0

Note : dans ce tableau, on a retenu la deuxième hypothèse présentée dans le tableau 3.

Conclusion

L'analyse présentée ici ne permet pas d'avancer sur la comparabilité des données de l'enquête IALS. Elle montre seulement que le même protocole d'exercices proposé dans des conditions de collecte différentes (et, postulons-le, de meilleures qualités) donne des résultats très sensiblement différents, la proportions de personnes en difficulté passant de 41 à 15 %. Les efforts pour rendre l'enquête moins lourde, la situation moins artificielle et préciser la mesure en réduisant le recours à l'écrit peuvent expliquer la variation des résultats. Dans quelle mesure les autres pays participants auraient subi les mêmes difficultés ? Pourquoi ces variations induisent-elles aussi des changements dans les corrélations observées (voir ici l'exemple de l'âge) ? Les expertises ultérieures permettront d'avancer sur ces questions.

¹⁴ On notera au passage l'ironique paradoxe de l'enquête : alors que les résultats pourraient suggérer une amélioration nette du niveau de compétences de la population et indirectement de son système éducatif (bien sûr, si l'on néglige imprudemment un effet de cycle de vie dû à l'éloignement du système éducatif), les journaux ont surtout utilisé les données IALS pour dénoncer la faillite de l'école, « devenue une usine à chômeurs et à illettrés » (c'est moi qui souligne).

¹⁵ Il y aurait donc une augmentation du niveau moyen de compétences de la population entièrement expliquée par l'augmentation du niveau de formation, chaque niveau étant toujours associé au même niveau de compétence. Ce résultat rappelle ce qui a été observé sur les tests des 3 jours (Baudelot et Establet 1988) mais doit être pris avec beaucoup de précaution car il ne tient pas compte d'éventuels effets de cycle de vie et pourrait être sensible aux épreuves utilisés.

Bibliographie

- [1] P. Dickes, & P. Vrignaud (1995) « *Rapport sur les traitements des données françaises de l'enquête internationale sur la littératie* », Rapport pour le Ministère de l'Education Nationale. Direction de l'Evaluation et de la Prospective.
- [2] A. Blum & F. Guérin-Pace (1999) « *L'illusion comparative : les logiques d'élaboration et d'utilisation d'une enquête internationale sur l'illettrisme* », *Population*, 54 (2), pp. 271-302
- [3] A. Blum & F. Guérin-Pace (2000), *Des lettres et des chiffres*, Fayard.
- [4] C. Baudelot & R Establet (1988), « *Le niveau intellectuel des conscrits ne cesse ne s'élever* », Education et Formations, Ministère de l'Education nationale, mars 1988
- [5] X. d'Haultfoeuille, F. Murat et T. Rocher (2002) : « *La mesure des compétences, les logiques contradictoires des évaluations internationales* », Actes des Journées de Méthodologie Statistique (Décembre 2000), Insee.
- [6] ONS (2000), *Measuring Adult Literacy* (sous la direction de Siobhàn Carey), ONS
- [7] NCES (1998), *Adult Literacy in OECD countries : Technical Report on the first International Adult Literacy Survey*, NCES.
- [8] OCDE – Statistique Canada (2000), *La littératie à l'ère de l'information : rapport final de l'enquête internationale sur la littératie des adultes*, OCDE
- [9] OCDE (2002), *Reading for Change. Performance and engagement across countries. Results from PISA 2000*, OCDE.
- [10] Vrignaud, P. (2001). « *Evaluations sans frontières : comparaisons interculturelles dans le domaine de la cognition* ». in M. Huteau & J. Lautrey (Eds). *Les figures de l'intelligence*. Paris : Editions et Applications Psychologiques.

