

L'évaluation des compétences des adultes : des méthodes en plein développement

Fabrice Murat

Bureau des études statistiques sur l'enseignement scolaire
Direction de l'évaluation, de la prospective et de la performance

L'évaluation des compétences des adultes est un domaine complexe, dont l'émergence est assez récente. On n'interroge pas un adulte chez lui comme on évalue un élève dans sa classe. La première expérience en France d'une certaine ampleur a montré les difficultés que l'on pouvait rencontrer : selon cette enquête (IALS), 40 % des adultes de 18-65 ans en France étaient en grande difficulté face à l'écrit en 1994. Des expertises méthodologiques ont remis en cause la fiabilité de l'enquête, justifiant le retrait de la France de l'opération. Cet article reprend l'un de ces travaux, en montrant l'importance des conditions de collecte : en utilisant un protocole mieux adapté, la proportion de personnes en difficulté passe à 15 %. Ceci montre la nécessité de construire un protocole qui maintienne la motivation des enquêtés, en limitant la longueur de l'épreuve et en adaptant la difficulté des exercices au niveau de la personne, par exemple. Ces résultats prennent une importance particulière dans la perspective de la mise en œuvre de l'enquête PIAAC, prévue par l'OCDE en 2011.

Or la France faisait partie des premiers pays participant à IALS, en 1994, mais 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 illettrés. Un certain nombre de problèmes méthodologiques sont apparus lors des expertises, justifiant le retrait de la France de cette enquête et la non-participation à l'enquête ALLS. C'est pourquoi les instances concernées par le sujet² ont décidé en 2000 de privilégier une approche nationale, pour développer nos connaissances dans ce domaine très particulier. Une enquête spécifique

(Information et vie quotidienne [IVQ]) a été conçue et menée une première fois sur le terrain fin 2002. À visée principalement méthodologique, cette première collecte a montré la faisabilité de ce type d'enquête et a été suivie d'une opération de plus grande envergure à la fin 2004. L'enquête IVQ devrait être rééditée en 2010, ce qui permettra de mesurer l'évolution des compétences.

La proximité des opérations IVQ 2010 et PIAAC 2011 amène à s'interroger sur de possibles convergences. L'enquête PIAAC s'inscrit dans le prolongement de IALS, certains pays ayant participé à cette dernière enquête souhaitant que les résultats de 2011 leur permettent d'avoir une

NOTES

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. L'enquête IVQ dont on va parler est issue de la collaboration de plusieurs organismes : l'ANLCI (Agence nationale de lutte contre l'illettrisme), le CGP (Commissariat général au plan), le CREST (Centre de recherche en économie et en statistique), la DARES (Direction de l'Animation de la recherche et des études sociales du ministère du Travail), la DEPP (Direction de l'évaluation, de la prospective et de la performance du ministère de l'Éducation nationale), la DGEFP (la Délégation générale à l'emploi et la formation professionnelle), DGLFLF (la Délégation générale à la langue française et aux langues de France du ministère de la Culture) la DIV (Délégation interministérielle à la ville), l'INED (Institut national des études démographiques), l'INETOP (Institut national d'étude du travail et de l'orientation professionnelle), l'INSEE (Institut national de la statistique et des études économiques), l'ONPES (Observatoire national de la pauvreté et de l'exclusion sociale). Plusieurs équipes de recherche universitaires ont aussi participé à la construction des tests d'évaluation.

idée de l'évolution des compétences de leur population. Le dispositif IVQ s'est en partie construit sur l'analyse des défaillances de l'enquête IALS en termes de méthodologie d'évaluation et de collecte, repérées lors des nombreuses expertises. Quel est l'impact des choix méthodologiques sur les résultats ? Dans quelle mesure les enseignements d'IVQ peuvent-ils être transposés au niveau international et guider l'élaboration de PIAAC ? Comment articuler en France deux opérations si proches dans le temps et portant à peu près sur le même thème, mais avec des perspectives et des choix méthodologiques qui resteront sans doute un peu différents ? Toutes ces questions nous poussent à revoir les différences entre les deux dispositifs et à reprendre l'analyse d'un protocole particulier de l'enquête IVQ de 2002, qui permet de faire un pont avec IALS.

