

Évaluation de la compréhension de l'écrit chez l'adulte

Hakima Megherbi*, **Thierry Rocher****, **Valérie Gyselinck*****,
Bruno Trosseille** et **Hubert Tardieu*****

Mesurer la compréhension de l'écrit nécessite la définition d'un certain nombre de concepts que nous situons dans un cadre théorique provenant des travaux de la psychologie du langage. Comprendre un texte ou un document est une activité complexe dont le but ultime est l'élaboration d'une représentation mentale cohérente du contenu (Gernsbacher, 1994, Kintsch, 1994 et 1998). Au-delà de la lecture des mots qui composent le document, la compréhension sollicite des processus de haut niveau qui assurent l'analyse syntaxico-sémantique de la phrase et la prise en charge de l'ensemble du document. La cohérence est recherchée par le lecteur pour des documents composés d'informations exclusivement verbales, et pour des documents associant des informations verbales et visuo-spatiales.

L'analyse des dimensions sous-jacentes du module Haut de l'enquête *IVQ* 2004, à partir des données recueillies sur 7 389 individus qui y ont participé, si elle doit être interprétée avec beaucoup de prudence en raison du faible nombre d'items composant le module, distinguerait en fin de compte deux seuls facteurs : l'un visuel et spatial (cartes spatiales et graphique), l'autre strictement verbal. La pertinence de ces deux dimensions serait en faveur d'une spécialisation du traitement des informations selon leur nature, verbale ou visuo-spatiale.

La spécialisation paraît confirmée par l'étude des fonctionnements différentiels liés au sexe. Un résultat classique qui se dégage est le « biais » - défini par la réussite de chaque item selon le sexe à niveau de performance égale entre les deux groupes - systématique sur les épreuves visuo-spatiales en faveur des hommes (plan du village et graphique). Nous observons des biais en faveur des femmes sur quelques items qui sollicitent des informations de nature verbale, mais ils sont peu nombreux.

* Université Paris 13, UTRPP, EA 3413, Villetaneuse

** Ministère de l'Éducation Nationale, DEPP, Paris

*** Université René Descartes, UMR 8189 CNRS, Boulogne-Billancourt

Nous remercions les deux rapporteurs anonymes pour leurs remarques qui ont permis d'améliorer notre article. Nos remerciements s'adressent également à Marie-France Ehrlich, Directeur d'Études Honoraire EPHE-Université Paris 5, et à Daniel Verger, responsable de l'Unité Méthodes statistiques à l'Insee, pour leurs précieux conseils.

L'élaboration des épreuves du module Haut a bénéficié de deux contrats financés par la Direction de l'Évaluation et de la Prospective et de la Performance, Ministère de l'Éducation Nationale, France.

La compréhension de documents écrits est une activité individuelle inscrite dans un contexte social. Elle met en jeu des processus psycholinguistiques, cognitifs, motivationnels dont la réalisation dépend de facteurs affectifs, sociologiques et culturels. La maîtrise de la compréhension est donc sous la dépendance de facteurs complexes et multidimensionnels qui interagissent, ce qui rend difficile l'étude de la contribution respective de chacun d'entre eux.

L'outil d'évaluation de la compréhension de l'écrit qui est l'objet de notre étude a été baptisé « module Haut », parce qu'il est destiné à toucher une population n'ayant pas de difficultés dans la maîtrise des mécanismes de base de la lecture. Rappelons que le module d'orientation (Degorre et Murat, ce numéro) opère une sélection parmi les répondants à l'enquête sur la base d'une épreuve de lecture de mots et de réponses à des questions de compréhension. Le support utilisé est une présentation visuelle d'un programme télévisé. Les questions sont relativement simples puisqu'elles font référence au traitement d'informations explicites présentes dans le document. Les participants ayant des difficultés sont orientés vers le module ANLCI (Agence nationale de lutte contre l'illettrisme).

Le module Haut a été conçu de sorte à mesurer le niveau de compréhension à l'aide de documents de différente nature que la population rencontre dans son quotidien. Cinq épreuves suffisamment attrayantes ont été proposées : trois textes, un graphique et un plan de village.

Ces documents font appel à des processus de différente nature : les trois textes ne comportent donc que des informations de nature verbale, les deux autres documents associent des informations visuo-spatiales et des informations verbales. Les trois textes sont de différente longueur et appartiennent à trois types distincts : un récit, un texte expositif-explicatif et un texte théorique scientifique. Les deux autres documents sont un graphique représentant des données numériques associées à une légende, et enfin un plan d'un village sur lequel apparaissent différentes indications de nature verbale.

Trois types d'analyses ont été effectués à partir des premiers résultats recueillis sur plus de 7 000 personnes âgées de 18 à 65 ans : analyse descriptive des performances observées pour les différentes épreuves (taux de réussite et pouvoir discriminant), analyse de la structure des réponses afin d'identifier les dimensions cognitives sous-jacentes aux performances (analyse

factorielle) et analyse des fonctionnements différents en fonction du sexe des participants.

Un cadre théorique provenant des travaux de la psychologie cognitive du langage

En dépit d'une apparence de grande simplicité, la compréhension d'un texte est une activité complexe qui met en jeu un ensemble de processus de différente nature, dont le but ultime est la construction d'une représentation mentale cohérente. Face à un document écrit, le lecteur doit extraire du sens, et pour y parvenir, il s'engage dans une activité de construction de représentations. Comme le rappelle Le Ny (2005, p. 104) dans son dernier ouvrage, « *La compréhension d'un énoncé, c'est-à-dire son traitement cognitif dans l'esprit/cerveau d'un compreneur, y produit causalement, d'abord une suite de sous-processus et d'états mentaux transitoires, qui aboutissent finalement à la construction d'une représentation sémantique terminale, mentale, qui est composée et structurée : c'est cette représentation qui constitue le sens de l'énoncé* ». La compréhension est donc une activité de « construction de sens ».

Pour parvenir cette construction, le lecteur doit lire les mots composant les phrases et le texte, et il doit en comprendre l'ensemble. La finalité de l'acte de lire est bien de comprendre le texte. Il existe un consensus entre les auteurs pour accorder un rôle important à l'identification des mots dans l'activité de compréhension de l'écrit. Sans maîtrise des mécanismes de base de la lecture, le sujet ne peut parvenir à comprendre, tout au moins dans une situation de langage écrit. Le débat se situe du côté de la « relative » indépendance entre la lecture et la compréhension. Alors que certains postulent une intrication entre l'identification des mots et la compréhension (Perfetti et Hart, 2001 ; Perfetti, 2007), d'autres défendent une relative indépendance entre les deux activités (Yuill et Oakhill, 1991). Il est nécessaire de faire la part des choses entre lecture et compréhension (cf. encadré 1). Puisque les participants du module Haut ont réussi l'épreuve d'orientation comportant le déchiffrement de mots, nous nous sommes centrés sur la caractérisation des processus de compréhension pour concevoir ce module.

Ainsi, la compréhension d'un texte écrit est étroitement dépendante de la lecture des mots qui le composent et de l'accès à leur signification. Au-delà du mot, la compréhension met en jeu des processus de « haut niveau » qui assu-

LECTURE ET COMPRÉHENSION

Pour Perfetti (1994), la « lecture » est un ensemble de processus permettant d'extraire la signification du texte qui comprend les processus d'identification des mots écrits et ceux alloués à la compréhension. Ainsi, ne distingue-t-il pas vraiment les processus de lecture et les processus de plus haut niveau engagés dans la compréhension. Il réaffirme les grandes lignes de sa position en soulignant le lien causal entre lecture de mots et compréhension : « *Les compétences lexicales permettent la compréhension, la compréhension permet la pratique de la lecture, la pratique de la lecture renforce les compétences lexicales* » (Perfetti et Hart, 2001, p. 67 ; cf. aussi Perfetti, 2007). Ainsi, les « bons compreneurs » à la différence des « faibles compreneurs » disposent de représentations lexicales - phonologiques, orthographiques et sémantiques - de haute qualité qui leur permettent de former des représentations nouvelles pour des mots nouveaux par exemple. Plus récemment, avec des techniques électrophysiologiques, Perfetti va plus loin en montrant que le traitement des mots est plus lent chez les faibles compreneurs comparés aux bons, les différences entre les deux groupes s'observant dès les premières millisecondes après le début du mot (Landi et Perfetti, 2007).

Une deuxième approche provenant des travaux réalisés essentiellement chez l'enfant stipule que la maîtrise des mécanismes de base de la lecture ne garantit pas, à elle seule, un bon niveau de compréhension (e.g. : Cain et Oakhill, 2006 ; Yuill et Oakhill, 1991) : certains ont un faible niveau de compréhension alors que leurs performances en lecture sont normales. Pour ces auteurs, ce sont les processus de haut niveau qui sont défectueux : traitement des marques linguistiques assurant la cohérence textuelle, mise en jeu des inférences, traitement des métaphores.

Ce qui apparaît certain est que la finalité de l'acte de lire est bien de comprendre le texte. La maîtrise des mécanismes de base de la lecture apparaît, de manière consensuelle, indispensable à la réalisation de la finalité de cette activité complexe. Alors que Perfetti stipule une intrication entre lecture et compréhension, Oakhill et ses collègues conçoivent une relative indépendance entre lecture et compréhension. De part leur position respective, les faibles et les bons compreneurs comparés dans leurs études ne présentent pas tout à fait les mêmes caractéristiques : alors que Perfetti contraste les deux groupes uniquement sur une épreuve de compréhension globale, Oakhill et ses collègues contrôlent, en plus, les niveaux de vocabulaire et de déchiffrement, de sorte que les deux groupes sont contrastés spécifiquement sur des épreuves de compréhension de textes.

La relative indépendance entre lecture et compréhension est soulignée dans le modèle « simple » de Gough et Tunmer, 1986 (cf. aussi Hoover et Gough, 1990) proposé pour des langues ayant un système d'écriture alphabétique. « *The Simple View* » est un

modèle multiplicatif qui conçoit que la compréhension de l'écrit résulte du produit de deux composantes, l'identification des mots et la compréhension du langage (oral). La première est une activité spécifique de l'écrit, tandis que la compréhension du langage ne l'est pas, puisque les processus de haut niveau - ceux alloués à l'analyse syntaxico-sémantique et ceux assurant l'organisation textuelle - sont dans une certaine mesure communs à l'écrit et à l'oral. La relation multiplicative entre les deux stipule que si l'une ou l'autre des composantes - ou *a fortiori* les deux - n'est pas maîtrisée, le niveau de compréhension à l'écrit sera affecté. Soulignons quand même que les études qui apportent une certaine validation empirique à ce modèle sont issues des travaux portant chez des sujets de langue anglaise, pour la plupart des enfants. La trajectoire développementale prédit que lorsque l'enfant maîtrise les mécanismes de base de la lecture, c'est la compréhension de l'oral qui contribue le plus à expliquer le niveau de compréhension à l'écrit (Gough *et al.*, 1996). Notons cependant qu'une étude réalisée chez de jeunes enfants de langue française montre que, dès le cours préparatoire (6 ans 8 mois), le décodage mais également la compréhension de l'oral sont deux prédicteurs spécifiques de la compréhension de l'écrit, la part de variance spécifique étant même plus forte pour la compréhension de l'oral (Megherbi *et al.*, 2006). Ce résultat a été interprété en termes de degré de transparence des règles graphèmes-phonèmes qui n'est pas le même en français et en anglais, l'enfant anglophone ayant besoin de plus de temps pour apprendre ces règles en raison d'une faible transparence. Ainsi, en français, la compréhension de l'oral serait un prédicteur spécifique dès la première année de primaire, au moins aussi puissant que le décodage.

Dans la perspective du modèle de Gough, la corrélation entre compréhension de l'écrit et compréhension de l'oral devrait être très forte chez l'adulte. C'est ce que Gernsbacher *et al.* (1990) observent chez de jeunes étudiants anglophones (corrélation de 0.92). Notons que la corrélation entre la compréhension de l'écrit et la compréhension des images est assez forte également (elle est de 0.80). Les auteurs concluent que la compréhension chez l'adulte est une activité complexe dépendante d'un système *général* impliqué dans le traitement des informations écrites, orales et imagées (Gernsbacher, 1990 ; Graesser *et al.*, 2002 ; Traxler et Gernsbacher, 2006). Soulignons que ces résultats ne sont pas observés avec une population tout-venant et que les textes à l'écrit et à l'oral étaient de même nature (des récits), ce qui signifie que la structure des textes ne tenait pas compte de la spécificité des formes de langage, écrit et oral (e.g. les textes de type conversationnel pour l'oral n'ont pas été considérés). Enfin, les auteurs n'ont pas évalué l'identification des mots isolés, ni la lecture de textes.

rent l'analyse syntaxico-sémantique des propositions et des phrases, et la mise en place de la double cohérence, locale et globale, afin d'élaborer une représentation mentale de la signification du contenu du texte (Ehrlich, 1994 ; Kintsch, 1998 ; Gernsbacher et Foertsch, 1999 ; Perfetti, 1994 ; Kintsch, 1998). La réalisation de ces processus (cf. encadré 2) qui s'appuie sur une coordination complexe de traitements et de stockages d'informations est contrainte par l'architecture du système cognitif et notamment par les capacités limitées de la mémoire de travail (système décrit quelques lignes ci-dessous) (Caplan et Waters, 1999 ; Carpenter *et al.*, 1994 ; Gathercole et Baddeley, 1993).

