



L'ÉVALUATION DES COMPÉTENCES DES ADULTES

Quelles contraintes ? Quelles spécificités ?

Fabrice Murat

MENESR-DEPP, bureau des études sur les établissements et l'éducation prioritaire

Thierry Rocher

MENESR-DEPP, bureau de l'évaluation des élèves

Évaluer les compétences des adultes est une opération bien plus complexe qu'évaluer les élèves. Les adultes sont sortis souvent depuis fort longtemps de l'école, ne sont plus habitués à la situation d'évaluation et en gardent parfois un mauvais souvenir. De plus, les conditions de passation, au domicile des personnes interrogées, sont aussi plus difficiles que les évaluations dans les salles de classe. Pour résoudre ces problèmes, des méthodologies spécifiques ont été développées ces dernières années, la demande de statistiques sur ce sujet ayant émergé au début des années 1990. Deux dispositifs d'enquêtes se dessinent : les enquêtes internationales (IALS, *International Adult Literacy Survey* ; ALLS, *Adult Literacy and Life Skills Survey* ; Piacac, *Program for the International Assessment of Adult*) sous l'égide de l'OCDE et les enquêtes françaises (IVQ, Information et vie quotidienne) pilotées par l'Insee. Cet article fait le point sur la méthodologie, les contraintes communes à ce type d'enquête et les points où les deux dispositifs se distinguent.

La mesure des compétences des adultes est à la jonction de deux traditions d'enquêtes : les « enquêtes-ménage », organisées notamment par l'Insee sur des thèmes aussi variés que l'emploi, le logement ou la santé, et les évaluations standardisées de compétences, plus généralement pratiquées sur des élèves. Ces deux modes de collecte d'informations statistiques ont leurs spécificités et l'enjeu d'une évaluation des compétences des adultes est d'en tenir compte, dans un protocole bien adapté. La qualité des résultats obtenus est très sensible aux choix faits pour répondre à cet objectif.

Jusqu'au début des années 2000, les enquêtes statistiques menées par l'Insee abordant le thème des compétences des adultes, et plus précisément celui de l'illettrisme, avaient une base déclarative : la personne enquêtée indiquait si elle éprouvait des

difficultés à lire des journaux, à remplir un chèque, etc. Une telle approche est très subjective et l'usage de tests pour vérifier ces déclarations est vite apparu nécessaire. Les évaluations de compétences des élèves, bien plus développées, ont pu servir d'exemple pour construire des outils de mesure, en tenant compte de la contrainte d'une enquête à domicile, dans un cadre moins standardisé que celui de la classe, avec des personnes moins disposées que des élèves à entrer dans une logique d'évaluation.

Le développement des évaluations de compétences des adultes ne s'est pas fait sans difficulté, comme le montre le bilan de la première opération d'envergure ayant eu les mêmes objectifs, l'enquête *International Adult Literacy Survey* (IALS) menée par *Statistic Canada* et *Educational Testing Service* (ETS, organisme responsable de nombreux tests d'évaluation aux États-Unis), dont les résultats ont été diffusés par l'Organisation de coopération pour le développement économique (OCDE). Les conclusions étaient pour la France très surprenantes [NCES, 1998 ; OCDE, 2000] : 40 % des Français entraînent dans la catégorie des plus mauvais lecteurs (vite assimilée à celle des illettrés), ce qui situait la France bien loin derrière la plupart des pays participants (entre autres, l'Allemagne, la Suède, les États-Unis). Des investigations ont mis en évidence un certain nombre de problèmes méthodologiques (traductions non équivalentes des textes et des questions en termes de difficulté, plan de sondage et correction de la non-réponse peu satisfaisants, conditions de passation peu adaptées, etc.) qui ont justifié le retrait de la France de l'opération et la non-diffusion officielle des résultats [BLUM et GUÉRIN-PACE, 2000 ; MURAT, 2008].

Depuis cette date, de gros progrès ont été faits dans les dispositifs nationaux et internationaux. La France a développé son propre dispositif d'enquêtes (les enquêtes Information et vie quotidienne – IVQ) en réponse aux problèmes identifiés dans l'enquête IALS et elle a également accepté de participer au programme Piac (Programme for the International Assessment of Adult Competencies). Cet article décrit les principes et les caractéristiques des évaluations de compétences des adultes, par comparaison avec les évaluations des élèves, et fait un point sur les opérations existantes et à venir [voir aussi DEGORRE et MURAT, 2009 pour une présentation générale sur ces questions].

POURQUOI ÉVALUER LES COMPÉTENCES DES ADULTES ?

L'intérêt d'évaluer les compétences des adultes est multiple et relève d'objectifs un peu différents de ceux des évaluations en milieu scolaire. En effet, on justifie celles-ci soit dans un but diagnostique pour repérer les élèves en difficulté et les aider, soit pour valider les acquis par un examen final, soit pour donner une image globale du système éducatif, en termes de niveau moyen, de disparité, de points forts et de points faibles [TROSSEILLE et ROCHER, dans ce numéro, p. 15]. Les finalités d'une évaluation sur population adulte sont un peu différentes et mettent plutôt l'accent sur l'importance de la lecture et du calcul dans la vie personnelle et professionnelle. Les personnes vivant en France ont-elles un degré de maîtrise suffisant pour faire face aux mutations du monde du travail et de la vie quotidienne, qui nécessitent un accès bien maîtrisé à une information de plus en plus complexe ?

