

Cette collection propose, en quatre pages, des résumés de travaux d'analyse conduits à l'Insee, généralement publiés sous forme de documents de travail, articles ou dossiers. Elle vise à mieux faire connaître leurs résultats.

Mesurer les compétences des adultes : comparaison de deux enquêtes

Nicolas Jonas et Alexandre Lebrère (INSEE),
Patrick Pommier (DARES) et Bruno Trosseille (DEPP)

- **Les compétences à l'écrit et en calcul de la population adulte ont fait récemment l'objet de deux enquêtes : l'enquête IVQ dont les résultats avaient été publiés fin 2012, et l'enquête internationale PIAAC dont les résultats paraissent cet automne.**
- **L'architecture de la nouvelle enquête IVQ permet de calculer un score de compétence de même type que celui mis en œuvre par PIAAC. Les caractéristiques des populations présentes aux deux extrêmes de la distribution de ces scores sont similaires.**
- **Mais les deux enquêtes n'identifient pas de la même façon les populations en difficulté. À l'aide d'une approche multidimensionnelle, combinant trois dimensions fondamentales de la maîtrise de l'écrit (compréhension, production et identification de mots), IVQ estime à 16 % la part des adultes en difficulté dont 11 % de personnes en difficulté grave ou forte.**
- **PIAAC estime à 21,5 % la part de la population en difficulté vis-à-vis de l'écrit, mais en privilégiant une analyse unidimensionnelle et à partir d'un protocole de collecte nécessitant une certaine familiarité avec l'outil informatique. Son enseignement est donc différent : elle s'attache à évaluer la maîtrise des processus qui sont sollicités lors de l'utilisation des technologies de l'information et de la communication.**

En 2011, d'après l'enquête Information et Vie Quotidienne (IVQ), 16 % des personnes de 16 à 65 ans résidant en France métropolitaine connaissaient des difficultés en calcul (Jonas, 2012). Elles étaient également 16 % à éprouver des difficultés dans les domaines fondamentaux de l'écrit. Pour 11 % ces difficultés étaient graves ou fortes.

En 2012, d'après l'enquête internationale sur les compétences des adultes (PIAAC – *Programme for the International Assessment of Adult Competencies*), 28 % des personnes de 16 à 65 ans se situent dans le groupe de niveau 1 ou inférieur à 1 de compétence en « numératie » (*sources et méthodes*), à un niveau leur permettant de poser une opération arithmétique simple avec des nombres

entiers. En « littératie » (*sources et méthodes*), 21,5 % des personnes se situent dans les mêmes groupes, soit à un niveau de compétence permettant de localiser une information dans un texte bref, mais plus rarement de comprendre parfaitement un texte long ou à la structure élaborée (OCDE, 2013).

Sur des domaines qui semblent si voisins (numératie/calcul, littératie/maîtrise de l'écrit) comment peut-on expliquer que deux enquêtes portant sur la population des 16-65 ans résidant en France dans la même période produisent des résultats apparemment discordants ?

Comment mesure-t-on les compétences dans les deux enquêtes IVQ et PIAAC ?

L'enquête IVQ 2011, prolongeant l'édition de 2004, a été menée par l'Insee en collaboration avec la Direction de l'animation de la recherche, des études et des statistiques (Dares) du ministère du Travail, de l'emploi, de la formation professionnelle et du dialogue social et la Direction de l'évaluation, de la prospective et de la performance (Depp) du ministère de l'Éducation Nationale. La collecte s'est déroulée en 2011, auprès de 14 000 personnes âgées de