Comme le précise le modèle de Kintsch (1988 ; 1998), la représentation mentale mise en place au cours de la lecture s'établit via des processus de construction-intégration. Pour aboutir à une représentation cohérente, le compreneur extrait et traite les informations linguistiques contenues dans le document, et il mobilise ses connaissances (du monde) pour enrichir la représentation. Ces connaissances sont en fait des représentations cognitives générales propres à chacun. Le Ny (2005) fait une distinction entre deux sources d'informations nécessaires : la première dite *externe* renvoie à l'information linguistique composant l'énoncé, et l'information *interne* concerne toutes sortes de connais-

Encadré 2

PROCESSUS IMPLIQUÉS LORS DE LA COMPRÉHENSION ET REPRÉSENTATIONS COGNITIVES

Les différents niveaux de représentation : au niveau du mot, de la phrase, du texte

Les processus sollicités durant l'activité de compréhension de documents écrits prennent en compte les différentes dimensions du langage : lexicale, syntaxique, sémantique et pragmatique. L'activation des processus d'identification et de reconnaissance des mots consistent en l'activation des représentations orthographiques (orthographe du mot), phonologiques (sa prononciation) et sémantiques (sa signification) : le lecteur déchiffre les mots et a accès à leur signification. L'étendue du vocabulaire ici joue un rôle déterminant. Le vocabulaire est généralement évalué par des épreuves spécifiques où l'on demande aux sujets de travailler sur la synonymie, sur la définition de mots, etc. Dans notre module, nous avons des questions qui ciblent particulièrement la connaissance du vocabulaire (par exemple, *Dans la phrase 1, quel mot parmi les quatre proposés pourrait remplacer le mot « chute »*). Dans certains cas, le participant peut y répondre sur la seule analyse du mot, tandis que dans d'autres, il doit intégrer le mot dans la phrase.

Au-delà de ce niveau, le compreneur doit intégrer ces mots dans des groupes syntaxiques et les rattacher à des unités de rang supérieur jusqu'à la phrase où ils sont interprétés syntaxiquement, sémantiquement et fonctionnellement. Ces analyses traduisent la signification de l'ensemble de la phrase (« Qui fait quoi ? » ; « Où ? » ; « Quand ? » ; « Comment ? »). Dans notre module nous avons des questions qui ciblent la compréhension de tournures syntaxiques complexes comme les phrases négatives ou les phrases passives (qui sont syntaxiquement plus complexes que les phrases affirmatives). Le compreneur s'engage ensuite dans des calculs de liens sémantiques et pragmatiques que cette phrase peut entretenir avec celles déjà traitées ou celles qui sont à venir. À ce niveau, il s'agit d'organiser les phrases de sorte qu'elles forment un texte. Il n'existe pas vraiment de règles formelles (comme pour

la syntaxe d'une phrase) qui déterminent précisément comment le texte doit être construit. La continuité entre les phrases est assurée par des marques linguistiques spécifiques telles que les connecteurs (e.g. *Paul est gentil avec Jacques. Mais le problème est que ...*), ou encore les anaphores. Une anaphore est une expression qui renvoie à un référent précédemment cité dans le texte ; c'est le cas des pronoms (e.g. *Paul est gentil avec Jacques. Il veut le protéger*). Les pronoms sujet *Il* et objet *le* réfèrent respectivement à *Paul* et à *Jacques*). La cohésion textuelle est également assurée par les marques de temps qui décrivent la séquentialité des événements (e.g., marques aspectuelles sur les verbes, adverbies), les marques de cohésion spatiale qui situent spatialement au fur et à mesure le récit (e.g. « *Sous le chapiteau ...* ») etc. Il est crucial pour un lecteur de faire des connexions entre les phrases. L'épreuve 1 du module a pour objectif précis d'inviter le participant à analyser les marques de cohésion de sorte à remettre dans l'ordre une suite de sept phrases présentées dans le désordre. Pour réaliser cette épreuve, l'enquêté doit chercher la cohérence entre les phrases, par le biais notamment, de l'analyse de la cohésion textuelle (cf. l'exemple fictif en encadré 3). Parfois, la présence de marques de cohésion est soit inexistante, soit insuffisante pour arriver à construire une représentation cohérente. Le compreneur met en jeu d'autres opérations comme la mise en jeu d'inférences causales ou de liaison. La cohésion participe de l'élaboration de la cohérence mais cette dernière peut être établie en sollicitant d'autres informations extralinguistiques comme la mise en jeu d'inférences.

Sur le plan de la cohérence globale du document, plusieurs processus sont mis en jeu. En particulier, la sélection des informations en fonction de leur importance eu égard au thème du texte contribue à appréhender son organisation globale. En fonction des connaissances, des croyances et des objectifs du compreneur, celui-ci élabore une représentation plus ou moins riche de la situation décrite dans le document. →

sances aussi bien celles concernant le monde, que les connaissances lexicales, grammaticales, sémantiques qui sont en mémoire dans l'esprit du comprendreur.

Kintsch propose un modèle de fonctionnement des processus qui permettent au lecteur de construire les différentes représentations. La représentation la plus riche intègre les connaissances et les croyances du lecteur et fait largement appel à la mise en jeu d'opérations complexes telles que les inférences (pour plus de détails, cf. encadré 2). Concernant cette représentation, elle est mise en œuvre avec plusieurs types de

textes. Ainsi, elle peut être élaborée avec des documents qui décrivent des unités concrètes du monde physique, que celles-ci soient réelles ou imaginaires, avec des textes abstraits, scientifiques et/ou théoriques (modèle de situation conceptuel), avec des documents décrivant des unités associant des informations visuo-spatiales et des informations verbales (modèle de situation visuo-spatial) etc. Des représentations linguistiques et les représentations du modèle de situation sont donc mobilisées lors de la compréhension de textes narratifs, expositifs, explicatifs, théoriques, descriptifs, de modes d'emploi, de graphiques, de plans etc. (Ehrlich *et al.*, 1993).

Encadré 2 (suite)

Représentation de la situation décrite dans le document : le « modèle de situation »

Kintsch propose un modèle de fonctionnement des processus qui permettent au lecteur de construire les différentes représentations (Kintsch et van Dijk, 1978 ; Kintsch, 1988, 1994 et 1998 ; en français on peut consulter l'ouvrage de Ehrlich *et al.*, 1993). L'auteur définit trois niveaux de représentation : le niveau de surface, la base de texte propositionnelle et le modèle de situation. Le niveau le plus superficiel est l'analyse des caractéristiques de surface qui consiste en une sorte de traduction strictement linguistique et littérale des mots et des phrases. La base propositionnelle, qui à partir de la représentation de traitement de surface, est aboutie grâce aux traitements sémantiques et conceptuels que le lecteur engage. Ainsi au cours de la lecture d'un texte, le lecteur construit une liste hiérarchisée de propositions, avec une proposition importante qui situe la thématique du contenu du texte, et des propositions subordonnées qui sont également reliées les unes aux autres. Kintsch définit la macrostructure et la microstructure. La première renvoie à la proposition thématique qui est une sorte de résumé de l'information du texte et la seconde comporte l'ensemble des propositions du texte. Dans notre module, nous questionnons la macrostructure par exemple par une question du type : *Choisissez un titre parmi quatre proposés qui conviendrait le mieux au récit que vous venez de lire* (les propositions faites sont toutes plausibles, mais l'une d'entre elles résume mieux la situation décrite dans le récit, ce qui rend le choix plus difficile).

Les deux premiers niveaux de représentation, caractéristiques de surface et base propositionnelle, sont réalisés grâce au traitement des informations linguistiques contenues dans le message, rendu possible grâce aux compétences lexicales et syntaxico-sémantiques dont dispose le lecteur. Ainsi, lorsque le sujet maîtrise bien la langue, ces deux niveaux sont réalisés de manière rapide, automatique et non volontaire. Ces deux niveaux sont évalués à l'aide de questions reprenant l'information littérale du texte (caractéristiques de

surface) et à l'aide de question de paraphrases où le traitement syntaxico-sémantique est ciblé (base propositionnelle).

La représentation la plus élaborée est le modèle de situation qui inclut la mise en jeu d'informations non strictement linguistiques, mobilisant les connaissances du comprendreur. À partir de la base propositionnelle, le comprendreur engage des opérations inférentielles qui lui permettent d'enrichir la représentation, qui devient alors une représentation non plus strictement sémantique, mais une représentation de faits, c'est-à-dire des actions, des états, des événements, etc. Ce niveau, qui va au-delà des entités linguistiques du message en intégrant les connaissances générales du comprendreur, est plus riche. Pour Kintsch, les connaissances sont en même temps le produit et la source des inférences. Ce dernier niveau peut donc être adressé à l'aide de questions de type inférences où le lecteur doit intégrer des informations linguistiques et ses connaissances (cf. en encadré 3 l'exemple d'un texte scientifique accompagné de questions).

Pour finir signalons que les processus comme la reconnaissance des mots écrits et l'analyse syntaxico-sémantique des phrases simples sont exécutés de façon automatique et irrépessible par les lecteurs qui ont une bonne maîtrise de la langue. Ils échappent à la conscience du lecteur. Les autres processus - le modèle de situation - présentent un caractère optionnel, nécessitent la mobilisation de ressources attentionnelles et prennent du temps. Leur mise en œuvre dépend en partie des objectifs du lecteur : lire pour se distraire, lire pour rechercher une information précise, lire pour acquérir de nouvelles connaissances, lire pour être évalué (comme c'est le cas présent du module) etc. Des procédures de régulation métacognitives intervenant avant, pendant et après la lecture permettront de moduler les processus en fonction d'objectifs particuliers. En conséquence, le modèle de situation est lui-même dépendant de ces objectifs.

Les processus mis en jeu dans la compréhension se développent en étroite relation avec la mémoire du lecteur. Il s'agit de la mémoire *permanente* (à long-terme) qui stocke l'ensemble des connaissances linguistiques et extralinguistiques d'une part, et de la mémoire *de travail* qui assure la gestion des processus et le maintien temporaire d'un nombre limité d'informations au fur et à mesure que le sujet lit le document d'autre part. D'après Baddeley (1986) la mémoire de travail est un système de capacité limitée, chargée du maintien temporaire et du traitement des informations pendant la réalisation de tâches cognitives complexes telles que la compréhension ou le raisonnement (Gathercole et Baddeley, 1993). Ainsi, le compreneur au fur et à mesure qu'il prend connaissance des informations du texte sollicite sa mémoire de travail puisqu'il doit à la fois traiter les informations et les stocker au moins temporairement, dans le but de comprendre le texte ou le document. La mémoire de travail est un système complexe doté de plusieurs composants, dont un processeur chargé de la gestion de l'attention et du contrôle et de deux sous-systèmes impliqués l'un dans le traitement des informations verbales, et l'autre dans le traitement des informations spatiales et imagées. Ces deux sous-systèmes seraient indépendants et spécialisés (Baddeley et Logie, 1999 ; Logie, 1995), ce qui signifie que le traitement des informations verbales est réalisé indépendamment du traitement des informations spatiales. Des recherches ont montré que la présence d'illustrations dans un texte écrit facilite la compréhension. Cependant, ce bénéfice n'est possible que chez les individus disposant d'un bon niveau de capacité de mémoire de travail (Krulley *et al.*, 1994 ; Gyselinck *et al.*, 2002).

Trois contraintes de l'évaluation

Le module Haut vise donc à évaluer la compréhension de l'écrit chez une population ayant subi avec succès le module d'orientation. Ce dernier (pour une description, cf. Degorre et Murat, ce numéro) consiste en la lecture d'un programme de télévision : l'enquêté répond à des questions faisant appel aux processus d'identification de mots, et à des questions relativement simples de compréhension et d'extraction d'informations (pas de mise en jeu d'inférences par exemple).