Décrire le niveau de compétence de la population

Contrairement aux évaluations d'élèves, ces opérations peuvent assez difficilement être utilisées pour évaluer le fonctionnement du système éducatif (même si de nombreux journalistes ne se privent pas de le faire) : les personnes évaluées sont sorties parfois depuis très longtemps d'un système scolaire très différent de celui d'aujourd'hui. Cependant, pour faire un tel usage de ces évaluations, on peut se restreindre aux personnes les plus jeunes. L'étude du lien entre les caractéristiques sociodémographiques des individus et les compétences permet aussi de renouveler les connaissances sur les inégalités sociales de réussite scolaire [PLACE et VINCENT, 2009]. De plus, on peut envisager une certaine circularité de cette relation : les parents les moins compétents risquent de ne pas être en mesure d'aider leurs enfants dans leurs études, provoquant chez eux un déficit de compétences [MURAT, 2009] ► **Encadré.**

L'étude de la répartition de la population dans les différents niveaux de compétences fait l'objet de publications nationales [MURAT, 2005 ; JONAS 2012 ; JONAS 2013], mais aussi dans les nombreuses régions, de métropole ou des DOM, qui ont augmenté la taille de l'échantillon pour pouvoir publier des chiffres locaux (tous les DOM ont procédé à ce type d'extension et c'est le cas aussi de l'Aquitaine, du Nord-Pas-de-Calais et des Pays de la Loire en 2004, de la Haute-Normandie, de l'Île-de-France, du Nord-Pas-de-Calais, de la Picardie et de la Provence-Alpes-Côte d'Azur en 2011). Des analyses plus approfondies ont aussi été menées selon le sexe [DJIDER et MURAT, 2006] ou l'âge des individus [MICHEAUX et MURAT, 2006].

LE LIEN ENTRE COMPÉTENCES DES PARENTS ET SCOLARITÉ DES ENFANTS

La sociologie de l'éducation a beaucoup mis en avant – à côté du milieu social mesuré par la profession du père – l'importance du « capital culturel », souvent appréhendé par le diplôme de la mère, pour étudier la réussite scolaire des enfants. L'enquête IVQ propose un indicateur différent de ce capital culturel : le niveau de compétence à l'écrit, en calcul et en compréhension orale. Une analyse menée sur les données de l'enquête IVQ de 2004 montrait l'importance de ces caractéristiques en utilisant le retard scolaire comme indicateur de performance scolaire [MURAT, 2009] : la moitié des enfants de 7 ans à 18 ans dont les parents avaient de faibles compétences en lecture avaient pris au moins une année de retard scolaire contre un cinquième quand les résultats des parents aux tests d'IVQ étaient très satisfaisants. L'écart était à peu près équivalent avec les compétences en

calcul, un peu moindre avec celles en compréhension orale. La prise en compte des autres caractéristiques du ménage (professions et diplômes des parents, revenus, etc.) réduisait ces écarts, mais ils restaient significatifs, indiquant une autre source d'inégalités à l'école que celles classiquement mesurées.

Il a paru utile de reproduire cette analyse sur l'édition de 2011 de l'enquête, car entre-temps, du fait d'une politique de réduction des redoublements, en particulier dans l'enseignement élémentaire, le taux de retard a sensiblement diminué : il est passé de 34 % pour IVQ 2004 à 26 % pour IVQ 2011. La corrélation entre les compétences des parents et le retard scolaire des enfants reste très nette : 40 % des enfants dont les parents ont eu de faibles performances en lecture ont pris du retard contre 14 % pour les enfants dont les parents sont parmi les plus compétents.

► **Tableau 1 Retard scolaire des enfants en fonction des compétences des parents (en %)**

Quartile	Lecture	Calcul	Compréhension orale	Les trois compétences
1	40	40	33	41
2	26	25	26	23
3	22	21	23	23
4	14	16	20	15
Ensemble	26	26	26	26

Lecture : 40 % des enfants se trouvant dans le premier quartile de compétences parentales en lecture (c'est-à-dire les 25 % des enfants dont les parents ont les scores les plus bas dans ce domaine) ont au moins un an de retard scolaire.

Champ : enfants de 7 à 18 ans vivant chez leurs parents en France métropolitaine.

Source : enquête Information et vie quotidienne 2011, Insee.

Du fait de la baisse du taux de retard, la comparaison avec IVQ 2004 n'est pas immédiate : comme ce taux se rapproche de 0, les écarts sont mécaniquement moins marqués. Le recours à une modélisation logistiqu permet de résoudre ce problème et d'introduire des variables de contrôle dans l'analyse pour tenir compte des corrélations entre les deux variables et d'autres caractéristiques de l'enfant ou des parents (sexe, milieu social, revenus, etc.). La prise en compte de l'âge de l'enfant est ainsi indispensable, car il joue fortement sur la probabilité d'avoir redoublé par le passé. Un premier modèle a donc été construit en reliant le retard scolaire avec l'âge de l'enfant et le score moyen des parents dans les trois domaines. Le coefficient associé au score dans IVQ 2011 est de - 0,64. C'est légèrement inférieur à la valeur de 2004 (- 0,72).

Cependant, la modélisation peut être enrichie avec d'autres variables, à la fois liées aux compétences et au retard scolaire. C'est le cas par exemple du diplôme des parents, dont on sait l'importance pour l'étude de la scolarité, et qui est assez logiquement lié à leurs compétences. Cette variable a donc été ajoutée dans l'analyse, ainsi que l'âge, le sexe, le pays de naissance,

la profession des parents, l'âge et le sexe de l'enfant et le revenu du ménage. Le coefficient associé au score dans les trois domaines passe à - 0,34, ce qui est encore une fois un peu inférieur à la valeur de 2004 (- 0,41). Ce coefficient reste toutefois statistiquement significatif et indique une corrélation avec le retard scolaire d'une ampleur à peu près équivalente à celle concernant le diplôme.

Ces résultats pourraient être approfondis comme en 2004 pour tester des effets plus subtils en fonction du sexe du parent par exemple. Rappelons qu'en 2004, une spécialisation était apparue selon la discipline : les compétences en calcul paraissaient plus importantes pour les pères, tandis que pour les mères, ce sont les compétences en lecture qui jouaient le plus. Ce résultat n'est pas apparu clairement dans les premières analyses sur 2011. Plus précisément, des coefficients ne sont significatifs que dans les régressions non pondérées. En 2004, une certaine sensibilité des résultats à l'usage ou non des pondérations avait déjà été mise en évidence, ce qui incite à la prudence et à poursuivre l'expertise méthodologique.