EXPERTISE DE IALS

Malgré le retrait de la France de l'enquête IALS, plusieurs journaux ont repris un résultat, qui en serait tiré : 40 % des personnes vivants en France seraient illettrés. En fait, il s'agissait là du résultat sur l'une des trois échelles de l'enquête, la compréhension de *Textes Suivis* ou de *Prose*, les résultats sur les échelles *Textes schématiques* et *Textes quantitatifs*, étant un peu moins mauvais. D'autre part, la définition de l'OCDE des personnes en difficulté ne correspond que partiellement à la notion d'illettrisme et se fonde sur la notion plus complexe d'échelle de compétence issue des modèles de réponse à l'item, que nous présenterons plus loin. Les questions sont classées par ordre de difficulté

sur une échelle de difficulté variant de 0 à 500³. Parmi ces questions, on isole les plus faciles, celles dont le paramètre de difficulté est inférieur à 225, que l'on qualifie de « niveau 1 » ; les questions de niveau 2 vont jusqu'à la valeur 275 ; le niveau 3 de 275 à 325, etc. On attribue comme score à une personne le niveau de difficulté d'une question qu'il a 80 % de chance de réussir. Ainsi, une personne avec le score de 250 a 80 % de chances de réussir une question de cette difficulté ; il a sensiblement plus de 80 % de chances de réussir une question de niveau 1 ; en revanche, il a nettement moins de 80 % de chances de réussir une question de niveau 3. On dira que cette personne est « de niveau 2 ». Ainsi, les personnes « de niveau 1 » sont celles qui ont 80 % de chances de réussir une question de ce niveau, c'est-à-dire parmi les plus simples que contenaient les questionnaires. Il existe donc une grande hétérogénéité dans cette population, entre les personnes qui réussissent généralement les questions de niveau 1 les plus difficiles (mais bien moins souvent 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 ». De plus, la France se trouvait nettement derrière la plupart des autres pays participant à la première vague d'enquête : la proportion de personnes en difficulté sur l'échelle de *Prose* était de 21 % aux États-Unis, 14 % en Allemagne, 13 % aux Pays-Bas, par exemple. Ces résultats assez surprenants ont conduit à un ensemble de travaux, qui ont justifié le retrait de la France de l'en-

quête (Dickes et Vrignaud 1995, Blum et Guérin-Pace 1999, Blum et Guérin-Pace 2000, ONS 2000, Vrignaud 2001). Les principaux enseignements de ces travaux vont être rappelés ici et on reviendra plus en détail ensuite sur ceux qui ont servi lors de la construction de l'enquête IVQ.

L'échantillonnage était fondé sur la liste des numéros téléphoniques et la méthode des itinéraires pour régler le problème des numéros en liste rouge : le numéro téléphonique désignait une adresse et en suivant des instructions précises (prendre la deuxième rue à droite en allant vers le nord, puis le deuxième immeuble à gauche, etc.), l'enquêteur déterminait le logement à interroger. L'enquêteur pouvait remplacer le logement en cas d'impossibilité à joindre le ménage de façon persistante. 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 termes de diplôme notamment, sont assez mal corrigés par le calage des pondérations 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

NOTE

3. Il importe de signaler que, dans les modèles de réponse à l'item, les valeurs sont en fait partiellement arbitraires et peuvent être changées par n'importe quelle transformation linéaire. En pratique, on définit l'échelle de façon univoque en fixant la moyenne et l'écart-type de la distribution (par exemple, dans le cas de PISA ou de TIMSS, qui utilisent ce type de méthode, la moyenne internationale est de 500 et l'écart-type de 100). Aussi la valeur 0 n'a aucune signification particulière et ne correspond bien sûr pas à l'item le plus facile qui existe ou à la personne la moins compétente possible...

toutes les personnes interrogées étaient germanophones. Dans le cas d'IVQ, l'usage de l'échantillon maître de l'Insee, tiré du recensement, garantit une meilleure représentativité de l'échantillon obtenu.