Les épreuves du module Haut ont été exclusivement créées pour les besoins de l'enquête *Information et Vie Quotidienne*. Plusieurs contraintes étaient imposées dans ce contexte. La première concerne les types de documents

que nous avons sélectionnés dans un respect de proposition de différents formats, de sorte à évaluer la mobilisation des processus de compréhension avec différents documents que la population peut rencontrer dans son quotidien. Il était donc important de proposer des supports familiers qui mobilisent des processus verbaux et des processus visuo-spatiaux. Par ailleurs, la diversité dans les types de réponses a également été préférée de sorte à éviter d'introduire de la monotonie dans le mode de réponse de la part des participants : questions ouvertes, questions à choix multiples, etc.

La deuxième contrainte découle de la procédure dans la passation des épreuves. Réalisée par un nombre important d'enquêteurs (environ 300), il était nécessaire de limiter la difficulté à ce niveau. Les épreuves devaient être faciles à faire passer, et le recueil des réponses durant la passation également. Cette contrainte, liée à la précédente, a des conséquences directes sur le choix du type de questions. Ainsi, la majorité d'entre elles sont de type choix multiples où la personne devait choisir dans un ensemble de propositions, la ou les propositions les plus appropriées ou la proposition correcte, eu égard au contenu du document. Ceci dit, nous avons aussi quelques questions « ouvertes » avec un pré-encodage fait par l'enquêteur.

La troisième contrainte est d'ordre temporel. En raison de la lourdeur de l'enquête - dans sa conception globale avec l'ensemble des modules - nous ne pouvions dépasser une demi-heure de passation par participant pour ce module. Cette contrainte temporelle a donc eu des conséquences sur les choix des épreuves et sur le nombre d'items. Il fallait donc faire des choix non seulement en termes de longueur des épreuves, mais également en termes de types de documents. Les types de documents proposés ne couvrent donc pas l'ensemble des formats que la population utilise régulièrement (pas de document type mode d'emploi, recette de cuisine, etc.).

Tous les documents du module Haut sont proposés aux participants sélectionnés. Ils disposaient de trente minutes pour passer les cinq épreuves et aucun temps limite n'était imposé par épreuve. L'enquêteur étant présent, il est évident qu'en cas de difficulté face à une question, celui-ci invitait l'enquêté à passer à la suivante (notons que les enquêteurs ont bénéficié d'une formation spécifique pour la passation de cette enquête).

L'ordre des épreuves était fixe et imposé : il a été fixé de manière à alterner entre les textes et les autres documents : le récit (épreuve 1), le graphique (épreuve 2), le texte expositif-explicatif (épreuve 3), le texte scientifique (épreuve 4) et le plan d'un village (épreuve 5).

Les cinq épreuves du module Haut

Nous donnerons pour chacune des cinq épreuves du module Haut des indications sur les processus sollicités à la fois lors de la compréhension du document lui-même, mais également lors de la compréhension des questions. Les épreuves 1, 3 et 4 concernent les textes, et les épreuves 2 et 5 concernent les documents associant informations visuo-spatiales et informations verbales (cf. encadré 3 pour quelques exemples de textes).

Le récit, première épreuve

Le récit est une structure de texte qui comporte diverses entités, des personnages, des objets qui participent à des états, des actions qui évoluent dans un espace et dans le temps. Des relations entre les entités de type causal, spatial, temporel sont décrites dans les récits. Pourquoi avoir retenu ce type de documents ? D'une part, parce que ce document est largement utilisé dans les documents de la vie quotidienne de type journaux écrits (rubrique de faits divers, etc.). Par ailleurs, la structure de récit permet d'évaluer non seulement le contenu de la représentation globale élaborée par le compreneur, mais elle permet également de mesurer la compréhension de marques spécifiques de cohésion. Ainsi, le récit permet de mesurer la capacité à se construire une représentation mentale cohérente des faits décrits dans le récit à partir de l'analyse des différentes relations entre les entités. Le lecteur doit comprendre le vocabulaire, compren-

Encadré 3

DEUX EXEMPLES D'ÉPREUVES : UN RÉCIT ET UN TEXTE SCIENTIFIQUE

Récit d'un fait divers (épreuve 1)

Le récit que nous avons retenu (épreuve 1) est celui qui pourrait être écrit dans une rubrique de faits divers où la plausibilité réelle du scénario peut être questionnée. Il est relaté dans ce récit que deux ouvriers travaillant dans les souterrains d'une ville auraient aperçu un lion. Le principe de l'épreuve est simple : on présente dans le désordre sept phrases écrites les unes à la suite des autres dans un ordre fixe (notées de A à G) et la première tâche que le participant a à effectuer est d'ordonner ces phrases de manière à constituer une suite de phrases respectant la cohérence. Celles-ci sont construites de telle sorte qu'un seul ordre est possible. Dans une première étape, l'enquêté doit rechercher la phrase introductive qui situe le thème du récit : « *La nouvelle est tombée ce matin dans la dépêche du journal de la ville* ». Les autres phrases sont exclues, par exemple le sujet peut exclure d'emblée le fait qu'un texte débute par « *En effet* » etc. Il engage un certain nombre de traitements allant des unités mots, aux phrases puis à l'organisation du récit pour rechercher la cohérence entre les phrases (sur le plan local) et sur l'ensemble (sur le plan global). Dans l'exemple ci-dessous, le connecteur « *En effet* » (phrase C) permet de faire un lien entre deux phrases : il introduit une confirmation de la dépêche du matin et la précise. Plusieurs autres marquages doivent être pris en compte pour agencer les phrases. Ainsi par exemple, le pronom *Ils* (au pluriel) fait référence à des personnages précédemment cités (*les ouvriers*) et donc la phrase contenant le pronom (phrase B) ne peut venir qu'après la phrase contenant les référents du pronom

(phrase C). De plus, nous apprenons qu'un lion s'est échappé d'un zoo (phrase B) et qu'il faut le rendre à ses propriétaires (phrase A). L'enquêté ici doit établir une inférence en partie causale : un lion est présent parce qu'il s'est échappé du zoo et donc il faut l'attraper pour le rendre à ses propriétaires (si on ne l'attrape pas, il pourrait faire des dégâts dans la ville et devenir dangereux pour la population etc.). Nous ne donnons ici que quelques exemples d'éléments, que le participant a à traiter. L'ordre des phrases ici est donc : D, C, B, A.

Exemple fictif limité à quatre phrases :

- A. *Ensuite, une équipe de pompiers a été déployée par la mairie pour tenter d'attraper cet animal et le rendre à ses propriétaires.*
- B. *Prévenant les services de la municipalité, ils apprenaient qu'un lion s'était échappé du zoo situé à quelques kilomètres de la ville.*
- C. *En effet, deux ouvriers de la mairie ont découvert un animal sauvage qui errait dans les souterrains de la ville.*
- D. *La nouvelle est tombée ce matin dans la dépêche du journal de la ville !*

etc.

Dans cette épreuve, le participant répond ensuite à deux autres questions : l'une d'entre elles vise à extraire la proposition thématique du récit ciblant donc l'ana-



dre les phrases et assembler les phrases en un tout cohérent, en analysant aussi les marques de la cohésion textuelle (cf. encadré 3).

La lecture et l'interprétation d'un graphique, deuxième épreuve

Il s'agit d'un graphique de type histogramme décrivant la répartition du trafic, du kilométrage, des victimes et des tués selon le type de voies (autoroutes, routes nationales, routes départementales et voies urbaines). Le lecteur doit donc combiner des informations visuo-spatiales et des énoncés verbaux. Comme précisé dans la partie théorique, les graphiques permettent de renforcer la représentation mentale élaborée sur la base des énoncés verbaux et ont donc une fonction facilitatrice. Ils sont susceptibles de participer à la construction d'un modèle de situation

qui est élaboré sur la base d'informations figuratives et verbales. Le participant doit répondre à trois questions tout en gardant sous les yeux le graphique. Pour inciter le lecteur à se référer au graphique (et non à ses connaissances), les questions commencent toutes par « *D'après le graphique ...* ». Elles varient en termes de difficulté : les deux premières font appel à des recherches d'informations avec une analyse des caractéristiques de surface, tandis que la troisième sollicite des processus plus complexes où le lecteur doit mettre en relation les deux variables et doit faire appel à une inférence simple (l'information n'est pas donnée par une simple lecture). Notre choix s'est orienté vers ce type de document en raison de l'utilisation fréquente que l'on peut en faire au quotidien, mais également parce qu'il était intéressant d'évaluer la compréhension associant des énoncés verbaux et des informations visuo-spatiales.

Encadré 3 (suite)

lyse globale du texte et la représentation mentale de la situation décrite dans le texte (choix d'un titre parmi quatre propositions toutes plausibles : *Dépêche insolite ! / Un lion dans la ville ! / Les pompiers étaient là ! / Le zoo a déployé des recherches !*). La dernière question concerne la compréhension d'un mot de vocabulaire (on extrait un mot du texte « *échappé* » puis on propose quatre autres mots (*enfui, perdu, arrivé, regagné*) et le participant doit choisir lequel de ces mots pourrait remplacer le mot, cette question pouvant être réussie même si le participant échoue dans la mise en ordre des phrases. Classiquement les connaissances de vocabulaire sont évaluées sur l'analyse de la signification de mots isolés, ce qui n'est pas le cas ici, puisque les mots cibles (e.g. « *échappé* ») sont insérés dans l'une des phrases du texte. L'enquêteur peut donc prendre en compte l'analyse syntactico-sémantique de la phrase pour répondre à cette question.

Texte théorique scientifique (épreuve 4)

Le texte est ici assez court comparé aux autres textes comme les épreuves 1 et 3. Les participants découvrent un texte concernant un aspect du fonctionnement de la mémoire humaine. Voici le texte présenté et les questions.

La mémoire humaine : l'effet de primauté

Un exercice de mémoire consiste à lire à une personne une liste de quinze mots et à lui demander de rappeler ces mots immédiatement après. On constate alors que les mots qui ont été présentés en premier sont bien rappelés. On appelle cela « l'effet de primauté ». Une explication est qu'il dépend du processus d'auto-répétition, qui consiste à se répéter mentalement les mots de la liste à apprendre, au fur et à mesure de leur présentation. Ainsi, les premiers mots présentés auraient

plus de chance d'être transférés dans une mémoire permanente, peu sensible à l'oubli.

Question 1. Les mots présentés en premier sont :

- bien rappelés ;
- moyennement rappelés ;
- peu rappelés.

Question 2. Les mots qui ont une forte probabilité de se trouver en mémoire permanente sont ceux qui ont été auto-répétés :

- un grand nombre de fois ;
- un petit nombre de fois ;
- peu importe le nombre de fois.

Question 3. Si l'on empêche l'auto-répétition, l'effet de primauté devrait :

- augmenter ;
- rester intact ;
- disparaître.

La première question fait appel à l'analyse de surface du texte puisque l'enquêteur doit repérer dans le texte le segment suivant : « les mots qui ont été présentés en premier sont bien rappelés ». L'information est donnée d'emblée et respecte la forme (appariement littéral entre la question et le texte). La deuxième question est de type paraphrase ciblant les processus syntactico-sémantiques : l'information est donnée dans le texte sous une autre forme. Le participant, grâce à ses connaissances lexicales et syntactico-sémantiques, n'aura aucune difficulté à répondre à cette question. Enfin la troisième question nécessite un calcul supplémentaire qui doit être mis en œuvre à partir des informations du texte. C'est une question introduite par *si ...* invitant le participant à raisonner sur une proposition.

Lire et comprendre un texte explicatif-explicatif dont la longueur est plus importante que les autres, troisième épreuve

Le texte inclut six paragraphes composés de 20 phrases et 500 mots. Ce texte décrit les caractéristiques d'une terre découverte récemment dans l'océan Indien. La découverte de ce territoire questionne beaucoup les géographes, les historiens, les archéologues etc. Les scientifiques posent des questions et tentent d'apporter des réponses (végétation curieuse, y'a-t-il eu des hommes sur ce territoire ? Etc.). Ce texte explicatif-explicatif décrit un problème (ici les mystères et questionnements), les observations et actions menées pour tenter de résoudre ces mystères, et une ébauche de résolution du problème (ébauches de réponses aux questionnements des scientifiques). Il décrit donc un certain nombre de relations spatiales, séquentielles, causales entre des états, des événements. Ce type de texte et les questions qui l'accompagnent sont classiquement employés lorsque l'objectif est de recueillir un niveau global de compréhension. Il comporte six questions ciblant différents processus : connaissance du vocabulaire, analyse syntaxico-sémantique et organisation textuelle. Certaines questions plus complexes visent à évaluer la construction de la cohérence via l'élaboration d'inférences, l'enquêté devant mobiliser ses connaissances et les mettre en relation avec les informations du texte.