Guider la politique de lutte contre l'illettrisme

Mieux connaître la population des personnes le plus en difficulté face à l'écrit est aussi tout à fait crucial. Il faut distinguer cette population en fonction du degré de difficultés et de leur origine (absence de scolarité, maîtrise de la langue française insuffisante, etc.) afin de cibler l'action de formation sur les différents groupes.

C'est un des objectifs suivis par l'enquête IVQ qui a développé un module spécifique pour mesurer précisément les compétences des personnes en situation d'illettrisme [BESSE, LUIS *et alii*, 2009] et le parcours individuel de ces personnes a fait l'objet d'analyses spécifiques [GUÉRIN-PACE, 2009]. L'Agence nationale de lutte contre l'illettrisme (ANLCI) utilise en effet l'enquête IVQ pour chiffrer l'illettrisme dans ses différents rapports. L'intérêt des institutions locales pour cette enquête s'explique aussi en partie par l'implication de ces institutions dans la lutte contre l'illettrisme et la nécessité de quantifier ce phénomène de façon fine. L'illettrisme a aussi été confronté à d'autres formes de difficultés sociales : la pauvreté [MURAT, 2006] ou le fait d'habiter dans un quartier défavorisé [MURAT, 2007].

Affiner l'analyse du marché du travail

Ces évaluations peuvent aussi servir, en s'inspirant de différentes théories économiques, à mieux comprendre le fonctionnement du marché du travail, en donnant des compétences des personnes interrogées un indicateur plus direct que le diplôme. On peut très rapidement rappeler les deux positions qui s'opposent dans les travaux concernant l'influence de l'éducation sur le marché du travail : d'une part, la théorie du capital humain, qui postule que les études permettent d'accroître les compétences des individus, ensuite valorisées sur le marché du travail, et d'autre part la théorie du signal, qui avance que les compétences sont une donnée préexistante, qui permettent d'atteindre un niveau d'études plus ou moins élevé, les employeurs utilisant alors justement ce niveau d'études comme indicateur de ces compétences. Ces deux théories impliquent une corrélation entre compétences et marché du travail et se distinguent surtout sur la question de la formation de ces compétences, en amont de ce que les évaluations d'adultes permettent d'étudier. Les enquêtes de l'OCDE, IALS et Piacac, ont ainsi donné lieu à de nombreuses analyses sur le lien entre littératie et marché du travail [GREEN et RIDDLE, 2001].

LES SPÉCIFICITÉS D'UNE ÉVALUATION D'ADULTES

Quoi évaluer ?

Dans la phase d'élaboration d'une évaluation de compétence, la question centrale est : que veut-on mesurer ? Les limitations dues aux temps d'enquête sont très fortes : une durée d'évaluation d'environ trois quarts d'heure semble difficile à dépasser si l'on souhaite disposer de réponses données dans de bonnes conditions. C'est pourquoi les évaluations des compétences des adultes se restreignent le plus souvent aux disciplines « fondamentales », la lecture et le calcul, en privilégiant souvent le premier domaine. La compréhension orale a cependant aussi été évaluée en mode mineur dans l'enquête IVQ. Dans Piacac, des compétences plus larges, mobilisées dans le cadre professionnel, sont approchées par un questionnaire déclaratif (*Job Requirement Approach*).

Pour définir plus précisément les dimensions à mesurer, il faut aussi tenir compte des objectifs de l'enquête. Cela fixe le cadre de définition et de mesure des compétences. La question se pose de façon cruciale pour les évaluations d'élèves : dans une optique nationale d'évaluation du système éducatif, il apparaît logique de se référer aux programmes en cours pour construire les épreuves, comme c'est le cas pour le

programme d'évaluations Cedre [TROSSEILLE et ROCHER, dans ce numéro, p. 15]. Dans un cadre international, on a souvent une vision plus large, en se référant plutôt à l'usage des compétences dans la vie personnelle et professionnelle, comme c'est le cas par exemple dans l'évaluation internationale PISA.

Dans le cas des évaluations d'adultes, nationales ou internationales, la référence aux programmes scolaires paraît peu pertinente et c'est donc la deuxième approche qui a été retenue. Ces enquêtes ont ainsi recours aux concepts de « littératie » et de « numératie » pour désigner les compétences de lecture et de calcul mobilisées sur des supports de la vie quotidienne. Ces dimensions sont très larges, elles renvoient à de nombreuses compétences, envisagées dans une perspective utilitariste. Le cadre de construction des épreuves reste donc assez empirique bien qu'il intègre les apports de la psychologie cognitive, s'agissant de notions telles que la compréhension de l'écrit [BESSE, LUIS *et alii*, 2009 ; MEGHERBI, ROCHER *et alii*, 2009] ou la dyscalculie [FISHER et CHARRON, 2009] dans le cas d'IVQ. L'effort porte en fait surtout sur la variété des supports et sur la bonne acceptation des exercices lors des tests.

Cependant, la référence à l'usage des compétences dans la vie quotidienne peut poser problème. En effet, la tentation est grande alors pour certains enquêtés de répondre aux questions à partir de leurs connaissances et expériences personnelles. Selon que ces connaissances et expériences correspondront ou non aux informations données par les supports d'évaluation, la personne interrogée donnera ou non la bonne réponse, sans que ses compétences de compréhension et de raisonnement soient mobilisées. Une des questions de l'enquête Piacac portait sur la langue parlée à la Guadeloupe : on peut se demander combien de personnes en France se sont référées au dépliant touristique faisant office de support pour y répondre.

Prendre en compte la situation d'évaluation

Pour essayer de diminuer ce problème, la compréhension et la restitution des objectifs de l'enquête par les enquêteurs sont fondamentales. D'autant plus que les enquêteurs en charge de l'enquête ne sont pas des spécialistes de l'évaluation. Il est donc important que le protocole (présentation des supports d'évaluation, recueil des réponses, réactions aux demandes de précisions éventuelles, etc.) soit bien standardisé, mais tienne aussi compte des incertitudes d'une enquête au domicile des personnes interrogées, cadre bien moins approprié que la classe dans le cas des évaluations d'élèves. La présence des autres membres de la famille peut ainsi être source de nombreuses perturbations. Ils peuvent ne pas bien prendre l'« intrusion » assez longue de l'enquêteur dans le logement. Ils peuvent aussi vouloir participer à l'évaluation, en aidant la personne interrogée ou en cherchant à prendre sa place. Les enquêteurs doivent savoir gérer ces difficultés, en rappelant la nécessité d'une mesure individuelle des compétences ou par exemple en donnant une copie des épreuves au « perturbateur », en lui demandant d'y répondre de son côté. Proposer de baisser le son de la télévision est aussi un réflexe que les enquêteurs acquièrent assez vite.