16 à 65 ans. Le questionnaire débute par les exercices de mesure des compétences. La première série, appelée module « d'orientation » identifie les adultes en difficulté dans les domaines les plus fondamentaux de la compréhension de l'écrit : l'identification de mots et la compréhension de textes courts. Si l'enquête commet peu d'erreurs, il est orienté vers des exercices plus complexes (module « Haut ») mobilisant des textes de structure et de

nature différentes pour évaluer des compétences élaborées, comme la capacité à produire des inférences. Si les résultats au module d'orientation sont faibles, il passe un test assez simple, le module « ANLCI » (Agence nationale de lutte contre l'illettrisme), pour affiner le diagnostic sur ses difficultés qui, outre de nouvelles questions en lecture de mots et en compréhension, porte sur la capacité à produire des mots écrits. Si l'enquêté obtient des résultats moyens au test d'orientation, il passe une épreuve « intermédiaire » pour déterminer laquelle des deux voies précédentes est la plus adaptée. A l'issue de ces parcours différenciés, les enquêtés répondent à un exercice commun. Le questionnaire inclut également un exercice de compréhension orale, situé après le module d'orientation, et de courts problèmes de calcul posés oralement à la suite du module commun. Il se termine par des questions biographiques qui recueillent des informations sur l'enfance, la scolarité et la vie professionnelle de l'enquêté.

L'enquête PIAAC est une enquête internationale coordonnée par l'Organisation pour la Coopération et le Développement Économique (OCDE). Elle a été menée conjointement dans 24 pays. En France, l'enquête a été conduite en 2012 auprès de 7 000 personnes âgées de 16 à 65 ans. L'interrogation commence par le questionnaire biographique qui détaille notamment les compétences mobilisées au travail. Puis les exercices sont proposés aux enquêtés. Cette partie débute par une évaluation (déclarative et en pratique) de la capacité des enquêtés à utiliser un ordinateur pour répondre aux exercices. Les personnes jugées suffisamment compétentes (environ

Le calcul d'un score de compétence

Dans PIAAC, les scores sont estimés par une méthode statistique appelée « Modèle de réponse à l'item à deux paramètres » (MRI, *sources et méthodes*). Cette méthode, qui fixe la moyenne des compétences à 250 et leur écart-type à 50, attribue, à chaque personne, un score (exprimé en points) dans les domaines évalués permettant de classer les individus sur une seule échelle de performance. Les enquêtés sont ensuite répartis dans 6 groupes dans le domaine de l'écrit, et autant dans le domaine des chiffres, renvoyant, selon l'OCDE, à des tâches-types qui peuvent être remplies avec succès pour chacun de ces 6 niveaux de compétence (Jonas, 2013).

Dans IVQ, les groupes de compétence sont élaborés selon le pourcentage d'exercices réussis. Le module ANLCI répartit les personnes selon leur degré de difficulté à l'écrit, définit en fonction de leur performance minimale dans les trois domaines fondamentaux mesurés par ce module. C'est ainsi qu'on identifie les 3 groupes présentant des difficultés graves, fortes ou partielles. Pour ceux n'ayant pas de difficultés, on retient le pourcentage de réussite aux

Les profils des individus les mieux et les moins bien classés sont similaires dans les deux dispositifs

Les comparaisons des adultes appartenant aux 1^{er} et 10^e déciles de la distribution des scores en numératie de PIAAC d'une part et d'IVQ d'autre part dessinent des profils de population très semblables. Parmi les 16-65 ans, les femmes sont moins souvent à l'aise que les hommes. Selon les deux enquêtes, 11,1 % d'entre elles appartiennent au 1^{er} décile contre un peu moins de 9 % des hommes. À l'autre extrême de la distribution, les ordres de grandeur sont inversés. Un homme sur huit appartient au groupe le mieux classé (12,6 % pour PIAAC et 12,8 % pour IVQ) contre moins d'une femme sur treize (7,6 % et 7,4 %). Dans le domaine de l'écrit, selon les deux enquêtes, les hommes sont proportionnellement plus souvent que les femmes dans le dernier décile de la

90 % des répondants) passent ensuite à des exercices interactifs à résoudre sur un ordinateur dont la difficulté est adaptée automatiquement selon les échecs/réussites de l'enquêté, mais aussi selon son niveau d'études et la langue apprise pendant l'enfance. Ces exercices ont été développés pour intégrer les fonctionnalités d'un ordinateur (utilisation d'une souris, surlignage, etc.) et de l'environnement informatique habituel (navigation sur Internet, utilisation de liens hypertextes, etc.). Les personnes moins à l'aise en informatique répondent sur un support plus traditionnel, un cahier d'exercices qu'elles doivent remplir seules puis rendre à l'enquêteur. Que ce soit sur support numérisé ou sur support papier, l'évaluation commence par un court module d'orientation. Si l'enquêté ne parvient pas à répondre correctement à la moitié de ces exercices, l'interrogation s'interrompt.