Lire et comprendre un texte théorique de nature scientifique, quatrième épreuve

L'épreuve fait appel à un certain degré d'abstraction dont le thème porte sur le fonctionnement d'une capacité humaine : la mémoire. Le modèle de situation élaboré est un modèle conceptuel, qui se veut étudier la capacité à construire une représentation mentale cohérente en l'absence *a priori* de connaissances spécifiques du participant. Ce dernier doit répondre à trois questions plus ou moins complexes : la première porte sur l'extraction d'informations de surface, la seconde porte sur l'analyse syntaxico-sémantique d'une phrase et pour la dernière, et la troisième adresse le modèle de situation (pour un exemple analogue, cf. encadré 3).

Évaluer la capacité à former une représentation de type modèle visuo-spatial, cinquième épreuve

Cette épreuve originale associe des informations visuo-spatiales et des énoncés verbaux. Le par-

ticipant est invité tout au long de l'épreuve à se déplacer mentalement dans un village. Six étapes sont proposées et pour chacune d'elles, une phrase introductive située en haut de la page lui indique l'itinéraire à prendre (e.g. : « *Suivez la rue des Hêtres puis passez devant le cimetière et vous arriverez devant la place de l'Église* »). Puis, il doit répondre à une question (e.g. : « *Lequel de ces trois itinéraires correspond à celui que vous venez de lire ?* »). Les questions sont plus ou moins complexes. En effet, pour certaines, l'enquêté doit rechercher l'information qui est en quelque sorte donnée d'emblée dans le document. Par contre pour d'autres, il doit mettre en relation, l'énoncé verbal de la question, la description imagée du plan et la légende (l'une des questions faisait appel au repérage des points cardinaux par exemple). Ainsi pour répondre correctement, le participant doit appairer des énoncés verbaux (la phrase introductive et la question) et la représentation spatiale du plan de village (elle-même accompagné de mots écrits : des noms de rue par exemple). Il était important de proposer une telle épreuve car dans la vie quotidienne, nombre de fois, nous sommes sollicités à comprendre des informations spatiales pour se déplacer en voiture, à pied, ou à vélo. Les personnes parviennent avec plus ou moins de difficulté à se former une représentation correcte du lieu de destination, et donc à élaborer un modèle spatial cohérent.

Trois analyses : discrimination, dimensionnalité et fonctionnements différentiels

Pour chaque épreuve, le nombre de questions variait de 3 à 6, ce qui est peu. Chaque question donnait lieu à une cotation binaire : 1 point pour la réponse correcte et 0 pour toute réponse erronée. Au total, 21 questions ont été posées, donc sur l'ensemble, le score maximal théorique est de 21 pour chaque participant. En raison du faible nombre d'items, nous serons prudents dans l'interprétation des résultats. Il est vrai que les épreuves sont diversifiées - et nous l'avons voulu - mais l'exploitation de cette diversité sera difficile compte tenu du nombre d'items par épreuve. L'analyse des dimensions sous-jacentes à ce module devra donc être interprétée avec prudence.

Trois types d'analyses des données ont été effectuées sur les 7 389 *individus* qui ont été orientés vers le module Haut : en premier une analyse

classique descriptive de l'épreuve, puis une analyse des dimensions sous-jacentes et enfin une analyse concernant les fonctionnements différentiels liés au sexe des participants, puisque notre module s'est révélé sensible à cette variable. L'analyse de la structure du module Haut s'inscrit dans une démarche de validation théorique mais également dans la perspective de la prochaine vague de l'enquête *IVQ* qui devrait intégrer un nombre d'items plus important dans ce module. En effet, si l'étude de la dimensionnalité du module haut est fragilisée par le faible nombre d'items (21), elle doit néanmoins permettre d'apporter quelques indications pour élaborer d'autres items.

Taux de réussite et discrimination : premières analyses

Le taux de réussite et le $r_{bis-point}^*$ permettent de mesurer la difficulté et la discrimination des 21 items du module Haut. En particulier, le

$r_{bis-point}^*$ indique dans quelle mesure l'item s'inscrit dans la dimension générale (corrélation entre l'indicatrice de réussite à l'item et le score global obtenu aux autres items de l'évaluation). Il indique également la différence de performance constatée entre les individus qui réussissent l'item et ceux qui l'échouent (cf. encadré 4).

Le taux de réussite moyen sur l'ensemble du module s'élève à près de 70 %, ce qui indique que le module est une épreuve plutôt facile (cf. tableau 1). Conformément à nos attentes, les items sont de difficulté variable : le taux de réussite est compris entre 31,7 % et 92,5 %, taux qui varie selon la complexité de la question et les niveaux de représentation sollicités. Ainsi, les questions adressant les caractéristiques de surface, autrement dit les questions faisant appel à l'extraction littérale des informations du texte, sont mieux réussies que celles qui sollicitent les relations entre langage et cognition comme les questions inférentielles.

Encadré 4

ANALYSES CLASSIQUES DES ITEMS

On note n le nombre d'individus ayant passé une évaluation composée de K items.

On note Y_i^j la réponse de l'individu i à l'item j . Dans notre cas, les items sont dichotomiques, c'est-à-dire qu'ils ne prennent que deux modalités (la réussite ou l'échec) :

$$Y_i^j = \begin{cases} 1 & \text{si l'individu } i \text{ réussit à l'item } j \\ 0 & \text{si l'individu } i \text{ échoue à l'item } j \end{cases} \quad (i = 1, \dots, n) \quad (j = 1, \dots, K)$$

Le *taux de réussite* à l'item j est le pourcentage d'individus ayant réussi l'item j . Il est noté p_j :

$$p_j = \frac{1}{n} \sum_{i=1}^n Y_i^j$$

Le *score observé* à l'évaluation par l'élève i , noté S_i , correspond au nombre d'items réussis par l'individu i :

$$S_i = \sum_{j=1}^K Y_i^j$$

Le $r_{bis-point}$ est le coefficient de corrélation linéaire entre la variable indicatrice de réussite à l'item Y et le score S . Cet indice est appelé « $r_{bis-point}$ » pour le distinguer du coefficient r biserial (r_{bis}) qui est le coefficient de corrélation linéaire entre le score et une variable latente, régie par une loi normale, conditionnant la réussite à l'item. Les deux indices entretiennent une relation fonctionnelle et le r_{bis} est en général plus élevé, plus particulièrement pour les items très réussis ou très échoués.

C'est un indice de *pouvoir discriminant* : il indique dans quelle mesure l'item s'inscrit dans la dimension générale (corrélation item-test). Il indique également la différence de performance constatée entre les individus qui réussissent l'item et ceux qui l'échouent.

En effet, on peut montrer que

$$r_{bis-point}(j) = \text{cor}(Y^j, S) = \frac{m_j - m^n}{\sigma_S} \cdot \sqrt{p_j(1-p_j)}$$

où m_j est le score moyen sur l'ensemble de l'évaluation des individus ayant réussi l'item, m^n celui des élèves l'ayant échoué et σ_S est l'écart-type des scores.

De manière générale, on préfère s'appuyer sur l'indice corrigé $r_{bis-point}^*$ qui est le coefficient de corrélation linéaire entre la variable indicatrice de réussite à l'item Y^j et le score ($S - Y^j$), c'est-à-dire le score obtenu aux items de l'évaluation, sauf l'item j . En pratique, une valeur inférieure à 0,2 indique un item peu discriminant (Laveault et Grégoire, 2002).

Le coefficient α de *Cronbach* est un indice mesurant la consistance interne d'un ensemble d'items censés mesurer la même dimension. Il est très lié à la moyenne des corrélations inter-items. Compris entre 0 et 1, il traduit un degré d'homogénéité des items d'autant plus élevé que sa valeur est proche de 1 (Dickes et al., 1994).

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum_j p_j(1-p_j)}{\sigma_S^2} \right)$$

Les indices de discrimination ($r_{bis-point}^*$) s'étalent de 0,17 à 0,42. Généralement, un item dont le $r_{bis-point}^*$ est inférieur à 0,20 est considéré comme peu discriminant. De nombreuses raisons peuvent expliquer une faible discrimination : mauvaise formulation de la question, ambiguïté pour certaines questions sur les différents choix proposés, tâche faisant appel à une compétence différente de celle visée par les autres items, etc. (Laveault et Grégoire, 2002). Par exemple, l'item h2 était une question concernant la thématique du récit (une sorte de résumé) à savoir « *D'après vous quel titre résume le mieux le texte ?* ». Plusieurs choix de titres étaient proposés dont l'un était le plus approprié mais les autres pouvaient être plausibles (mais moins appropriés) (pour un exemple, cf. encadré 3). L'indice à 0,20 sur cet item indique que le choix d'un de ces titres ne présage pas assurément de la réussite ou de l'échec aux autres items du module. Les distracteurs (les autres choix possibles pour répondre) ici ne sont sans doute pas suffisamment « éloignés » de la proposition la

plus appropriée. À l'inverse, l'item h22 - qui concernait la description d'un itinéraire sur la base d'indications variées - avec un indice de 0,42, est très lié à la dimension générale qui est mesurée dans cette épreuve. C'est aussi le cas de l'item h1 (remise en ordre de sept phrases).

Dans l'ensemble, les valeurs des $r_{bis-point}^*$ sont satisfaisantes mais pas très élevées, ce qui témoigne d'un niveau moyen d'inter-corrélations entre les items. D'ailleurs, la valeur de l'indice α de Cronbach qui est de 0,74 révèle un degré moyen d'homogénéité de l'épreuve (cf. encadré 4).

Dimensionnalité de l'épreuve, deuxièmes analyses

L'homogénéité de l'épreuve, telle que mesurée par l'indice α de Cronbach, renvoie également au caractère unidimensionnel de l'épreuve. L'unidimensionnalité de l'épreuve est vérifiée si les performances des participants aux items proposés sont expliquées par le degré de maîtrise d'une seule et même compétence. Pour de nombreuses évaluations, l'unidimensionnalité de l'épreuve est une propriété recherchée (Rocher, 2003). Cette propriété permet en effet d'appliquer des modèles statistiques tels que les modèles de réponse à l'item (cf. Murat et Rocher, ce numéro). Plus simplement, le calcul d'un score « brut » - nombre de bonnes réponses - suggère qu'il est possible d'additionner les résultats obtenus aux items et donc que chacun d'entre eux mesure la même dimension.

Le module Haut cible différents processus mis en œuvre dans la compréhension de documents écrits allant du niveau du mot (vocabulaire par exemple) aux phrases puis à l'organisation du document dans son ensemble, où le sujet doit établir la cohérence locale et globale (cf. encadré 2). Dès lors, il est intéressant d'étudier dans quelle mesure ces différents aspects sous-tendent les réponses observées. Autrement dit, la structure des relations entre les items reflète-t-elle les différents processus mis en jeu dans la compréhension de documents de différents formats ? La dimensionnalité de l'épreuve peut ainsi être entendue comme le nombre de « compétences » (ou dimensions latentes) nécessaires pour répondre correctement aux items proposés (Nandakumar, 1994).

Pendant, la dimensionnalité d'un ensemble d'items n'est pas facile à apprécier car elle dépend de nombreux facteurs extérieurs liés aux

Tableau 1
Taux de réussite et pouvoir discriminant ($r_{bis-point}^*$) des items du module Haut

	Items (1)	Taux de réussite (%)	$r_{bis-point}^*$
Récit	h1	43,6	0,42
	h2	78,5	0,20
	h3	84,7	0,29
Graphique	h4	83,0	0,31
	h5	88,7	0,32
	h6	72,6	0,23
Texte expositif	h8	84,0	0,25
	h9	86,2	0,33
	h10	38,3	0,27
	h11	31,7	0,26
	h12	74,4	0,36
h13	61,6	0,35	
Texte théorique	h14	78,5	0,31
	h15	61,1	0,17
	h16	52,2	0,22
Cartes spatiales	h17	80,0	0,30
	h18	60,2	0,28
	h19	92,5	0,37
	h20	63,7	0,36
	h21	79,1	0,37
	h22	70,0	0,42

1. L'item h7, avec un taux de réussite de 11,3 % et un $r_{bis-point}^*$ de 0,04 a été écarté de l'analyse. Cet item, qui nécessite une lecture croisée de différentes données présentes dans le graphique (épreuve 2), comportait des ambiguïtés dans sa formulation.

Lecture : l'item h1 est réussi par 43,6 % des individus ayant passé le module Haut d'IVQ. Avec un $r_{bis-point}^*$ de 0,42, cet item affiche un bon pouvoir discriminant (cf. encadré 4).

Champ : individus ayant passé le module Haut de l'enquête IVQ.