Les formations aux enquêteurs permettent de bien définir les enjeux de l'enquête et de donner quelques consignes pour gérer ce type de situation. La qualité des résultats dépend du professionnalisme des enquêteurs dans le respect des consignes et

de leur capacité à s'adapter aux situations imprévues. En particulier, les critiques adressées à l'enquête IALS concernant les conditions de passation [BLUM et GUÉRIN-PACE, 2000 ; CAREY, 2000] ont bien montré qu'une attention particulière doit porter sur les relations entre enquêteurs et enquêtés, afin de favoriser et de maintenir la motivation des enquêtés au cours de l'évaluation. La création de l'enquête IVQ a été une réponse à ce problème, en proposant un protocole plus naturel [VALLET, BONNET *et alii*, 2002]. En particulier, il a été décidé de donner les exercices les uns après les autres et non pas globalement dans un livret. Le test paraît alors moins lourd et les échanges, bien cadrés, entre enquêteurs et enquêtés entre deux exercices, permettent d'atténuer l'impression d'examen.

Si des efforts importants peuvent être faits pour obtenir des conditions de passation les plus standardisées et confortables possibles, la situation sur le terrain n'est pas toujours idéale. C'est pourquoi les enquêteurs doivent recueillir des informations sur le déroulement de l'enquête. Cela peut prendre la forme d'une grille d'observation, où l'enquêteur va rendre compte, parfois exercice par exercice, de la façon dont l'enquêté s'est investi dans l'évaluation. L'informatisation des épreuves permet aussi facilement d'enregistrer les temps passés sur chaque exercice, qui, trop courts, peuvent être le signe d'un manque de motivation [MURAT et ZAMORA, 2002].

Adapter l'épreuve

Les enquêtes auprès des adultes portent sur une population beaucoup plus hétérogène, en termes de niveaux de compétences, que celles portant sur des élèves, généralement menées à un niveau scolaire donné ou à un âge donné. Les personnes interrogées ont connu des parcours scolaires très variés, parfois hors de France, dans une autre langue que le français, à des époques plus ou moins éloignées. Il faut donc construire un protocole d'évaluation qui puisse s'adapter à toutes les situations. C'est pourquoi les évaluations d'adultes comportent généralement un processus d'« orientation », qui permet d'adapter la difficulté des épreuves au niveau de compétences de la personne. Il ne faut pas décourager les personnes en difficulté par des questions trop difficiles ; il ne faut pas non plus perturber les personnes très compétentes avec des questions trop faciles, car elles risquent de se démotiver ou de chercher des difficultés où il n'y en a pas. Pour les plus en difficulté, c'est la nature même des épreuves qui peut être modifiée, la mesure de la compréhension de texte pouvant être utilement complétée par celle du décodage des mots et des procédures élémentaires d'écriture, comme c'est le cas dans IVQ.

Cependant, cette nécessité d'adapter les épreuves conduit à des protocoles d'évaluation assez complexes, composés de filtres permettant l'orientation des individus vers des épreuves différentes, en fonction de leurs résultats obtenus, au fur et à mesure qu'ils avancent dans l'évaluation. Avec cette procédure, appelée *multistage adaptive testing*, l'estimation des niveaux de compétences des individus est complexe [MURAT et ROCHER, 2009]. En effet, par construction, le niveau de difficulté des épreuves proposées dépend du niveau de compétence des individus. Les individus les plus performants passent des épreuves difficiles ; les individus les moins performants des épreuves faciles. Dès lors, il n'est pas possible de comparer directement les individus en fonction de leurs résultats obtenus aux différentes épreuves. Les modèles psychométriques permettent précisément d'envisager de

façon séparée les deux concepts que sont le niveau de compétence des individus et le niveau de difficulté des épreuves [ROCHER, dans ce numéro, p. 37]. Pour ce faire, il est nécessaire de disposer d'épreuves communes à des individus de niveaux différents. Les enquêtes IVQ et Piac adoptent ce principe, cependant avec des modalités différentes que nous détaillons plus loin.

DEUX FAMILLES D'ÉVALUATION DES ADULTES

Les enquêtes internationales

La première enquête à grande échelle sur les compétences en littératie des adultes a pris place en 1985 aux États-Unis. Le *Young Adult Literacy Survey* (YALS) a été organisé par ETS avec le soutien du *National Center for Educational Statistics* (NCES). Restreinte au champ des jeunes adultes, l'enquête YALS a permis d'expérimenter pour la première fois un dispositif d'évaluation dont les épreuves sont fondées sur des supports diversifiés, avec des niveaux de difficulté divers. S'appuyant sur les résultats de cette première enquête, le *National Adult Literacy Survey* (NALS) a été conduit en 1989-1990 sur un échantillon couvrant l'ensemble de la population adulte des États-Unis. Trois grandes familles d'épreuves ont été proposées. La dimension *Prose Literacy* mesure la capacité à comprendre et à utiliser de l'information organisée à travers des phrases, elles-mêmes structurées en paragraphes. Des textes narratifs, mais aussi de la poésie, ont ainsi été repris à partir de journaux, magazines ou brochures, en préservant la typographie et la mise en page originale. La dimension *Document Literacy* s'intéresse à l'aisance pour manipuler de l'information structurée en matrice, c'est-à-dire à travers des lignes et des colonnes. Les supports reprennent des tableaux, des tickets, des graphiques, des grilles horaires, etc. La dimension *Quantitative Literacy* correspond à la notion d'arithmétique quotidienne : additions, soustractions, multiplications et divisions prennent place à travers des mises en situation (calculer une remise sur un achat, trouver le coût d'un emprunt, etc.).