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 d'une collecte assistée par ordinateur (CAPI) dans IVQ permet des interactions plus naturelles entre enquêteurs et enquêtés. De plus, les exercices étant posés un à un, sur des feuillets séparés, 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 a conduit pour IVQ 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 le module d'orientation pour faire accepter plus facilement le processus d'évaluation aux personnes en difficulté.

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'étude des données et le retour aux questionnaires ont montré qu'il existait un flou assez important dans la correction et qu'un retour sur les réponses pour une analyse fine était parfois utile. En effet, d'une part, les exercices étaient souvent très ambigus et il est apparu possible dans de nombreux cas 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 ils portaient 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.

Alain Blum et France Guérin-Pace, ainsi que Pierre Vriгдаud 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 multidimensionnel 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.

IALS ET PISA

L'OCDE a mené une étude qui indirectement peut servir d'expertise des résultats de IALS (OCDE 2002).

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 moins de 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).

Lecture – 13 % des jeunes de 15 ans en France, d'après PISA 2000, se trouvent au plus bas niveau de compétences dans l'échelle de Prose.

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) (tableau 1). 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, puisque c'est la version suisse qui a été utilisée). À l'aide de ces items, un ancrage psychométrique des données de PISA a pu être effectué sur l'échelle de IALS⁴ (Yamamoto 2002).

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, assez performante à 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 1 rappelle combien les résultats à IALS sont différents. Alors que 41 % de la population française se trouvaient au plus bas niveau de lit-

tératie 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. À 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.

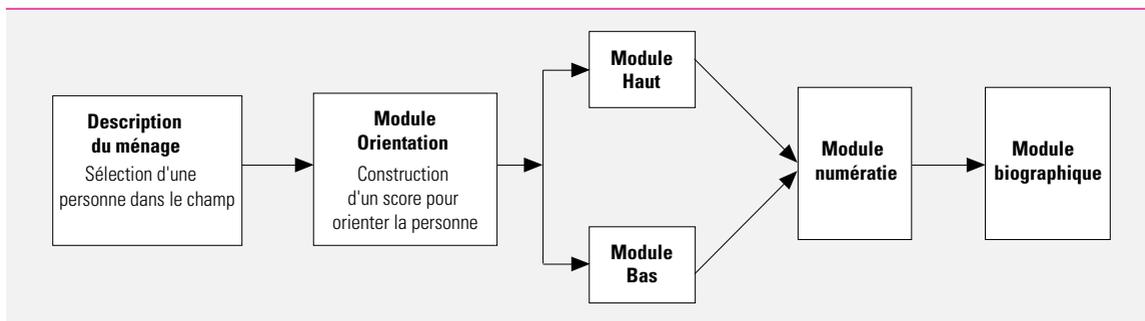
L'ENQUÊTE IVQ

L'enquête IVQ est le résultat de la collaboration d'institutions concernées par l'évaluation des compétences des adultes et d'équipes de recherche universitaire. Deux tests ont été organisés fin 2000 et mi-2002 pour vérifier, après les déboires de l'enquête IALS, qu'un tel type d'interrogation était possible. Un document de travail n° 0202 de la série Méthodologie de collecte de l'Insee : « Enquête méthodologique Information et Vie quotidienne, Tome 1 : bilan du test 1 » présente plus en détail le cadre de l'enquête et les études faites sur le premier test. Sur ce test et sur le deuxième, une contribution a aussi été faite aux journées de la méthodologie statistique en 2002 (Les performances d'adultes à des tests en lecture : comment séparer motivation et compétences ?). À la fin 2002, l'enquête a été entreprise sur un échantillon initial de 4 000 logements.