Source : calculs à partir des données de l'enquête Information et Vie Quotidienne 2004, Insee.

individus et aux items (Juhel, 1999). La motivation, la préférence pour certains thèmes, le format des items, etc., sont autant de paramètres qui influencent la structure des réponses. C'est pourquoi de nombreuses méthodes ont été mises au point pour étudier la dimensionnalité d'un ensemble d'items (Tate, 2003) et l'analyse de la structure des réponses aux items est l'objet de recherches récentes s'appuyant sur des modélisations complexes (Goldstein *et al.*, 2007). Deux approches sont envisagées ici (cf. encadré 5) : une approche exploratoire reposant sur l'analyse factorielle des corrélations entre items et une approche confirmatoire basée sur des méthodes non paramétriques.

Analyses factorielles exploratoires

Les analyses factorielles (cf. encadré 5) sont employées ici de manière exploratoire. Un premier facteur « dominant » apparaît clairement (cf. graphique) : la première valeur propre

vaut 5,9, la deuxième 1,4 et la troisième 1,3. Quantité d'indices ont été proposés pour juger de l'unidimensionnalité à partir de ces résultats (Hattie, 1985). Par exemple, le rapport des deux premières valeurs propres vaut ici 4,2 ce qui tend à indiquer une structure unidimensionnelle. Ce premier résultat, en faveur d'une structure unidimensionnelle, traduirait le fait que les processus ciblés par nos items ne sont pas indépendants les uns des autres. Leur mise en œuvre dépend d'une compétence langagière générale. Elle indique également que ces processus sont mis en jeu dans la compréhension des deux formats verbal et visuo-spatial. Comme le soulignent Gernsbacher, 1990 et 1994 ; Graesser *et al.*, 2002 ; Traxler et Gernsbacher, 2006, la compréhension est une activité complexe dépendante d'un système général impliqué dans le traitement des informations verbales et imagées.

Cependant, comparées aux valeurs obtenues sur des données simulées sur le même nombre

Encadré 5

DEUX APPROCHES POUR ÉTUDIER LA DIMENSIONNALITÉ D'UN ENSEMBLE D'ITEMS

Approche exploratoire : analyse factorielle sur la matrice des coefficients de corrélation tétrachoriques et rotation *Promax*

Pour déterminer les dimensions sous-jacentes à un ensemble de variables, l'analyse factorielle est une méthode classiquement employée. Dans le cas de l'analyse factorielle d'un ensemble d'items dichotomiques, le coefficient de corrélation linéaire entre deux items - qui correspond au coefficient $\varphi^2 = \chi^2 / n$ où n est la taille de l'échantillon - montre certaines limites, en supposant une relation linéaire. Il est préférable de se baser sur le coefficient de corrélation tétrachorique, pour ses propriétés supposées d'invariance selon les groupes d'individus (Lord et Novick, 1968).

Soient deux items Y^1 et Y^2 ayant respectivement pour taux de réussite p_1 et p_2 . On considère alors l'existence de deux variables latentes continues Z^1 et Z^2 , dont le couple (Z^1, Z^2) suit une loi normale bivariée, et telles que

$$Y_i^1 = \begin{cases} 1 & \text{si } Z_i^1 \geq \alpha_1 \\ 0 & \text{si } Z_i^1 < \alpha_1 \end{cases} \text{ et } Y_i^2 = \begin{cases} 1 & \text{si } Z_i^2 \geq \alpha_2 \\ 0 & \text{si } Z_i^2 < \alpha_2 \end{cases}, (i = 1, \dots, n)$$

où $\alpha_1 = \psi^{-1}(p_1)$ et $\alpha_2 = \psi^{-1}(p_2)$ avec ψ la fonction de répartition de la loi normale centrée réduite.

Le coefficient de corrélation tétrachorique est le coefficient de corrélation linéaire entre Z^1 et Z^2 .

Approche confirmatoire : la procédure *Dimtest*

Stout (1987) a proposé et raffiné (Stout *et al.*, 2001) un test non paramétrique de l'unidimensionnalité d'un ensemble d'items.

L'idée principale consiste à tester l'indépendance locale d'un ensemble d'items. L'indépendance locale implique que les réponses aux items sont indépendantes, conditionnellement au niveau de compétence des individus. Cette hypothèse est la conséquence directe de l'idée que, sous l'hypothèse d'unidimensionnalité, seul le niveau de compétence conditionne la réussite aux items.

La procédure *Dimtest* teste l'hypothèse faible d'indépendance locale, c'est-à-dire que pour chaque paire (j, k) d'items et chaque niveau de compétence : θ

$$\text{Cov}(Y^j, Y^k / \Theta = \theta) = 0$$

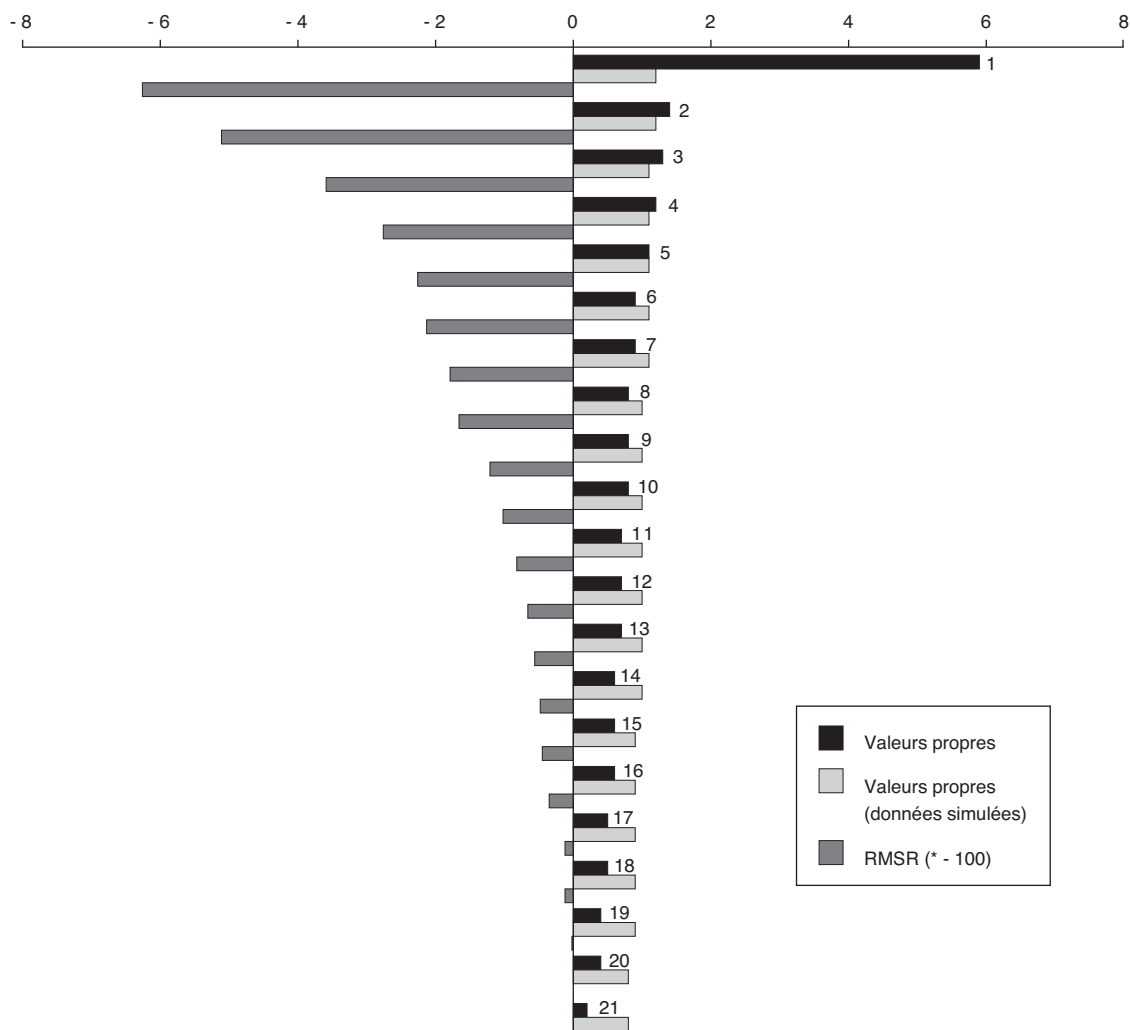
En pratique, on choisit un groupe d'items (*AT*, « *assessment test* ») dont on veut vérifier s'ils forment une dimension. Le groupe d'items restant (*PT*, « *partitioning test* ») sert à fixer les niveaux de performances. Une statistique T_L est alors construite à partir des covariances observées entre les items *AT*, conditionnellement au score observé sur l'ensemble des items *PT*. Cette statistique est enfin comparée à la même statistique, notée T_G , obtenue sur des données simulées selon un cadre unidimensionnel. La différence standardisée entre les deux statistiques, notée T , suit asymptotiquement une loi normale centrée réduite.

d'individus et sur le même nombre d'items de réussites équivalentes, quatre valeurs propres sont supérieures à 1, ce qui conduirait à étudier les quatre premiers facteurs (Tate, 2003). Pour approfondir cette analyse exploratoire, nous présentons les valeurs de l'indice RMSR (*Root Mean Square Residual*, (cf. graphique). Il s'agit de l'écart résiduel standardisé (moyenne géométrique des résidus) entre la matrice des corrélations d'origine et la matrice des corrélations « reconstruite » à partir de la factorisation retenue. La décroissance de cet indice conforte le choix de retenir quatre facteurs.

Une analyse factorielle considérant quatre facteurs communs est donc engagée, avec une rota-

tion oblique (*Promax*) des axes, car l'hypothèse la plus crédible est que ces facteurs ne sont pas indépendants. L'examen des corrélations entre les items et les quatre facteurs (cf. tableau 2 où elles sont surlignées en gras quand elles sont supérieures ou égales à 20) fait clairement apparaître que les items portant sur le traitement du plan de village (cartes spatiales, épreuve 5) forment une dimension se détachant des autres items (facteur 1, items de h17 à h22). Le quatrième facteur rassemble les trois items portant sur le graphique (h4, h5 et h6, épreuve 2) qui associe des informations verbales et visuo-spatiales. Les facteurs 2 et 3 regroupent les items portant sur les épreuves exclusivement verbales à savoir les épreuves 1, 3 et 4 (à l'exception de l'item 6 sur l'épreuve 2).

Graphique
Analyse en composantes principales à partir des coefficients de corrélation tétrachoriques



Lecture : la partie droite du graphique représente les valeurs propres de l'analyse en composantes principales des items du module Haut à partir des coefficients de corrélation tétrachoriques (cf. encadré 5). En blanc, il s'agit des valeurs propres obtenues par une analyse identique effectuée cette fois sur des données simulées prenant en compte les taux de réussite aux items et considérant que les items sont indépendants. La partie gauche représente l'indice RMSR – Root Mean Square Residual.
Champ : individus ayant passé le module Haut de l'enquête IVQ.
Source : calculs à partir des données de l'enquête Information et Vie Quotidienne 2004, Insee.