Ces méthodes ont également inspiré, au niveau international, l'enquête *International Adult Literacy Survey* (IALS), organisée par Statistique Canada et ETS, puis coordonnée par l'OCDE dans une vingtaine de pays entre 1994 et 1999. En se fondant sur des travaux de chercheurs américains spécialisés en psychométrie, l'enquête IALS a donné lieu à des études comparatives [OCDE, 2000] sur les niveaux de compétences observés dans les pays ayant participé au dispositif. Une approche fondée sur les modèles de réponse à l'item (MRI) a ainsi été mobilisée pour construire une échelle internationale commune, à partir d'items traduits dans chaque langue [ROCHER, dans ce numéro, p. 37].

L'OCDE, assistée de Statistique Canada et de ETS, a décidé de lancer en 1999 une autre enquête sur les compétences des adultes, *Adult Literacy and Life Skills Survey* (ALLS), étendant son champ d'investigation à d'autres compétences comme la « résolution de problèmes » (*Problem Solving*). Les deux dimensions *Prose Literacy* et *Document Literacy* ont été fusionnées en une seule échelle. Une évaluation en numératie, portant plus spécifiquement sur les compétences de calcul, s'est substituée à la *Quantitative Literacy*. Les principes de conception du questionnaire IALS ayant été repris à l'identique, malgré

les critiques adressées, la France a décidé de ne pas participer à ce projet et de développer en premier lieu un cadre national d'évaluation des compétences.

En 2007, l'OCDE a lancé l'opération Piacac (*Program for the International Assessment of Adult*) auquel ont participé 24 pays. Les compétences mesurées restent les mêmes que pour ALLS (avec une informatisation du protocole sur laquelle nous allons revenir). Cependant, le protocole tenait compte en partie des enseignements de l'expérience française. La France a décidé de participer à cette enquête en 2012, qui a concerné 7 000 personnes âgées de 16 ans à 65 ans. Les premiers résultats ont été diffusés à partir de la fin de l'année 2013 [OCDE, 2013 ; JONAS, 2013].

L'enquête IVQ

Suite aux problèmes rencontrés par l'enquête IALS et au refus de la France de participer à l'enquête ALLS, un comité de pilotage a été institué pour mettre au point un protocole rigoureux et adapté à la réalisation d'une enquête ménage en France¹. Les objectifs définis dans le cadre de ce partenariat ont façonné la forme prise par l'enquête Information et vie quotidienne. La rencontre entre la recherche en psychologie² et l'ingénierie statistique a permis de donner corps à ces objectifs, en adaptant des exercices d'évaluation des compétences puis en les orchestrant sous la forme d'une enquête ménage, tout en prenant soin de répondre aux biais potentiels posés par le protocole et en particulier les relations entre enquêteurs et enquêtés. Un premier test de l'opération a été mené en décembre 2000 sur quelques centaines d'individus pour s'assurer que le principe même d'une évaluation à domicile était possible. On a constaté que les personnes interrogées se prenaient au jeu, grâce à des supports variés et ancrés dans la vie quotidienne (un programme TV, un CD, des cartes routières, etc.), mais que les problèmes de motivation (attention accordée au questionnaire, ou lassitude en fin d'épreuve) se posaient toujours de façon cruciale. C'est pourquoi un effort particulier a été fait pour capter de l'information sur le degré de motivation de chaque enquêté. Le deuxième test, sur un échantillon du même ordre, en avril 2002, a servi à choisir et améliorer des épreuves pour une première édition de l'enquête sur un échantillon important [VALLET, BONNET *et alii*, 2002]. L'expertise fine des données a permis de valider le protocole et d'envisager la mise en place d'une enquête IVQ 2004 sur un échantillon étendu à l'ensemble de la France métropolitaine. Le projet a été présenté devant les instances nationales en charge de coordonner les travaux de la statistique publique et d'en vérifier la qualité – le Conseil national de l'information statistique et le Comité du label. Confortée par les premiers résultats méthodologiques, l'enquête IVQ 2004 a bénéficié du label d'intérêt général et de qualité statistique.

1. Dans le groupe de pilotage, se trouvent représentés l'ANLCI (Agence de lutte contre l'illettrisme), 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, de l'Enseignement supérieur et de la Recherche), 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).

2. Plusieurs équipes universitaires ont travaillé sur les exercices proposés : PsyEf (Lyon 2) sur le module d'orientation et module « illettrés », EVA (Rennes 2) sur le module d'orientation et module « numératie », Paris V et Paris XII sur le module « haut ».

Conduite dans l'ensemble des régions de France métropolitaine, l'enquête IVQ 2004 s'est appuyée sur un échantillon de 17 300 logements, avec des critères de pondération qui augmentaient les probabilités de trouver une personne en difficulté (notamment, chef de ménage peu diplômé ou né à l'étranger). Un sur-échantillonnage a été réalisé pour obtenir des résultats sur les zones urbaines sensibles (ZUS). Le protocole a également été repris dans les DOM par la suite avec des adaptations pour préserver la proximité sémantique des exercices à la vie quotidienne des populations enquêtées : par exemple, les noms des villes dans les textes ont été modifiés, tout en préservant les difficultés formelles qu'ils doivent présenter.

En 2011, une nouvelle édition de l'enquête a eu lieu en France métropolitaine. Elle a concerné 14 000 personnes et a permis d'établir des comparaisons avec l'enquête de 2004 [JONAS, 2012].

Comparaison entre les deux enquêtes

La participation de la France à l'enquête Piacac montre que, depuis IALS, les évaluations internationales d'adultes ont sensiblement amélioré leur méthodologie et pris en compte les travaux en France sur IVQ. Cependant, il reste un certain nombre de différences, qui peuvent expliquer des divergences dans les résultats [JONAS, LEBRÈRE *et alii*, 2013].