L'un des principes-clé de l'enquête est d'adapter la passation des épreuves au niveau de la personne. Ainsi, on sélectionne une personne parmi les individus de 18 à 65 ans vivant dans le logement. Si la personne se déclare immédiatement illettrée, l'enquêteur insiste et évoque l'exercice de compréhension orale ; si la personne ne parle pas français, on se contentera de lui poser les questions biographiques. Sinon, on lui propose un module d'orientation avec des questions simples de lecture de mot et de compréhension sur un texte court. Si la personne n'a pas eu de résultats suffisants, elle passera les exercices du module ANLCl ; sinon, on lui propose les exercices du module Haut. Viennent ensuite des questions de « numératie », visant la maîtrise des compétences de base en calcul et en raisonnement logique (là encore, le niveau de difficulté des exercices est adapté en fonction des réponses à quelques questions simples insérées dans le module d'orientation). Ensuite, on recueille un certain nombre d'informations sur le parcours familial, scolaire et professionnel de la personne interrogée, sur ses pratiques de lecture et sur d'éventuelles difficultés à accomplir certains actes de la vie quotidienne, pour les personnes ayant eu des performances médiocres aux tests. À tout moment, l'enquêteur a la possibilité de changer de module, dans le cas où l'enquête se passe mal

NOTE

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).



(problème de temps, trop grande facilité ou difficulté des exercices, etc.). Du fait d'un problème informatique, cette possibilité a eu des conséquences néfastes : les données du module quitté ont été irrémédiablement perdues et il est alors difficile de s'assurer par exemple que la personne était en difficulté sur le module Haut au moment de la réorientation.

Le module d'orientation et le module ANLCl ont été élaborés par l'équipe PsyEf de J.-M. Besse (Université de Lyon II). L'équipe EVA (Université de Rennes et Hôpital Sainte Anne) de C. Charon et C. Meljac ont constitué le module numératie (et les dernières questions du module d'orientation) et enfin, C. Chabrol (Université Paris III) et P. Vrignaud (INETOP) ont élaboré quelques-unes des épreuves du module Haut. Tous les membres du comité de pilotage ont participé également à cette élaboration.

- Le module d'orientation se réduit à un exercice sur une page de programme TV mais comporte plusieurs étapes. Dans un premier temps, on propose des questions de mise en situation, assez floues (difficiles à coder, elles n'interviennent pas dans le score d'orientation) mais non scolaires, ce qui permet de faire accepter par la personne le principe d'évaluation (Qu'est-ce que c'est ? À quoi cela sert-il ?). Ensuite, viennent des questions d'identification de mots (noms d'émission ou d'invité) et un exercice

de compréhension portant sur le film du soir.

- Les questions de numératie sont des problèmes d'une ou deux phrases, donnés oralement pour ne pas créer d'interférence avec la littératie. La lecture de deux nombres et trois petits exercices sont insérés dans le module d'orientation, qui conditionnent la passation de l'épreuve proprement dite composée de 13 questions classées par ordre de difficulté : les personnes n'ayant réussi qu'une question au plus, commencent la chaîne de questions au début, par les plus faciles, tandis que celles qui ont réussi au moins deux questions passent directement à la question 8. Au bout de trois erreurs, on arrête le questionnement. La passation de ce module est indépendante des résultats en lecture.

- Le module Bas, appelé module ANLCl du fait de la collaboration avec cet organisme, commence par un exercice de compréhension orale puis la personne passe une « dictée » (le terme, très connoté, n'est bien sûr pas employé ; il s'agit d'une liste de courses) et des questions d'identification de mot et de compréhension de texte sur un support de la vie quotidienne (un CD de musique).

- Pour le module Haut, 5 textes ont été retenus en plus de l'exercice de compréhension orale commun avec le module ANLCl. Il porte volontairement sur des supports variés : un extrait d'article sur les familles recompo-

sées, une page du *Guide du Routard*, un texte de Victor Hugo sur la peine de mort, des tableaux et graphiques sur les accidents de la route et un récit de match de football (ce dernier texte n'était passé que s'il restait suffisamment de temps). En outre, un tiers de l'échantillon s'est vu proposer une série de six exercices issus de l'enquête IALS de 1994, ce qui va permettre une comparaison des résultats avec des protocoles de collecte différents, dans la suite de cet article.