Ainsi, il apparaît que la représentation mentale élaborée au cours de la compréhension ne présente pas les mêmes caractéristiques lorsqu'il s'agit de traitement de textes écrits (épreuves 1, 3, et 4) que lorsqu'il s'agit de traitement de graphiques (épreuve 2) ou de cartes spatiales (épreuve 5). Dans les épreuves verbales (1, 3 et 4 avec des énoncés écrits uniquement), les modèles sollicitent des processus linguistiques sur la base des compétences langagières du sujet (lexicales, syntaxico-sémantiques et organisation textuelle), tout en mobilisant les connaissances du monde du sujet. En dépit du fait que ces trois épreuves sont constituées de textes de structure différente - l'épreuve 1 étant un récit, l'épreuve 3 un texte théorique, et l'épreuve 4 un texte expositif - ce n'est pas sur cette base que les facteurs 2 et 3 se différencient. Il n'est d'ailleurs pas aisé d'expliquer l'émergence de ces deux dimensions. Parmi les items qui saturent sur le facteur 2, un bon nombre d'entre eux (mais pas tous) adresse les compétences mesurant strictement la maîtrise des informations

verbales (compétences lexicales et syntaxico-sémantiques) alors que les items qui saturent sur le facteur 3 ont tendance à cibler le modèle de situation où le compreneur sollicite d'autres processus comme la mise en jeu de connaissances du monde. D'ailleurs nous pouvons y trouver là une explication à la saturation de l'item h6 sur le facteur 3, lequel, bien que portant sur le graphique, sollicitait les connaissances du participant (il est probable que certains aient répondu uniquement sur la base de leurs connaissances). L'explication de la distinction entre les facteurs 2 et 3 n'est qu'une première tentative qui doit être considérée avec beaucoup de prudence parce que les niveaux de représentation ne sont pas si tranchés dans les questions (par exemple, le vocabulaire est évalué parfois en lien avec l'analyse syntaxico-sémantique de la phrase). La saturation de certains items sur les deux facteurs 2 et 3 pourrait appuyer cette remarque. En effet l'item h1 par exemple, qui concernait la remise en ordre de sept phrases, nécessite à la fois l'extraction des informations

Tableau 2
Analyse factorielle avec rotation oblique (*Promax*)

A - Corrélations semi-partielles entre les items et les facteurs de l'analyse factorielle en rotation oblique en quatre facteurs

	Items	Facteur 1	Facteur 2	Facteur 3	Facteur 4
Récit	h1	0,10	0,28	0,24	0,02
	h2	0,05	0,20	0,06	0,00
	h3	0,07	0,08	0,32	0,03
Graphique	h4	0,05	0,01	0,06	0,59
	h5	- 0,01	0,06	- 0,05	0,87
	h6	0,05	- 0,18	0,37	0,22
Texte expositif	h8	- 0,07	0,33	0,10	0,08
	h9	0,00	0,35	0,15	0,05
	h10	0,04	- 0,03	0,42	- 0,05
	h11	0,01	0,11	0,29	- 0,02
	h12	- 0,04	0,15	0,47	- 0,02
	h13	0,01	0,26	0,25	- 0,03
Texte théorique	h14	0,00	0,50	- 0,06	0,00
	h15	0,05	0,24	- 0,06	- 0,01
	h16	0,03	0,08	0,14	0,10
Cartes spatiales	h17	0,36	0,09	- 0,03	0,03
	h18	0,43	- 0,13	0,10	0,02
	h19	0,44	0,26	- 0,05	0,02
	h20	0,41	0,04	0,04	0,03
	h21	0,37	0,13	0,07	- 0,04
	h22	0,63	- 0,03	- 0,01	0,00

B - Corrélations entre les facteurs

	Facteur 2	Facteur 3	Facteur 4
Facteur 1	0,57	0,49	0,42
Facteur 2		0,53	0,41
Facteur 3			0,33

Lecture : les corrélations supérieures à 0,20 sont en gras.

Le facteur 1 comprend les items portant sur les cartes spatiales. Il est corrélé à 0,57 avec le facteur 2.

Champ : individus ayant passé le module Haut de l'enquête IVQ.

Source : calculs à partir des données de l'enquête Information et Vie Quotidienne 2004, Insee.

linguistiques contenues dans les énoncés mais aussi une mise en correspondance entre langage et cognition.

Dans l'épreuve de traitement de cartes spatiales (épreuve 5), le compreneur s'engage dans la construction de modèles d'environnements spatiaux élaborés à partir d'informations visuo-spatiales et d'énoncés descriptifs verbaux. Le degré d'élaboration de ces modèles est variable selon le type de description proposé, selon les compétences verbales et spatiales des participants, c'est-à-dire leur niveau de compréhension des termes spatiaux et des rapports topologiques, elle-même dépendante de l'expérience perceptivo-motrice du sujet. De plus, des recherches ont montré que la finesse de la représentation dépend de la capacité de la mémoire de travail, le bénéfice de la présence d'illustrations n'étant observé que chez des individus ayant une grande capacité de mémoire de travail (Kruley *et al.*, 1994 ; Gyselinck *et al.*, 2002).

Les informations verbales et les informations spatiales constituent deux modes de présentation qui ont des propriétés très distinctes, ce qui n'exclut pas que les deux puissent fonctionner en interaction lorsque à la fois langage et espace sont sollicités. Il est alors intéressant d'examiner la spécificité des processus engagés lors du traitement des données verbales et lors du traitement des données de nature spatiale. La question de la spécificité ou non des processus engagés dans l'une et l'autre forme revient à poser la question de l'indépendance ou non du traitement des activités langagières et du traitement des activités de nature spatiale. Si nous réalisons une analyse en rotation oblique *Promax* en deux facteurs, c'est bel et bien un facteur verbal d'une part et un facteur spatial (et

figuré) d'autre part qui se distingue (cf. annexe). Rappelons que les travaux issus de l'étude des composants de la mémoire de travail prônent une indépendance des traitements : le calepin visuo-spatial prendrait en charge les informations visuo-spatiales et la boucle articulatoire serait spécialisée dans le traitement des informations verbales (Baddeley et Logie, 1999 ; Logie, 1995). Un certain nombre d'auteurs défendent l'idée que le modèle spatial est construit sur la base d'un format non-propositionnel (Glenberg *et al.*, 1994 ; Haenggi *et al.*, 1995).

Procédure de validation *Dimtest*

Afin de valider les dimensions détectées par l'analyse factorielle, une méthode confirmatoire, la procédure *Dimtest*, permet de tester si un ensemble d'items forme une dimension spécifique, comparé à un autre ensemble (cf. encadré 4). Ainsi, les quatre dimensions précédentes ont été testées, une à une, à chaque fois en comparaison avec les items restants. De plus, la dimension spatiale a été testée par rapport à la dimension verbale. Les résultats de ces analyses (cf. tableau 3) font ressortir que seules les dimensions spatiales (cartes et graphiques) se distinguent du reste des items. Les deux facteurs concernant les épreuves verbales ne sont pas confirmés, séparément. En revanche, les items portant sur les épreuves spatiales peuvent être considérés comme formant une dimension spécifique par rapport aux items portant sur les épreuves verbales. Un tel résultat serait en faveur d'une spécialisation des traitements en fonction de la nature des informations à traiter, verbale ou visuo-spatiale. Comme nous l'avons précisé, les différences entre les hommes et les femmes seraient pour certains auteurs, une confirmation que ces deux dimensions sont distinctes.

Tableau 3
Tests de dimensionnalité selon la procédure confirmatoire *Dimtest*

	AT	T_L	T_G	T	P-value	Dimension Spécifique
4 facteurs	h17 h18 h19 h20 h21 h22	7,8	5,4	2,3	0,0096	Oui
	h2 h8 h9 h14 h15	3,5	3,2	0,2	0,4023	Non
	h3 h10 h11 h12	3,8	3,2	0,6	0,2750	Non
	h4 h5	5,0	1,7	3,3	0,0005	Oui
Verbal/Spatial	Graphique et cartes	9,5	7,8	1,7	0,0446	Oui

Lecture : première ligne : le groupe de 6 items h17, h18, h19, h20, h21 et h22 constitue le groupe AT (« assessment test ») dont on veut vérifier s'ils forment une dimension, le groupe de 15 items restant PT (« partitioning test ») servant à fixer les niveaux de performances. La statistique T_L est construite à partir des covariances observées entre les items du groupe AT, conditionnellement au score observé sur l'ensemble des items PT. Cette statistique est comparée à la même statistique, notée T_G , obtenue sur des données simulées selon un cadre unidimensionnel. La différence standardisée T entre les deux statistiques suit asymptotiquement une loi normale centrée réduite. Le test est unilatéral car il s'agit de tester la supériorité de T_L sur T_G .

Les items portant sur les cartes (h17 à h22), c'est-à-dire un des facteurs de l'analyse factorielle, forment une dimension spécifique (P-value < 0,05), cf. encadré 5.

Champ : individus ayant passé le module Haut de l'enquête IVQ.

Source : calculs à partir des données de l'enquête Information et Vie Quotidienne 2004, Insee.

Pour certains, l'observation de différences entre les hommes et les femmes dans le traitement des informations spatiales offre un argument en faveur d'une spécialisation des ressources allouées au traitement des informations spatiales, et ce dès l'adolescence (cf. par exemple Choi et Sylverman, 2003 ; Dabbs *et al.*, 1998).

Fonctionnements différentiels selon la variable sexe, troisièmes analyses

Un item présente un *fonctionnement différentiel* (FDI) par rapport à une caractéristique de groupes d'individus (hommes/femmes, Français/Suédois etc.) dès lors que l'on observe des réussites différentes selon les groupes, à niveau de performance égale à l'ensemble du test. Un *fonctionnement différentiel* est à distinguer de l'*impact* de l'item, qui renvoie à l'effet général de l'appartenance à un groupe sur la performance (Vrignaud, 2002). Par exemple, les cadres réussissent mieux les items du module Haut que les ouvriers (impact) mais des items peuvent présenter un fonctionnement différentiel en faveur de l'un ou l'autre de ces deux groupes, c'est-à-dire être réussi différemment, à niveau de performance égal entre les deux groupes.

Soit donc deux items (h13 et h17) pris comme exemple : on présente (cf. tableau 4) les taux de réussite des hommes et des femmes à ces items à tranche de score égal, ce qui permet d'identifier la présence de fonctionnements différentiels. L'item h13 (épreuve texte expositif) présente un fonctionnement différentiel en ce sens que, à tranche de score égal, il est mieux réussi par

les femmes ; en revanche, l'item h17 (épreuve des cartes) ne présente pas de fonctionnement différentiel.

L'identification des fonctionnements différentiels permet notamment de repérer les items biaisés, c'est-à-dire qui mesurent autre chose que ce qu'ils sont censés mesurer. Les FDI sont également utiles pour déterminer si tel ou tel groupe peut subir un désavantage. Par exemple, la recherche de FDI est particulièrement importante dans les comparaisons internationales où elle permet de mettre à jour des « biais culturels » (Rocher, 2003).

De nombreuses méthodes ont été proposées pour identifier des FDI (Millsap et Everson, 1993). La longueur du test, la taille de l'échantillon, le nombre d'items biaisés sont autant de facteurs pouvant expliquer la meilleure adaptation de telle ou telle méthode. C'est pourquoi il est préférable de confronter les résultats de différentes méthodes. Deux méthodes ont été retenues ici : la procédure classique dite de Mantel-Haenszel telle que développée par Holland et Thayer (1988) et la procédure non-paramétrique *Sibtest* proposée par Shealy et Stout (1993). Ces deux méthodes (cf. encadré 6) ont l'avantage de proposer une graduation de l'ampleur des FDI, au-delà des tests d'hypothèse, sensibles à la taille de l'échantillon.

De manière descriptive d'abord, il apparaît que, à l'exception des items h13 et h14, tous les items sont mieux réussis par les hommes que par les femmes (cf. tableau 5). La différence moyenne de réussite est de près de cinq points

Tableau 4
Exemples de fonctionnements différentiels d'items (FDI) selon le sexe

	Répartitions (%)		Taux de réussite (%)					
			h13			h17		
	Hommes	Femmes	Hommes	Femmes	Écart	Hommes	Femmes	Écart
score ≤ 12	19,2	30,5	28,8	34,3	- 5,5	55,4	57,5	- 2,1
12 < score ≤ 14	18,5	18,3	42,1	55,8	- 13,7	80,0	80,2	- 0,2
14 < score ≤ 16	22,5	21,4	60,0	71,7	- 11,7	84,3	86,2	- 1,9
16 < score ≤ 18	23,4	18,1	77,7	83,9	- 6,2	92,2	90,4	1,8
18 < score ≤ 21	16,4	11,7	95,2	96,9	- 1,7	96,7	97,6	- 0,9
Ensemble	100,0	100,0	60,6	62,5	- 1,9	81,8	78,5	3,3

Lecture : dans l'ensemble, l'item h13 est réussi par 60,6 % des hommes et par 62,5 % des femmes. Mais à tranche de score égale, la différence de réussite entre les hommes et les femmes est plus marquée. Ces écarts selon les tranches peuvent sembler élevés en comparaison de l'écart observé sur l'ensemble des individus. Il s'agit d'un effet de structure lié aux répartitions différentes des hommes et des femmes selon les tranches de scores. Cet item présente un fonctionnement différentiel (FDI) en faveur des femmes (cf. tableau 5). En revanche, l'item h17 ne présente pas de FDI car à tranche de score égal, les différences de réussites entre les hommes et les femmes ne sont pas significatives.

Champ : individus ayant passé le module Haut de l'enquête IVQ.

Source : calculs à partir des données de l'enquête Information et Vie Quotidienne 2004, Insee.

FONCTIONNEMENT DIFFÉRENTIEL DES ITEMS (FDI)

La procédure de Mantel-Haenszel (1959)

La statistique de Mantel et Haenszel (1959) vise à mesurer et à tester le lien observé entre deux variables qualitatives dans différentes strates, c'est-à-dire décliné selon une troisième variable qualitative. Elle est souvent utilisée dans le domaine biomédical, pour analyser par exemple le lien entre différents traitements et la rémission, après avoir tenu compte d'un facteur perturbateur pouvant influencer cette relation, comme l'âge, le centre de soins, etc.