Dans l'ensemble, pour ce qui concerne la maîtrise de l'écrit, les deux enquêtes s'appuient sur une définition assez proche de ce que l'on souhaite évaluer : la capacité à utiliser l'information dans la vie quotidienne. Les deux enquêtes se distinguent plutôt dans le traitement des personnes les plus en difficulté. Les responsables de Piacac ont développé un module spécifique pour les personnes qui montraient très vite des difficultés particulières aux premiers exercices, mais la France n'a pas souhaité utiliser ce module. En effet, l'étude des processus élémentaires de la lecture, incluant le déchiffrement de mots, est apparue peu pertinente dans une perspective internationale, avec des langues aux orthographes et aux alphabets très variés. L'enquête IVQ a développé des exercices spécifiques à ce sujet. Par ailleurs, la mesure des compétences en calcul, évaluée en mode mineur dans IVQ, s'appuie sur des principes théoriques assez différents de la numératie utilisée dans Piacac. Dans IVQ, il s'agit, à partir de problèmes courts et d'une forme un peu scolaire, de mesurer la maîtrise des outils de bases des mathématiques (additions, soustractions, proportionnalité, logique), alors que Piacac contextualise davantage les exercices, en essayant de les rapprocher de l'utilisation des compétences en mathématiques dans la vie quotidienne.

Si les deux enquêtes adoptent le principe d'une procédure adaptative, d'un point de vue pratique, le protocole adaptatif de Piacac est plus complexe que celui d'IVQ. L'enquête Piacac distingue dans un premier temps, à partir des réponses au questionnaire biographique et à travers un court test informatique, ceux qui maîtrisent l'outil informatique et les autres. Ces derniers passeront une version papier des épreuves, commençant par huit questions. À l'issue de ces huit questions, l'évaluation est interrompue pour les individus les plus faibles (dans certains pays, ils passent alors une évaluation spécifique). Les autres passent soit une épreuve en littératie, soit une épreuve en numératie. Pour les personnes qui ont des compétences suffisantes en informatique, l'évaluation se fait en cinq étapes : une première épreuve permet

d'interrompre l'évaluation pour les individus les plus faibles. Une deuxième épreuve est proposée en littérature ou en numératie (la passation d'un des deux domaines en premier est aléatoire) en fonction des résultats à l'épreuve précédente et du niveau d'études connu par le questionnaire biographique : les individus ayant eu de bons résultats et/ou un bon niveau d'études ont plus de chances de se voir proposer des exercices difficiles³. Une troisième épreuve est proposée dans le même domaine, là encore adaptée en fonction des réponses aux exercices précédents. Enfin, les deux dernières épreuves reprennent le même mécanisme pour le domaine qui n'a pas été évalué en premier. Cette procédure nécessite à la fois une estimation très précise en amont de l'enquête de la difficulté des exercices et une correction automatique des premières réponses des enquêtés.

Le protocole d'IVQ est plus simple dans la mesure où il ne comporte qu'un seul processus d'orientation (*two-stage adaptive test*). Après une première épreuve dite d'orientation, trois choix sont possibles en fonction des résultats obtenus : un module spécifique est destiné aux personnes en difficulté face à l'écrit ; un module « haut » est proposé aux personnes qui maîtrisent les compétences élémentaires de la lecture ; un module intermédiaire est proposé aux personnes ayant eu des résultats passables à l'épreuve d'orientation et permet de préciser laquelle des deux orientations ci-dessus est préférable. En 2011, deux innovations ont été introduites. D'une part, le module « haut » a été enrichi de manière à mieux décrire les compétences des individus qui n'ont pas de difficulté de lecture : ces individus sont alors orientés aléatoirement vers un module « haut A » ou bien vers un module « haut B », deux modules qui comportent un exercice en commun afin de relier les résultats. D'autre part, un « module commun » est proposé à tous les individus en fin d'évaluation. Ce module permet d'affiner la mesure du niveau de compétences des individus situés à la frontière du seuil d'orientation et de consolider la construction d'un score commun à tous les individus.

Les principales différences entre les deux enquêtes portent très certainement sur les conditions de collecte : durée de la passation, relations enquêteurs-enquêtés, support de l'évaluation (papier/informatique).

Notons tout d'abord que l'enquête Piac est particulièrement longue : elle peut facilement atteindre 2 heures, alors que l'enquête IVQ a été calibrée pour une passation d'un peu plus d'une heure. De plus, dans Piac, le choix a été fait de faire passer d'abord un long questionnaire biographique (une heure environ), au risque de proposer les exercices à des enquêtés déjà lassés. Le fait de faire passer le questionnaire biographique avant les épreuves cognitives apparaît en cohérence avec les principes d'estimation des compétences qui utilisent les informations biographiques et les résultats aux épreuves pour en déduire des « valeurs plausibles » des niveaux de compétences [ROCHER, dans ce numéro, p. 37].

Un point crucial concerne les relations entre enquêteurs et enquêtés. En effet, l'expertise française sur IALS avait montré que la passation sous forme de cahier avait un caractère peu naturel et stressant pour l'enquêté : pendant qu'il travaillait

3. L'orientation n'est pas déterministe et comporte un facteur aléatoire pour contourner le problème d'estimation évoqué dans la partie « adapter l'épreuve », p. 89. En outre, il existe de nombreux recouvrements entre les différentes épreuves qui comportent des questions communes.

sur les exercices, l'enquêteur était inactif et l'enquêté pouvait avoir l'impression de lui faire perdre son temps. Dans l'enquête IVQ, ce problème avait été contourné en faisant donner les exercices un par un par l'enquêteur et en l'impliquant dans la saisie des réponses (ainsi cela permet de tester plus précisément la compréhension sans interférence avec les capacités d'expression, qui sont sollicitées si l'enquêté doit écrire ou saisir sa réponse). De plus, cette procédure permet de concentrer l'attention de l'enquêté sur chaque exercice, sans qu'il soit tenté de faire un choix en feuilletant le cahier et en s'arrêtant sur les supports qui l'intéressent le plus. L'enquête Piac est restée sur le principe d'une évaluation auto-administrée, en passant sur un support électronique. Le rôle de l'enquêteur est volontairement limité au minimum et il reste un témoin silencieux, inutile, voire un peu gênant pendant la passation des épreuves. Cette approche interroge sur la possibilité de maintenir un degré suffisant de motivation des enquêtés au cours de la passation d'épreuves n'ayant aucun enjeu pour eux. Ainsi, dans IVQ, les enquêteurs suivent une formation préalable approfondie, pour que les interactions avec les enquêtés soient les plus naturelles possible. L'enquêteur doit maintenir la motivation de l'enquêté, sans pour autant lui donner des indications sur la qualité de ses réponses. En particulier, les demandes de correction et d'explication sont renvoyées à la fin de l'enquête, pour que les personnes passent les exercices dans les mêmes conditions : au sein d'un même exercice, corriger la première question peut ainsi aider ceux qui le demandent, à répondre aux questions suivantes.