L'architecture et le protocole de ces enquêtes diffèrent sur deux points principaux. En premier lieu, l'enquête PIAAC se construit surtout autour d'exercices interactifs sur ordinateur. Les enquêtés doivent résoudre les exercices, mais aussi être capables de restituer leurs réponses sur support informatique dans un format imposé. Or, il est possible que, entre deux personnes possédant les mêmes capacités, celle ayant plus de familiarité avec l'outil informatique obtienne de meilleures performances. Dans IVQ, au contraire, c'est l'enquêteur qui saisit les réponses sous la dictée de l'enquêté. En second lieu, la durée de l'enquête est sensiblement plus courte dans IVQ que dans PIAAC pour maintenir la motivation des enquêtés et éviter les abandons et les non réponses partielles qui peuvent fausser l'interprétation des résultats.

questions du module Haut pour distinguer 4 groupes supplémentaires de compétences. Comme l'orientation est précoce dans IVQ, il y a un saut qualitatif entre les scores des personnes passées par le module ANLCI et par le module Haut. Cependant, grâce à des innovations introduites en 2011, en particulier la création d'un module commun (*sources et méthodes*), il est possible d'appliquer aux résultats « bruts » d'IVQ une méthode d'estimation comparable à celle utilisée dans PIAAC. On a ainsi estimé un score continu de littératie (en utilisant les questions de compréhension du module d'orientation et du module ANLCI, et l'ensemble des questions du module Haut et du module commun) et de numératie (en utilisant le seul module consacré à ce domaine).

Partant de ces scores continus, une approche comparative des caractéristiques des personnes présentant des scores identiques dans les deux enquêtes est envisageable. Nous nous limiterons ici à la comparaison des 10 % d'adultes aux deux extrêmes de la distribution des scores.

distribution, mais aussi dans le 1^{er}. L'écart entre hommes et femmes est cependant plus accentué dans IVQ (environ 1,5 point) que dans PIAAC, où il est quasiment nul dans le 1^{er} décile.

En littératie, les caractéristiques scolaires des individus les plus performants et les moins performants sont très comparables entre les deux enquêtes. Par exemple, 23,8 % des personnes pas ou peu diplômées se situent dans le 1^{er} décile d'après PIAAC et 23,9 % d'après IVQ ; les titulaires du baccalauréat sont 3,0 % à être dans le même cas d'après PIAAC et 3,1 % d'après IVQ. En numératie, on observe le même gradient en fonction du diplôme, mais celui-ci est plus accentué dans PIAAC que

dans IVQ. Pour le 1^{er} décile, par exemple, la différence entre le pourcentage de personnes peu ou pas diplômées et de personnes diplômées de l'enseignement supérieur est de 24,2 points dans PIAAC alors qu'elle est de 17,7 points dans IVQ. Cet écart peut s'expliquer par le fait que le questionnaire de PIAAC est plus long et propose des exercices plus compliqués que celui d'IVQ.

Les caractéristiques des deux déciles extrêmes en termes d'âge sont également très proches entre les deux enquêtes. En calcul, la seule différence notable concerne la tranche d'âge la plus élevée. Les personnes âgées de 55 à 65 ans sont plus nombreuses dans le groupe des adultes les moins performants dans PIAAC que dans IVQ

Que mesure-t-on finalement ?