Cette statistique a été adaptée par Holland et Thayer (1988) pour détecter les fonctionnements différentiels d'items (FDI), c'est-à-dire les items pour lesquels deux groupes d'individus, à niveau de compétence égal, obtiennent des taux de réussite différents.

On considère deux groupes (par exemple les hommes et les femmes), qu'on découpe en J classes de compétences équivalentes. Le tableau ci-dessous présente les effectifs de chaque catégorie pour le niveau j :

Un cas de répartition des réponses à un item donné pour deux groupes d'égale compétence « j »

		Réponse à l'item		Total
		Correcte	Incorrecte	
Niveau j	Groupe 1	A_j	B_j	n_{1j}
	Groupe 2	C_j	D_j	n_{2j}
Total		m_{1j}	m_{2j}	T_j

La statistique de Mantel-Haenszel (Mantel et Haenszel, 1959) est donnée par :

$$MH = \frac{\left(\sum_{j=1}^J (A_j - E(A_j)) - \frac{1}{2} \right)^2}{\sum_{j=1}^J \text{Var}(A_j)}$$

$$\text{où } E(A_j) = \frac{n_{1j} m_{1j}}{T_j} \text{ et } \text{Var}(A_j) = \frac{n_{1j} n_{2j} m_{1j} m_{2j}}{T_j^2 (T_j - 1)}$$

$E(A_j) = n_{1j} \frac{m_{1j}}{T_j}$ peut être vu comme le nombre de réponses correctes « attendues » pour le groupe 1, compte tenu de son effectif (n_{1j}) et de la proportion de réponses correctes de tous les membres du niveau j : $(\frac{m_{1j}}{T_j})$.

$\text{Var}(A_j) = \frac{E(A_j)E(D_j)}{(T_j - 1)}$ peut être vu comme le produit, corrigé de $\frac{1}{T_j - 1}$, du nombre de réponses correctes

« attendues » dans le premier groupe et du nombre de réponses incorrectes « attendues » dans le deuxième groupe.

Sous l'hypothèse $H_0 : A/B = C/D$ (la réussite à l'item est indépendante du groupe considéré), on a approximativement : $MH \sim \chi_1^2$.

Holland et Thayer (1988) ont adapté cette procédure en calculant l'indice suivant :

$$\Delta = 2.35 \ln(\hat{a})$$

où $\hat{a} = \sum \frac{A_j D_j}{T_j} / \sum \frac{B_j C_j}{T_j}$, c'est-à-dire une forme de

rapport de cotes (*odds-ratio*) entre les deux groupes.

Le signe de Δ donne le sens de la différence. Une valeur positive indique que le groupe 1 est désavantagé (i.e. son taux de réussite est moins élevé que le groupe 2, à niveau de compétence égal), une valeur négative qu'il est avantagé. Une valeur absolue supérieure à 1 indique un fonctionnement différentiel important.

La procédure Sibtest (*Simultaneous Index of Bias*)

Shealy et Stout (1993) ont mis au point une procédure, Sibtest, pour détecter les fonctionnements différentiels d'items, qui s'appuient sur le cadre des modèles de réponse à l'item multidimensionnels (cf. Murat et Rocher dans ce numéro pour une présentation des modèles de réponse à l'item).

Le FDI est étudié entre deux groupes : le groupe focal (F) du groupe de référence (R). On distingue deux variables latentes : la compétence θ et une variable de nuisance η . La distribution de la variable η , conditionnellement à θ , est supposée différente selon le groupe. On considère alors que les items sensibles à cette variable de nuisance (i.e. dont les réponses dépendent de η , à θ fixé) présentent un fonctionnement différentiel.

En pratique, on détermine un ensemble d'items dits *valides* servant à fixer le niveau de compétence. Il s'agit ensuite d'étudier les différences de performances aux items *suspects* selon le niveau de compétence estimé.

La statistique suivante est calculée :

$$\hat{\beta}_U = \sum_{k=0}^n \hat{\rho}_k (\bar{Y}_{Rk}^* - \bar{Y}_{Fk}^*)$$

où $\hat{\rho}_k$ est la proportion d'individus du groupe de référence qui se situent au niveau k de score sur les n items valides, \bar{Y}_{Rk}^* - respectivement \bar{Y}_{Fk}^* - est le score moyen observé sur l'item (ou le groupe d'items) à l'étude, pour les individus du groupe de référence - respectivement du groupe focal - situés au niveau k de score sur les n items valides, \bar{Y}_{Rk}^* et \bar{Y}_{Fk}^* sont des transformations de ces deux scores afin de tenir compte du fait que les deux groupes ont des distributions différentes sur θ . Dans ce cadre, $\hat{\beta}_U / \hat{\sigma}(\hat{\beta}_U)$ suit asymptotiquement une loi normale centrée réduite.



de pourcentage. Ce résultat peut sembler surprenant étant donné les résultats supérieurs des filles en lecture enregistrés dans le cadre scolaire ou chez les jeunes de 17 ans lors de la Journée d'appel de préparation à la défense (voir la note d'information rédigée par de La Haye *et al.*, 2007). Si l'âge et le contexte pourraient expliquer ce décalage par rapport à d'autres enquêtes, il faut rappeler que les individus passant le module Haut sont pré-sélectionnés et que les hommes connaissent plus souvent que les femmes de graves difficultés

de lecture (Degorre et Murat, ce numéro). Ce « biais de sélection » explique en partie les meilleures performances des hommes aux épreuves du module Haut. Sur l'ensemble de l'échantillon, l'estimation d'un score global ne fait apparaître aucune différence significative entre les deux groupes (Murat et Rocher, ce numéro). En réalité, une étude récente a montré que si l'on prend en compte la distinction entre lecture et compréhension (cf. encadré 1), il s'avère que les femmes sont plus performantes dans la maîtrise des mécanismes de base de

Encadré 6 (suite)

En complément des tests statistiques, Roussos et Stout (1996) ont présenté une grille de classification empirique permettant d'apprécier l'ampleur des FDI :

- 1 (FDI négligeable) : le test est significatif à 5 % et la valeur absolue de $\hat{\beta}_U$ est inférieure à 0,059
- 2 (FDI modéré) : le test est significatif à 5 % et la valeur absolue de $\hat{\beta}_U$ est comprise entre 0,059 et 0,088
- 3 (FDI important) : le test est significatif à 5 % et la valeur absolue de $\hat{\beta}_U$ est supérieure ou égale à 0,088

Cette procédure est intéressante dans la mesure où elle permet de tester le fonctionnement différentiel d'un groupe d'items (« bundle ») et non pas uniquement d'un seul item.

Elle permet également de tester les FDI non-uniformes ou « croisés », c'est-à-dire qui diffèrent selon le niveau de compétence q. Les FDI « croisés » ne sont pas analysés ici, le faible nombre d'items ne permettant pas de les interpréter convenablement.

Tableau 5
Analyses des fonctionnements différentiels des items (FDI) selon le sexe

		Indices de fonctionnement différentiel d'item (FDI)						
	Item	Taux de réussite			Mantel-Haenszel		Sibtest	
		Hommes	Femmes	Écart	MH ($-\chi^2$)	Δ	$\hat{\beta}_U$	classe
Récit	h1	45,4	42,0	3,4	4,7*	- 0,3	- 0,050*	1
	h2	79,3	77,8	1,5	0,5	- 0,1	- 0,005	1
	h3	88,5	81,3	7,2	35,3**	1,0	0,038*	
Graphique	h4	85,1	81,3	3,8	0,9	0,2	- 0,003	
	h5	90,1	87,6	2,5	0,0	0,0	- 0,029*	1
	h6	77,6	68,1	9,4	47,1**	0,9	0,064**	2
Texte expositif	h8	85,2	82,9	2,3	0,0	0,0	- 0,024	
	h9	87,1	85,3	1,9	1,8	- 0,2	- 0,027*	1
	h10	42,7	34,5	8,2	22,7**	0,6	0,031**	1
	h11	34,4	29,3	5,1	3,5	0,2	- 0,018	
	h12	77,4	71,7	5,7	2,9	0,2	- 0,007	
Texte théorique	h13	60,6	62,5	- 1,9	42,1**	- 0,8	- 0,092**	3
	h14	77,5	79,3	- 1,9	36,7**	- 0,9	- 0,071**	2
	h15	61,3	61,0	0,3	2,7	- 0,2	- 0,032**	1
Cartes spatiales	h16	55,6	49,3	6,3	9,7**	0,4	0,021	
	h17	81,8	78,5	3,4	0,1	0,0	- 0,004	
	h18	65,6	55,5	10,0	39,2**	0,7	0,057**	1
	h19	93,5	91,5	2,0	0,6	- 0,2	- 0,006	
	h20	69,2	58,9	10,3	34,1**	0,7	0,038**	1
	h21	84,4	74,5	9,9	50,2**	1,1	0,055**	1
	h22	75,3	65,3	10,0	30,9**	0,7	0,031*	1
	Moyenne	72,3	67,5	4,7				

Lecture : l'item h3 est réussi par 88,5 % des hommes et 81,3 % des femmes. Les deux procédures pour tester le fonctionnement différentiel sont significatives (* : P-value < 0,05, ** : P-value < 0,01). Le signe positif de Δ et béta indique que le « biais » est en faveur des hommes. Les deux procédures diffèrent sur l'ampleur du biais : le coefficient delta de 1 indique un fort biais, mais selon la procédure Sibtest, l'item est dans la classe 1, ce qui indique un biais négligeable (cf. encadré 6).

Champ : individus ayant passé le module Haut de l'enquête IVQ.

Source : calculs à partir des données de l'enquête Information et Vie Quotidienne 2004, Insee.

la lecture alors qu'au niveau de la compréhension, cette différence n'apparaît plus.

Au-delà de cette différence de réussite globale, les deux analyses sont assez cohérentes et montrent que les items comportent des fonctionnements différentiels selon le sexe. Conformément aux données de la littérature, les items de l'épreuve des cartes spatiales sont biaisés en faveur des hommes ainsi que les items du graphique (Linn et Petersen, 1985 ; Stumpf et Eliot, 1995, 1999 ; Voyer, Voyer et Bryden, 1995 ; Voyer, Nolan et Voyer, 2000). Les auteurs ont tenté d'expliquer ce résultat et différentes propositions ont été faites, notamment en termes d'expérience perceptivo-motrice qui serait plus importante chez les hommes (cf. par exemple Rahman et Wilson, 2003), la société y contribuant largement avec notamment l'étalement de jouets pour les garçons mobilisant des compétences spatiales, par exemple. Cet écart devrait se creuser dans les années à venir si on prend en compte la familiarité qu'ont les jeunes adolescents avec les jeux virtuels, ce qui est moins le cas des adolescentes.

Quelques items cependant donnent un avantage aux femmes (h13, h14 et h1), qui sont des items qui ciblent le traitement des marques linguistiques particulières du texte et la production d'inférences, à partir des textes ne comportant que des informations verbales. Ces items concernent donc la dimension verbale.

Ainsi même si nous restons prudents (trop peu d'items), l'étude des fonctionnements différentiels a tendance à révéler une spécialisation des traitements : les épreuves verbales notamment le texte expositif et le texte théorique sont mieux réussies par les femmes, tandis que les épreuves visuo-spatiales sont mieux réussies par les hommes (cf. tableau 6).

* *
*

Nous pouvons insister sur le fait que le module Haut ainsi élaboré permet d'évaluer la compréhension à l'aide de divers formats présents dans le quotidien de la population (journaux quotidiens, magazines scientifiques, indications routières, lecture de cartes spatiales etc.). Le score maximal théorique était de 21, toutes épreuves associées : 3 questions pour le récit, 3 pour le graphique, 6 pour le texte expositif, 3 pour le texte scientifique et 6 pour le plan de village. Parmi ces épreuves, trois ne comportaient que des informations verbales (épreuves 1, 3 et 4), tandis que les deux autres associaient également des informations visuo-spatiales (épreuves 2 et 5). L'objectif de cet article était de présenter la démarche théorique qui a sous-tendu la conception des cinq épreuves du module Haut, et de donner les premiers résultats concernant notamment l'étude de la dimensionnalité du module et les effets liés au sexe des participants.

C'est la première fois que ce module dans cette conception, a été soumis à un échantillon de la population vivant en France, dans le cadre de l'enquête *Information et Vie quotidienne* en 2004. Ce sont 7 389 individus âgés entre 18 et 65 ans, hommes et femmes, qui ont été sélectionnés pour la passation de ce module à l'issue du module d'Orientation. Toutes ces personnes sont *supposées* ne pas avoir de grandes difficultés avec le langage écrit. En effet, à l'issue du module d'Orientation, elles ont été orientées vers le module Haut, en grande partie, parce qu'elles ne présentaient pas de difficultés dans la maîtrise de la lecture des mots.