Les procédures de passation constituent un élément clé de la qualité de ces enquêtes. MURAT [2009] a pu montrer qu'à partir des mêmes items utilisés par IALS, mais avec les procédures de passation d'IVQ, le taux de mauvais lecteurs s'établissait à 15 %, contre 40 % dans IALS. Ce résultat issu d'IVQ était d'ailleurs proche de celui obtenu à partir de PISA 2000 pour les élèves de 15 ans, ce qui renforce sa vraisemblance et confirme que la collecte de IALS n'a pas dû se faire dans les meilleures conditions. Le professionnalisme des enquêteurs et le soin mis à les former ont donc une importance fondamentale dans les résultats. Cependant, sur ce point, l'enquête Piac ne se distingue pas autant d'IVQ que IALS, car les enquêteurs sont issus du réseau des enquêteurs professionnels de l'INSEE, ce qui renforce la standardisation et limite les biais de passation. Les enquêteurs ont par ailleurs bénéficié d'une formation poussée pour les aider à gérer la situation particulière d'évaluation dans le cadre d'une enquête à domicile.

Le passage au support électronique est une innovation propre à l'enquête Piac. Les responsables de Piac ont ainsi fait l'hypothèse qu'il y a une équivalence parfaite entre passation électronique et passation sur support papier. En particulier, ils pensent être en mesure de comparer les résultats de Piac avec ceux de IALS et de ALLS, dont les épreuves ont été passées sur support papier. Or, des travaux réalisés sur des évaluations d'élèves montrent la fragilité de l'hypothèse d'équivalence entre les deux supports [BESSONNEAU, ARZOUManIAN, PASTOR, dans ce numéro, p. 159]. Ce problème a d'ailleurs été mis en exergue lors de l'expertise de l'enquête pilote française de Piac qui avait montré des différences de difficulté très sensibles pour certains exercices entre les personnes qui le passaient sous forme électronique et celles qui le passaient sous format papier (ces décalages apparaissaient beaucoup plus nets pour les exercices de littérature que de numératie). De plus, la procédure d'orientation de Piac peut sembler en contradiction avec ce principe d'équivalence : les individus qui montrent une maîtrise trop médiocre des procédures électroniques (cliquer, surligner, etc.),

passent une version papier aussi difficile que les exercices électroniques, en faisant l'hypothèse, par ailleurs vérifiée, qu'ils peuvent obtenir des résultats tout aussi bons que les personnes qui maîtrisent les procédures informatiques. Cela semble signifier implicitement qu'une dimension « maîtrise de l'informatique » peut perturber la mesure des compétences en compréhension. Il est regrettable que le protocole final n'ait pas prévu la passation d'exercices papier pour une partie au moins des personnes n'ayant pas de difficulté avec l'informatique.

Enfin, nous ne développons pas ici les aspects liés aux comparaisons internationales, qui sont de fait une différence entre les deux enquêtes. Cependant, il est à noter que ces problèmes d'équivalence entre les deux supports posent de sérieuses questions quant à la validité des comparaisons internationales. En effet, il existe des différences importantes entre pays du niveau de maîtrise, objectif ou déclaré, des compétences informatiques. Par conséquent, le nombre de personnes concernées par la passation informatique ou la passation papier-crayon peut être très variable d'un pays à l'autre : ainsi, il s'avère que 37 % des Japonais ont passé la version papier-crayon alors que cela n'a concerné que 10 % des Néerlandais. La validité des comparaisons internationales est donc fortement soumise à l'hypothèse d'équivalence des deux versions, dans chacun des pays.

QUEL AVENIR POUR LES ÉVALUATIONS D'ADULTES ?

Malgré une certaine proximité dans les principes et dans les méthodes, les enquêtes internationales et françaises se distinguent sur de nombreux aspects méthodologiques, en particulier pour ce qui concerne les conditions de collecte des données. Les résultats globaux des dernières enquêtes IVQ et Piacac apparaissent cependant relativement cohérents : par exemple le lien entre le niveau de compétence et le niveau de diplôme est assez comparable d'une enquête à l'autre [JONAS, LEBRÈRE *et alii*, 2013]. Au final, c'est sans doute la capacité de ces enquêtes à répondre ou non à certaines problématiques qui constitue la véritable ligne de démarcation.

Ainsi, la possibilité de situer la France par rapport aux autres pays est évidemment un atout des enquêtes internationales. Il faut cependant en rappeler la contrainte principale : une enquête internationale est issue d'un consensus entre pays qui empêche d'adapter au mieux le protocole à la population d'un pays particulier. Le chiffrage de l'illettrisme, par exemple, s'appuie sur la mesure de compétences élémentaires, étroitement associée aux caractéristiques de la langue française. Comme nous l'avons indiqué dans la partie « comparaison entre les deux enquêtes » p. 92, la recherche d'une comparabilité internationale dans ce domaine apparaît peu pertinente, en particulier en ce qui concerne le décodage de mots, dont la traduction ne rend pas compte des différences de difficulté de lecture d'un pays à l'autre, en raison notamment de la différence de prononciation, inhérente à la langue considérée.