Il importe de rappeler que l'effort a autant porté sur la constitution d'un protocole de collecte fiable que sur l'élaboration d'épreuves adaptées au public. La « capisation » du questionnaire a facilité la prise d'informations et permis des relations entre enquêteurs et enquêtés plus naturelles. Les réponses sont par ailleurs connues précisément, et surtout il est possible de recueillir deux types d'informations permettant de se faire une idée du degré de motivation de la personne interrogée : le temps qu'elle a consacré à répondre aux questions (la prise de temps se fait informatiquement) et une grille d'observation de son comportement remplie par l'enquêteur (la personne est-elle stressée ? était-elle en colère ? a-t-elle été aidée ? etc.).

Quant au module biographique, il recueillait un certain nombre d'informations sur :

- l'origine sociale de la personne (diplôme et profession des parents) ;

- la scolarité et les langues maternelle et d'apprentissage de la lecture ;
- les événements survenus dans la jeunesse (décès, maladie, situation matérielle, etc.) ;
- la situation professionnelle et l'usage de l'écrit dans le milieu professionnel ;
- les pratiques de lecture (livres, journaux) ;
- les stratégies de contournement des difficultés dans la vie de tous les jours pour les personnes ayant passé le module ANLCI.

Les analyses sur cette première enquête, à visée plutôt méthodologique, ont été plutôt satisfaisantes (Murat, 2004). Il a donc été décidé de la répéter très rapidement, fin 2004, sur un plus gros échantillon pour mener des analyses plus fines. Quelques aménagements ont alors été apportés au protocole pour tenir compte des problèmes rencontrés lors de l'analyse. En particulier, la difficulté à classer les personnes dont les compétences se situent autour de seuil de l'illettrisme a conduit à élaborer un module intermédiaire entre le module Haut et le module ANLCI, destiné aux personnes ayant eu des résultats moyens à l'exercice d'orientation. Ce changement a sans aucun doute amélioré la mesure des compétences, mais rend les résultats de l'enquête de 2004 difficilement comparables à ceux de 2002 (Murat 2005). Cette expérience illustre

à sa façon l'importance du protocole d'observation : un changement assez minime du protocole peut avoir un impact fort sur les résultats.

LA COMPARAISON IALS-IVQ

Une comparaison assez précise entre les méthodologies IALS et IVQ a été rendue possible dans l'enquête de 2002, par la reprise à l'identique d'exercices issus 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, six 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 (alors que PISA avait pris une version corrigée). 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 littératie : *Textes schématiques* ou *Textes quantitatifs*). Il ne serait sans doute pas impossible de contruire une échelle commune aux 16 items, les corrélations étant fortes entre les réussites aux différentes questions, mais pour appliquer la méthodologie de 1994, nous nous sommes vu contraints d'éliminer 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 modifi-

cation 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 importe 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 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.

NOTE

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).

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

- *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.

Sur les 2 086 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 (elles disent ne pas savoir lire ou ne parlent pas le français) : on va imputer des réponses fausses à l'ensemble des exercices IALS (en conformité avec la procédure suivie en 1994), ce qui conduira naturellement à les classer au niveau 1 de l'échelle. 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 en fin de procédure un score en fonction de leurs caractéristiques sociodémographiques¹⁰ (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

NOTES

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 sociodémographiques.

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 4 011 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 physique, absence de longue durée). Sur les 3 060 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 (si elle ne le réussit pas, 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.

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 (cf. *supra* note 6), on ne dispose plus que de leurs réponses à l'exercice d'orientation. Finalement, 547 personnes ont passé l'ensemble des épreuves IALS.