Historiquement, en France, l'émergence de la question des compétences des adultes, à travers celle de la mesure de l'illettrisme, est liée aux thématiques de l'exclusion sociale et de la pauvreté. L'accent a été mis sur l'identification des niveaux les plus bas de maîtrise de l'écrit. L'enquête IVQ a donc été notamment adaptée pour décrire la nature des difficultés de ces niveaux grâce au module ANLCI qui adopte une méthode multidimensionnelle. Dans PIAAC, la littératie et la numératie ne concernent pas seulement les capacités de compréhension, mais plus globalement les compétences fondamentales en traitement de l'information sur support informatique, dont la maîtrise permettrait aux individus d'atteindre des objectifs personnels et de développer des connaissances et des aptitudes professionnelles. L'accent est donc mis sur la maîtrise de processus spécifiques, transposables lors de l'usage d'un ordinateur.

Cette différence d'approche se traduit par une méthode distincte de constitution des groupes de compétence. L'utilisation d'une méthode MRI traite les compétences d'un point de vue unidimensionnel. Un score unique (trait latent) permet d'estimer le niveau de compétence. Les groupes sont ensuite constitués grâce à des seuils appliqués à la distribution des scores. Dans IVQ, au contraire, pour les niveaux les plus bas de compétence, la maîtrise de l'écrit est traitée de façon multidimensionnelle et ne se résume pas à un score unique. Trois domaines fondamentaux de l'écrit sont combinés : la production, l'identification et la compréhension de mots. On considère qu'une personne qui obtient des performances moyennes dans chacun de ces trois domaines a une communication à l'écrit plus efficace que celle qui obtient de bonnes performances dans deux domaines et des performances médiocres dans le troisième.

On peut se rendre compte des conséquences de ces logiques différentes de constitution des groupes en mettant en regard, sur les données d'IVQ, les groupes construits par la méthode multidimensionnelle et les groupes construits grâce aux scores MRI. On a fixé des seuils permettant de retrouver des groupes de compétence de même importance, soit 11 % de personnes en difficulté graves ou fortes. Il existe une divergence d'appréciation non négligeable entre les deux estimations. 9,6 % de la population seraient classés en situation préoccupante à la fois avec les scores MRI et avec l'approche multidimensionnelle d'IVQ. Par contre, 1,5 % des personnes seraient classées dans cette situation

(16,5 % contre 14 %). Cette différence est en partie attribuable à « l'effet support » induit par l'usage de l'ordinateur. La nécessité d'utiliser correctement les fonctionnalités informatiques pour répondre aux exercices crée un avantage comparatif pour les générations les plus jeunes. Cet effet est moins marqué en littératie. Les adultes âgés de 55 à 65 ans, sont 1,3 % de plus au sein du 1^{er} décile dans PIAAC que dans IVQ, et environ 1 % de moins au sein du 10^e décile.

Les extrêmes de la distribution des scores présentent donc des profils de population très comparables entre les deux enquêtes, surtout en littératie.

uniquement par l'approche unidimensionnelle et 1,4 % uniquement par l'approche multidimensionnelle. Cette divergence d'estimation est loin d'être négligeable si on considère qu'une personne sur huit en difficulté à l'écrit selon la méthode multidimensionnelle n'est pas reconnue comme telle avec la méthode MRI. Un score continu ne donne finalement pas les mêmes résultats qu'un classement établi à partir de l'approche multidimensionnelle qui est enrichie par la prise en compte de deux autres dimensions fondamentales de la maîtrise de l'écrit : la production de mots et l'identification de mots.

Par ailleurs, la part des personnes appartenant aux niveaux les plus élevés de compétence est très faible dans le dispositif PIAAC. Seuls 7,7 % de la population française en littératie et 8,3 % en numératie possèdent les compétences nécessaires pour être classés aux niveaux 4 ou 5, des proportions nettement plus faibles que celles obtenues dans IVQ. Ces écarts posent des questions sur la différence de nature des compétences évaluées mais aussi sur celle des protocoles mis en œuvre. Par exemple, les exercices les plus difficiles de PIAAC, permettant de confirmer la maîtrise des plus hauts niveaux de compétence, apparaissent en fin de questionnaire, alors que l'interrogation a déjà été longue. L'ampleur de ces différences et leurs conséquences possibles sur la mesure devront faire l'objet d'études plus approfondies.