Dans une enquête nationale, il est aussi plus facile d'intégrer dans le questionnaire biographique des questions précises répondant aux préoccupations des

institutions et des chercheurs français (lien avec la pauvreté, avec la politique de la ville, etc.). Un questionnaire biographique international doit à la fois garantir la comparabilité des concepts et répondre aux demandes des différents partenaires qui s'intéressent à des domaines divers (lien des compétences avec le marché du travail, avec la formation continue, avec la santé, etc.). Cela permet de découvrir des problématiques variées et parfois novatrices dans une perspective française (la mesure indirecte des compétences utilisées au travail, la *Job Requirement Approach* par exemple, est apparue très séduisante), mais cela conduit à proposer aux enquêtés un questionnaire très long, dépassant une heure, ce qui complique la collecte et peut provoquer la lassitude des enquêtés en particulier lors du passage des épreuves.

Enfin, la territorialisation des résultats est un point particulièrement important. L'enquête Piacac permet naturellement des comparaisons internationales, mais l'enquête IVQ a l'intérêt de proposer des déclinaisons régionales, répondant aux besoins des institutions locales. On ne pourrait pas régler le problème posé par la disparition d'IVQ en proposant des extensions régionales de Piacac. En effet, l'intérêt des institutions locales pour IVQ vient surtout de la mesure de l'illettrisme que cette enquête propose. L'enquête Piacac est, on l'a dit, beaucoup moins bien placée pour répondre à ce type de demande.

Les enquêtes IVQ et Piacac apparaissent donc éminemment complémentaires. Bien que le coût important des évaluations d'adultes amène régulièrement à s'interroger sur leur opportunité, la coexistence de ces deux dispositifs est justifiée, tout comme il existe à la fois des évaluations internationales et nationales d'élèves.

BIBLIOGRAPHIE

BESSE J.-M., LUIS M.-H., BOUCHUT A.-L., MARTINEZ F., 2009, « La mesure des compétences en traitement de l'écrit chez les adultes en grande difficulté », *Économie et statistique*, n° 424-425, Insee.

BLUM A., 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.

CAREY S., 2000, *Measuring Adult Literacy – The International Adult Literacy Survey in the European Context*, London, Office for National Statistics.

DEGORRE A., MURAT F., 2009, « La mesure des compétences des adultes, un nouvel enjeu pour la statistique publique », *Économie et statistique*, n° 424-425, Insee.

DJIDER Z., MURAT F., 2006, « Des chiffres pour les hommes... des lettres pour les femmes », *Insee première*, n° 1071, Insee.

FISCHER J.-P., CHARRON C., 2009, « Une étude de la dyscalculie à l'âge adulte », *Économie et statistique*, n° 424-425, Insee.

GREEN D. A., RIDDLE W. C., 2001, « Les capacités de lecture et de calcul et la situation sur le marché du travail », *Statistique Canada*, n° 8, coll. « Enquête internationale sur l'alphabétisation des adultes ».

GUÉRIN-PACE F., 2009, « Illettrisme et parcours individuels », *Économie et statistique*, n° 424-425, Insee.

JONAS N., 2013, « Les capacités des adultes à maîtriser des informations écrites et chiffrées – Résultats de l'enquête Piac 2012 », *Insee Première*, n° 1469, Insee.

JONAS N., 2012, « Pour les générations les plus récentes, les difficultés des adultes diminuent à l'écrit, mais augmentent en calcul », *Insee Première*, n° 1426, Insee.

JONAS N., LEBRÈRE A., POMMIER P., TROSSEILLE B., 2013, « Mesurer les compétences des adultes – Comparaison de deux enquêtes », *Insee Analyses*, n° 13, Insee.

MEGHERBI H., ROCHER T., GYSELINCK V., TROSSEILLE B., TARDIEU H., 2009, « Évaluation de la compréhension de l'écrit chez l'adulte », *Économie et statistique*, n° 424-425, Insee.

MICHEAUX S., MURAT F., 2006, « Les compétences à l'écrit, en calcul et en compréhension orale selon l'âge », *Données Sociales – La société française*, Insee.

MURAT F., 2009, « Le retard scolaire en fonction du milieu parental : l'influence des compétences des parents », *Économie et statistique*, n° 424-425, Insee.

MURAT F., 2008, « L'évaluation des adultes : des méthodes en plein développement », *Éducation & Formations*, n° 78, MEN-DEPP.

MURAT F., 2007, « Maîtrise du français et du calcul chez les adultes dans les ZUS », *rapport 2006 de l'ONZUS*.

MURAT F., 2006, « Les compétences des adultes et l'exclusion sociale », *Travaux de l'Observatoire – 2005-2006*, ONPES.

MURAT F., 2005, « Les compétences des adultes à l'écrit, en calcul et en compréhension orale », *Insee première*, n° 1044, Insee.

MURAT F., ROCHER T., 2009, « Création d'un score global dans le cadre d'une épreuve adaptative », *Économie et statistique*, n° 424-425, Insee.

MURAT F., ZAMORA P., 2002, « Les performances d'adultes à des tests en lecture : comment séparer motivation et compétences ? », *contribution aux journées de méthodologie statistique*.

NCES, 1998, *Adult Literacy in OECD countries: Technical Report on the first International Adult Literacy Survey*, NCES.

PLACE D., VINCENT B., 2009, « L'influence des caractéristiques sociodémographiques sur les diplômes et les compétences », *Économie et statistique*, n°424-425, Insee.

OCDE, 2013, *Perspectives de l'OCDE sur les compétences 2013 – Premiers résultats de l'évaluation des compétences des adultes*, Paris, OCDE.

OCDE, Statistique Canada, 2000, *La littératie à l'ère de l'information : rapport final de l'enquête internationale sur la littératie des adultes*, Paris, OCDE, Statistique Canada.

VALLET L.-A., BONNET G., EMIN J.-C., LEVASSEUR J., ROCHER T., BLUM A., GUÉRIN-PACE F., VRIGNAUD P., HAULTFOEUILLE X. D', MURAT F., VERGER D., ZAMORA P., 2002, « Enquête méthodologique Information et vie quotidienne – Tome 1 : bilan du test 1, novembre 2002 », *Document de travail, série Méthodologie statistique*, n° C0202, Insee.