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, apparues en difficulté sensible sur l'exercice d'orientation. Ce choix peut paraître trop sévère pour celles qui ont arrêté l'épreuve en cours : ces personnes sont un peu moins performantes à l'exercice d'orientation que celles qui vont au bout des exercices (elles 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 ANCLI (qui ont, 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 les personnes qui ne l'ont pas passé, c'est-à-dire à la fois pour celles qui l'ont abandonné en cours et pour celles qui ont passé le module ANCLI. Pour ces dernières bien sûr, l'imputation est très fragile et l'hypothèse qui les place directement au niveau 1 sera privilégiée. Il importe de souligner que pour les personnes ayant arrêté l'épreuve IALS en cours, les imputations relèvent de deux logiques : pour les questions que les personnes ont effectivement passées mais pour lesquelles 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'enquêté invoque la

trop grande difficulté des exercices pour abandonner.

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 2* donne l'ensemble des données disponibles (NCES 1998).

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 (en fait les 547 répondants au module IALS) 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 40,0 % 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

Tableau 2 – Réussites aux items repris de IALS dans IVQ et dans IALS (en %)

	IVQ (1)	IVQ (2)	IALS 94 France	IALS 94 autres pays
m1	56,1	49,6	55	77
m2	48,6	44,3	38	62
m4	74,7	67,6	57	77
m5	59,3	53,6	54	65
m6	82,7	78,5	43	69
m7	69,0	66,1	43	60
m9	68,9	63,3	54	60
m10	90,8	87,0	40	53
m11	86,0	78,6	40	73
m12	63,0	58,8	20	55
m13	42,9	38,1	22	27
m14	74,2	67,7	55	70
m16	80,5	75,1	71	87

Lecture – 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 le champ à l'ensemble des 663 personnes ayant passé les épreuves avec imputation à partir des résultats à l'exercice d'orientation. La quatrième 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, États-Unis, Suède, Suisse allemande, Suisse Française).

NOTE

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.

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). Par ailleurs, la suppression de ces trois items ne « favorise » aucune population (ce qui aurait pu être le cas, si l'on n'avait supprimé que des questions très difficiles pour une population, mais faciles pour une autre : la première population aurait été favorisée). En effet, avant suppression des items, le taux de réussite moyen sur les 13 questions est de 70 % pour les répondants au module IALS d'IVQ et de 64 % pour les pays participants à IALS en 1994 (sauf la France). Après suppression, le taux de réussite sur les 10 questions est 69 % pour les répondants au module IALS d'IVQ et de 64 % pour les pays participants à IALS en 1994 (sauf la France). Cette suppression jouerait très légèrement en défaveur de l'échantillon IVQ, puisque le taux de réussite moyen baisse pour lui de 1 point après la suppression, alors qu'il reste identique pour l'échantillon IALS.

NOTE

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.

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 problèmes, 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. Cependant, vu le faible nombre d'items, il convient de prendre les résultats avec précaution.

En 1994, les responsables de l'enquête IALS ont décidé d'utiliser un modèle MRI (modèle de réponse à l'item) à deux paramètres sur les données. L'un des avantages de ce type de méthodologie est de permettre des comparaisons fiables entre des enquêtes utilisant des protocoles d'évaluation partiellement différents. Or, il était souhaitable que les résultats soient comparables avec ceux des précédentes enquêtes américaines, alors que de nouveaux exercices avaient été créés pour tenter de tenir

compte de la diversité culturelle des pays participants.