Références

Jonas N., « [Pour les générations les plus récentes, les difficultés des adultes diminuent à l'écrit mais augmentent en calcul](#) », *Insee Première* n°1426, décembre 2012.

Jonas N., « [Les capacités des adultes à maîtriser les informations écrites ou chiffrées](#) », *Insee Première* n° 1467, octobre 2013.

Jonas N. et Lebrère A., « Le pilote PIAAC et les enquêtes internationales de mesure des compétences des adultes. Analyse des données françaises », *Document de travail DSDS*, Insee, 2013, à paraître.

Murat F. et Rocher T., « [Création d'un score global dans le cadre d'une épreuve adaptative](#) », *Économie et Statistique*, n°424-425, pp. 149-178, février 2010.

OCDE, *Technical Report of the Survey of Adult Skills (PIAAC)*, Éditions OCDE, 2013.

Ce document est un complément technique aux deux numéros d'*Insee Première* consacrés respectivement aux enquêtes IVQ et PIAAC, les numéros 1426 de décembre 2012 et 1467 d'octobre 2013. Des éléments techniques complémentaires sont fournis dans Jonas et Lebrère (2013, à paraître).

Sources et méthodes

Définitions

La **littératie** se définit comme la capacité à comprendre et à utiliser l'information contenue dans des textes écrits dans divers contextes quotidiens pour atteindre des objectifs personnels et pour développer des connaissances et des aptitudes. La **numératie** se définit comme la capacité à utiliser, appliquer, interpréter et communiquer des informations et des idées mathématiques.

Les modèles de réponse à l'item

Les modèles de réponse à l'item (MRI) sont des modèles logistiques qui expliquent la probabilité de réussite d'un individu à un exercice (item) par le niveau de compétence de l'individu et par les caractéristiques propres de l'item. Plus précisément, les individus sont caractérisés par leur niveau de compétence, qui correspond à leur position sur la dimension latente (ou trait latent) θ . Les items, quant à eux, sont caractérisés par leur difficulté (b), c'est-à-dire leur position sur cette même dimension latente, et leur pouvoir de discrimination (a), appelé aussi « pente » de l'item. Lorsque la discrimination est élevée, une légère variation du niveau de compétence de l'enquêté autour du niveau de difficulté de l'item entraîne une variation importante de la probabilité de répondre correctement. L'avantage de ces modèles est de séparer les concepts : l'aptitude d'un individu est définie indépendamment de la difficulté de l'épreuve et, inversement, la difficulté des items n'est pas fonction du niveau de compétence des individus. Plusieurs éléments limitent cependant la portée de ces modèles :

- a_i , b_j et θ ne sont pas uniques. Pour résoudre l'estimation il faut imposer des hypothèses sur la distribution des compétences. Ainsi, dans PIAAC, la moyenne des θ a été fixée à 250 et l'écart-type à 50, comme lors de l'enquête IALS (*International Adult Literacy Survey*).
- dans le contexte d'une évaluation internationale, il faut que les caractéristiques de chaque item soient stables entre les pays.
- ce type de modèle suppose que le trait modélisé (la compétence) est unidimensionnel.

Création d'un score global dans l'enquête IVQ

L'enquête PIAAC est conçue dans l'idée que les estimations de compétence sont comparables entre l'ensemble des pays participants à l'enquête. Le nombre important d'exercices proposés, la variabilité de leur niveau de difficulté et la logique de leur allocation autorisent en effet le recours à une méthode MRI, dès lors que, suite à une période de tests, les résultats ont montré une certaine stabilité des paramètres de la plupart des items au sein de chaque pays. Les estimations obtenues classent ainsi les individus sur une même échelle internationale de compétence. Pour comparer les propriétés des mesures réalisées grâce au dispositif PIAAC avec celles réalisées grâce au dispositif IVQ, il était nécessaire de construire également un score continu à partir des données d'IVQ.