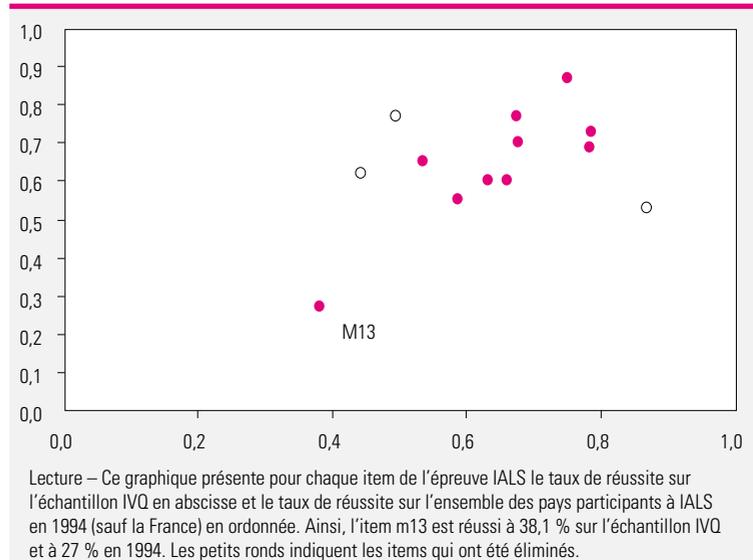
Le modèle MRI synthétise une matrice d'indicatrices de réussite avec en colonnes les items et en lignes les individus : dans la case (j,i) on trouve un 1 si l'individu *j* a réussi l'item *i* ; un 0 s'il l'a raté. Il s'agit de modéliser la présence d'un 1 en fonction de paramètres individuels (la compétence) et de paramètres d'items (la difficulté de l'item, sa discrimination, c'est-à-dire son degré de cohérence avec la dimension mesurée). Ainsi, on pose que la probabilité que l'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, en cherchant les valeurs qui permettent le mieux

Graphique 1 – Comparaison entre la réussite aux items IALS entre l'échantillon global d'IVQ et les autres pays participants à IALS en 1994



de rendre compte des données : la fonction ci-dessus doit être élevée dans les cases où il y a des 1 et faibles dans les cases où il y a des 0. Les techniques pour effectuer cette estimation sont nombreuses, comme celle de la maximisation de la vraisemblance (D'Haultfoeuille *et alii* 2002). Dans le cas présent, pour assurer la comparabilité entre les deux enquêtes, les paramètres a_i et b_i pour les 10 items retenus sont fixés à leur valeur de 1994 et seuls les θ_j sont estimés pour les individus de l'échantillon IVQ. Le score ainsi obtenu est donc sur l'échelle IALS de 1994 et permet la comparaison entre les deux enquêtes.

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.

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

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*.

Comme on l'a indiqué, l'hypothèse qui semble la plus vraisemblable est celle 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 passées) 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).

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.

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 ans¹³. 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'analyse¹⁴. Sur les données IALS, au contraire, elle résiste à l'introduction de ces variables même si le coefficient multiplicatif associé à l'âge est

NOTES

13. 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 ».

14. 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 trois 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.

tout de même divisé par trois (il passe de -1,42 par année supplémentaire à -0,47).

ENSEIGNEMENTS POUR LES ENQUÊTES À VENIR

L'évaluation des compétences est un exercice difficile, particulièrement dans un cadre international (Murat & Rocher 2004). La situation est encore plus complexe dans le cas d'une évaluation d'adultes, car le cadre de référence est moins bien défini pour déterminer ce que l'on souhaite évaluer (il n'y a pas de programmes scolaires) et la maîtrise des conditions de collecte est bien plus difficile que pour des élèves dans un établissement scolaire. Il ne s'agit pas là de précautions d'usage sur des points de détail. La comparaison de IALS avec PISA et celle de IALS avec IVQ montrent à quel point les résultats sont sensibles à la méthode utilisée pour les obtenir. La comparaison IVQ et IALS est particulièrement intéressante, car elle porte sur la même population. Le changement de protocole fait passer de 41 % à 15 % la proportion de personnes en difficulté en France. Cet écart peut s'expliquer par l'amélioration des conditions de collecte, par un protocole moins lourd et plus convivial. L'enquête IALS était sans doute trop longue, présentée sous une forme assez rébarbative (un gros livret d'exercice au lieu des feuillets successifs d'IVQ). Le processus d'orientation