Cependant, à l'origine, l'enquête IVQ n'a pas été élaborée dans cet objectif et l'architecture de son questionnaire n'entre pas dans ce cadre idéal de test à « cahiers tournants » où l'allocation des cahiers doit être en partie aléatoire. L'attribution du module Haut ou du module ANLCI, dépend en effet des résultats au module d'orientation. Ce processus d'orientation précoce, qui distingue nettement deux sous-populations, risque de provoquer des biais dans l'estimation qui va tenter l'ancrage entre les différents modules. Par ailleurs, le module ANLCI comprend trois domaines d'évaluation afin de mieux qualifier la nature des difficultés rencontrées par les personnes qui possèdent les plus bas niveaux de compétence en traitement de l'écrit, alors que le module Haut se concentre sur un seul domaine. Enfin, les exercices du module Haut, vers lesquels plus de 80 % des répondants sont orientés, n'étaient pas suffisamment nombreux, lors de la précédente campagne d'IVQ, pour distinguer correctement les performances des personnes réussissant le module d'orientation.

Deux évolutions de l'architecture de l'enquête IVQ en 2011 ont facilité la construction d'un score global. Le module commun, composé de 7 questions de difficulté variable, augmente le nombre d'exercices communs passés par l'ensemble de la population. Le module Haut B, cahier alternatif au module Haut d'IVQ 2004 (rebaptisé Haut A), diversifie les parcours des personnes ayant réussi le module d'orientation. Grâce à ces innovations, une estimation concurrente des items a été mise en œuvre selon l'approche définie par Murat et Rocher (2010) dans le cas d'une épreuve sur plusieurs groupes non-équivalents avec des items communs. Comme les performances aux exercices de production de mots (module ANLCI) et d'identification de mots (modules d'orientation et ANLCI) sont peu corrélées au score global, seules les questions portant sur la compréhension de l'écrit ont été retenues. Cette restriction de l'estimation à un seul domaine de l'évaluation, présent dans tous les modules, permet par ailleurs de se rapprocher au plus près de la définition de la littératie utilisée dans PIAAC.

Contexte d'une évaluation internationale

Dans le cadre de PIAAC, malgré les nombreux efforts réalisés depuis l'expérience de IALS pour assurer une qualité homogène des résultats (Jonas et Lebrère, 2013), il peut rester certains écarts entre les pays qui, sans remettre en cause la validité des données, nécessitent la poursuite d'investigations sur la portée et les limites d'une telle entreprise. En premier lieu, proposer à des adultes de répondre, de façon autonome, à un ensemble relativement long de questions scénarisées est une pratique qui peut être plus ou moins adaptée selon les contextes nationaux. Par exemple, les adultes qui auront eu, dès l'enfance, l'habitude d'être évalués sur des sujets qui ne renvoient pas à un programme scolaire (tests de QI, bilans de compétence, tests d'aptitude, etc.) trouveront *a priori* plus naturel ce type de pratique et y seront plus habitués que des adultes qui le découvrent au moment de l'enquête. De la même façon, les pays qui ont fréquemment recours à des évaluations sous forme de QCM auront une population potentiellement plus performante à ce genre de tests que des pays qui privilégient d'autres types d'évaluation (argumentations, commentaires). En second lieu, pour ce genre d'enquête, le protocole de collecte joue directement sur la qualité des données obtenues. Dans l'enquête PIAAC, malgré les efforts d'homogénéisation, des écarts de pratique sont encore notables. Par exemple, certains pays versent une compensation financière aux répondants, alors que d'autres, comme la France, ne le font pas. En troisième lieu, il est nécessaire de s'assurer que les questions posées possèdent les mêmes caractéristiques dans chaque pays. Cette condition est indispensable à la validité théorique des MRI. Si, par exemple, un item A est plus difficile qu'un item B dans la majorité des pays, le pays pour lequel cette hiérarchie sera inversée obtiendra une mauvaise estimation du niveau de compétence de sa population. Pour éviter cet écueil, l'OCDE a sélectionné, à partir d'une enquête pilote, les exercices dont les paramètres étaient les plus stables. Des travaux d'expertise complémentaire pourront être menés sur la robustesse de ces résultats ou la sensibilité des scores à certaines questions.