d'IVQ paraît aussi indispensable pour maintenir la motivation des enquêtés, en leur proposant des exercices adaptés à leur niveau. Il est possible aussi que la nature de la dimension évaluée dans les deux enquêtes soit un peu différente, du fait que les réponses sont recueillies à l'oral dans IVQ, alors qu'elles l'étaient à l'écrit pour IALS. Dans ce cas, l'approche IVQ semble préférable, car l'objectif de ce type d'enquête est de mesurer la compréhension des supports et non les capacités d'expression. Si l'usage de l'écrit introduit une difficulté particulière, la mesure de la compréhension est alors perturbée. Bien sûr, ces facteurs ont sans doute aussi affecté les résultats des autres pays participants à IALS, mais il n'est pas possible d'évaluer leur impact hors de France¹⁵ et de déterminer dans quelle mesure le classement français aurait été modifié.

Ces résultats prennent une importance particulière dans la perspective des enquêtes à venir. L'enquête IVQ 2010 va s'inscrire naturellement dans le prolongement d'IVQ 2002 et 2004. Des aménagements seront sans doute apportés, mais ils resteront sans doute mineurs : la difficile comparaison entre IVQ 2002 et IVQ 2004 montre à sa manière la sensibilité des résultats au protocole retenu. L'enquête PIAAC sera le fruit d'arbitrages plus complexes entre les différents pays participants. Quel type d'exercice sera retenu ? Quelle durée aura l'enquête ? Quel mode de recueil des réponses sera-t-il utilisé ? La comparabilité avec IALS est-elle indispensable pour les pays ayant participé à cette enquête et, dans ce cas, l'amélioration du protocole est-elle souhaitable ? De nouveaux tests de questionnaires ou de méthodes de collecte, comparant

les différents choix, seront sans doute nécessaires pour répondre à ces questions, guidant la mise en place d'une enquête de qualité, répondant aux attentes de tous ses participants. ■

NOTE

15. Par ailleurs, les problèmes de comparabilité internationale, en particulier à cause des biais culturels dans les exercices, n'auront ici été évoqués que rapidement, par les travaux de Blum, Gérin-Pace et Vrignaud.

À LIRE

- P. Dickes et P. Vrignaud** (1995) « Rapport sur les traitements des données françaises de l'enquête internationale sur la littératie », Rapport pour le ministère de l'Éducation nationale, Direction de l'évaluation et de la prospective.
- A. Blum et 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.
- A. Blum et F. Guérin-Pace** (2000), *Des lettres et des chiffres*, Fayard.
- C. Baudelot et R. Establet** (1988), « Le niveau intellectuel des conscrits ne cesse ne s'élever », *Éducation & Formations*, ministère de l'Éducation nationale, mars 1988.
- 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.
- F. Murat** (2004), « Les difficultés des adultes face à l'écrit », *Insee Première*, n° 959, avril 2004.
- F. Murat** (2005), « Les compétences des adultes à l'écrit, en calcul et en compréhension orale », *Insee Première*, n°1044, octobre 2005.
- F. Murat et T. Rocher** (2004), "The methods used for international assessments of educational competences", in *Comparing Learning Outcomes International Assessment and Education Policy*, Editors: Jay Moskowitz; Maria Stephens.
- ONS** (2000), *Measuring Adult Literacy* (sous la direction de Siobhán Carey), ONS
- NCES** (1998), *Adult Literacy in OECD countries: Technical Report on the first International Adult Literacy Survey*, NCES.
- OCDE - Statistique Canada** (2000), *La littératie à l'ère de l'information : rapport final de l'enquête internationale sur la littératie des adultes*, OCDE
- OCDE** (2002), *Reading for Change. Performance and engagement across countries. Results from PISA 2000*, OCDE.
- P. Vrignaud** (2001). « Évaluations sans frontières : comparaisons interculturelles dans le domaine de la cognition », in M. Huteau et J. Lautrey (Eds), *Les figures de l'intelligence*. Paris : Éditions et applications psychologiques.
- K. Yamamoto** (2002), *Estimating PISA Students on the IALS Prose Literacy Scale*. Princeton, NJ: Educational Testing Service.