L'approche théorique exposée est celle des travaux issus du champ de la psychologie cognitive du langage. La compréhension est une activité complexe dont le but est de construire une représentation mentale cohérente du contenu du texte. Cette activité sollicite un certain nombre de processus allant du traitement des mots, à la phrase puis à l'organisation textuelle. La représentation mentale la plus riche est l'élaboration

Tableau 6
Analyses des fonctionnements différentiels de groupes d'items selon le sexe

Supports	Items	$\hat{\beta}_U$	Classe
Récit	h1 h2 h3	- 0,011	
Graphique	h4 h5 h6	0,052*	1
Texte expositif	h8 h9 h10 h11 h12 h13	- 0,151**	3
Texte théorique	h14 h15 h16	- 0,070**	2
Cartes spatiales	h17 h18 h19 h20 h21 h22	0,278**	3

Lecture : les items du texte expositif forment un ensemble présentant un fonctionnement différentiel significatif (* : P-value < 0,05, ** : P-value < 0,01), important (classe 3) et en faveur des femmes ($\hat{\beta}_U$ négatif) (cf. encadré 6).

Champ : individus ayant passé le module Haut de l'enquête IVQ.

Source : calculs à partir des données de l'enquête Information et Vie Quotidienne 2004, Insee.

d'un modèle de situation, où les relations entre maîtrise du langage et mobilisation des connaissances sont cruciales. Ainsi, comprendre un texte nécessite de mettre en relation le langage et la cognition.

L'analyse descriptive du module montre tout d'abord que le niveau est relativement facile avec 70 % de réussite moyenne sur l'ensemble des questions tous participants confondus. Les items sont plus ou moins bien réussis révélant la complexité des items, et l'homogénéité interne - donnant le niveau d'inter-corrélations entre les items - est relativement moyenne.

Pour évaluer le caractère dimensionnel de l'épreuve, nous avons réalisé plusieurs types d'analyses factorielles. La première, à visée exploratoire, fait apparaître quatre facteurs, l'un faisant clairement référence au traitement des informations spatiales, l'autre aux informations contenues dans le graphique, et les deux autres font référence au traitement des informations linguistiques contenues dans les trois épreuves verbales. Il apparaît que la distinction entre ces deux derniers facteurs bien qu'elle ne soit pas tranchée, renverrait à la distinction entre la mobilisation des connaissances lexicales et syntaxico-sémantiques d'une part, et la mise en relation entre les informations verbales et la mobilisation des connaissances d'autre part. Cette analyse exploratoire est une première tentative qui permet d'étudier les dimensions qui émergent à partir du module. Cependant, en raison du nombre limité d'items, nous resterons prudents sur la validité de cette analyse. Une seconde analyse à l'aide d'une méthode confirmatoire (les quatre dimensions précédentes ont été testées une à une, à chaque fois avec les items restants) montre que seuls deux facteurs demeurent distincts : il s'agit d'un facteur visuel et spatial (cartes spatiales et graphique) et d'un facteur strictement verbal. L'établissement de ces deux dimensions serait en faveur d'une spécialisation du traitement des informations selon leur nature, verbale ou visuo-spatiale.

La question de la spécialisation paraît confirmée par l'étude des fonctionnements différentiels liés au sexe des participants. De manière générale, le module est mieux réussi par les hommes et ce, pour la grande majorité des items (en moyenne 72,3 % pour les hommes contre 67,5 % pour les femmes). Lorsque cette question est approfondie, en étudiant les « biais » (1) liés au sexe - c'est-à-dire à niveau de performance égale entre les deux groupes, on étudie la réus-

site de chaque item selon le sexe - on observe que la majorité des items sont « biaisés » en faveur des hommes. Un résultat classique qui se dégage est le « biais » systématique sur les épreuves visuo-spatiales en faveur des hommes (plan du village et graphique), expliquée dans la littérature en termes d'expérience perceptivo-motrice plus importante dès le plus jeune âge chez les garçons. Nous observons des « biais » en faveur des femmes sur quelques items qui sollicitent des informations de nature verbale, mais ces « biais » restent peu nombreux.

D'autres fonctionnements différentiels seraient intéressants d'étudier. En effet, la question de la langue maternelle pourrait être creusée pour étudier la répercussion à long-terme du fait que certains individus ne sont pas de langue maternelle française. Dans le même registre, il serait alors souhaitable d'étudier l'influence du temps d'exposition à la langue française chez les participants arrivés tardivement en France. Le module Biographique en l'état nous permet de considérer le premier aspect, mais le second devrait être affiné pour la prochaine enquête *IVQ* 2011. D'autres fonctionnements différentiels devront faire l'objet d'études : le rôle du diplôme de la mère, du père, de la personne enquêtée.

Enfin la question des profils de bons et de faibles compreneurs sur le module Haut devrait être considérée pour apprécier quels sont les processus qui sont défaillants chez les faibles compreneurs. Nous faisons l'hypothèse que les différences entre les bons et les faibles sont plus marquées sur les items sollicitant le traitement de la continuité référentielle (par exemple sur l'épreuve 1 spécifiquement construite pour mesurer cet aspect) et sur les items qui nécessitent de mettre en relation le traitement des informations verbales et la mobilisation des connaissances du monde : les inférences. En effet, les travaux provenant de la littérature sur les différences individuelles chez l'enfant montrent que les enfants faibles compreneurs (à différents âges étudiés) ont des déficits dans la prise en charge de la double cohérence locale et globale (pour les recherches anglo-saxonnes, cf. notamment Cain et Oakhill, 2006 ; Yuill et Oakhill, 1991 ; et en français Megherbi et Ehrlich, 2004 ; 2005 ; Megherbi *et al.*, 2006). Cependant, compte tenu du faible nombre

1. Le terme « biais » utilisé ici, qui ne correspond pas à la signification usuelle en statistique ou en économétrie, se réfère au fonctionnement différentiel des items (cf. encadré 6).

d'items sur l'ensemble du module, il nous est difficile de caractériser les profils de bons et de faibles compreneurs à ce stade.

Pour la prochaine enquête, nous souhaiterions augmenter le nombre d'items en créant une autre série de textes selon la procédure classique d'utilisation de « cahiers tournants ». Nous pourrions conforter et affiner les différentes hypothèses développées dans cet article sur la dimensionnalité du module et sur les fonction-

nements différentiels. Pour la seconde série, il nous semble important d'introduire des types de documents qui n'ont pas pu être pris en compte de manière à évaluer la compréhension sur plusieurs structures de textes comme par exemple les modes d'emploi, les recettes de cuisine, les règles de jeu, une ordonnance etc. Cette procédure nous permettra d'une part de faire des comparaisons dans le temps sur la série 1, et de réaliser des analyses plus assurées en tenant compte des deux séries. □

BIBLIOGRAPHIE

Baddeley A.D. (1986), *Working Memory*, Oxford University Press, Oxford.

Baddeley A.D. et Logie R.H. (1999), « Working Memory : The multiple-Component Model », in A. Miyake et P. Shah (éds.), *Models of working Memory : Mechanisms of Active Maintenance and Executive Control*, pp. 28-61, Cambridge University Press, Cambridge, UK.

Cain K. et Oakhill J. (2006), « Profiles of Children with Specific Reading Comprehension Difficulties », *British Journal of Educational Psychology*, vol. 76, n° 4, pp. 683-696.

Caplan D. et Waters G.S. (1999), « Verbal Working Memory and Sentence Comprehension », *Behavioral and Brain Sciences*, vol. 22, n° 1, pp. 77-94.

Carpenter P.A., Miyake A. et Just M.A. (1994), « Working Memory Constraints in Comprehension : Evidence from Individual Differences, Aphasia and Aging », in Gernsbacher M.A. (éd.), *Handbook of Psycholinguistics*, Academic Press, San Diego, CA, pp. 1075-1122.

Choi J. et Sylverman I. (2003), « Processes Underlying Sex Differences in Route-Learning Strategies in Children and in Adolescents », *Personality and Individual Differences*, vol. 34, n° 7, pp. 1153-1166.

Dabbs J.M., Chang E.L., Strong R.A. et Milun R. (1998), « Spatial Ability, Navigation Strategy and Geographic Knowledge Among Men and Women », *Evolution and Human Behavior*, vol. 19, n° 2, pp. 89-98.

de La Haye F., Gombert J.-E., Rivière J.-P. et Rocher T. (2007), « Les évaluations en lecture dans le cadre de la journée d'appel de préparation à la défense. Année 2006 », *Note d'information*, n° 07.25, MEN, Depp : <http://media.education.gouv.fr/file/25/6/5256.pdf>.

Dickes P., Tournois J., Flieller A. et Kop J.-L. (1994), *La psychométrie. Théorie et pratique de la mesure en psychologie*, Presses Universitaires de France, Paris.

Ehrlich M.-F. (1994), *Mémoire et compréhension du langage*, Presses Universitaires de Lille, Lille.

Ehrlich M.-F., Tardieu H. et Cavazza M. (éds.) (1993), *Les modèles mentaux, approche cognitive des représentations*, Masson, Paris.

Gathercole S.E. et Baddeley A.D. (1993), *Working Memory and Language*, Erlbaum, Hillsdale, N.J.

Gernsbacher M.A. (1990), *Language Comprehension as a Structure Building*, Erlbaum, Hillsdale, N.J.

Gernsbacher M.A. (éd.) (1994), *Handbook of Psycholinguistics*, Academic Press, San Diego, CA.

Gernsbacher M.A. et Foertsch J.A. (1999), « Three Models of Discourse Comprehension », dans S. Garrod et M. Pickering (éds.), *Language Processing*, Psychology Press, East Sussex, UK, pp. 283-300.

Gernsbacher M. A., Varner K. R. et Faust M. E. (1990), « Investigating Differences in General

- Comprehension Skill », *Journal of Experimental Psychology : Learning, Memory, and Cognition*, vol. 16, n° 3, pp. 430-445.
- Glenberg A.M., Kruley P. et Langston W.E. (1994)**, « Analogical Processes in Comprehension : Simulation of a Mental Model », in M.A. Gernsbacher (éd.), *Handbook of Psycholinguistics*, Academic Press, San Diego, CA, pp. 609-640.
- Goldstein H., Bonnet G. et Rocher T. (2007)**, « Multilevel Structural Equation Models for the Analysis of Comparative Data on Educational Performance », *Journal of Educational and Behavioral Statistics*, vol. 32, n° 3, pp. 252-286.
- Gough P.B., Hoover W.A. et Peterson C.L. (1996)**, « Some Observations on a Simple View of Reading », dans C. Cornoldi et J. Oakhill (éds.), *Reading Comprehension Difficulties : Processes and Intervention*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 1-13.
- Gough P.B. et Tunmer W.E. (1986)**, « Decoding, Reading and Reading Disability », *Remedial and Special Education*, vol. 7, n° 1, pp. 6-10.
- Graesser A.C., Gernsbacher M.A. et Goldman S.J. (éds.) (2002)**, *Handbook of Discourse Processes*, Erlbaum, Mahwah, N.J.
- Gyselinck V., Cornoldi C., Dubois V., de Beni R. et Ehrlich M.-F. (2002)**, « Visuospatial Memory and Phonological Loop in Learning From Multimedia », *Applied Cognitive Psychology*, vol. 16, n° 6, pp. 665-685.
- Hattie J. (1985)**, Methodology Review, « Assessing Unidimensionality of Tests and Items », *Applied Psychological Measurement*, vol. 9, n° 2, pp. 139-164.
- Haenggi D., Kintsch W. et Gernsbacher M.A. (1995)**, « Spatial Situation Models and Text Comprehension », *Discourse Processes*, vol. 19, n° 2, pp. 173-199.
- Holland P.W. et Thayer D.T. (1988)**, « Differential Item Performance and the Mantel-Haenszel Procedure », dans P. W. Holland et H. Braun (éds.), *Test Validity*, Lawrence Erlbaum Associates, Hillsdale, N.J., pp. 129-145.
- Hoover W.A. et Gough P.B. (1990)**, « The Simple View of Reading », *Reading and Writing*, vol. 2, n° 2, pp. 127-160.
- Juhel J. (éd.) (1999)**, *Le modèle de la réponse à l'item*, numéro spécial de Psychologie et Psychométrie, vol. 20, n° 2-3.
- Kintsch W. (1988)**, « The Role of Knowledge in Discourse Comprehension : A Construction-Integration Model », *Psychological Review*, vol. 95, n° 2, pp. 163-182.
- Kintsch W. (1994)**, « The Psychology of Discourse Processing », dans M.A. Gernsbacher (éd.), *Handbook of Psycholinguistics*, Academic Press, San Diego, CA, pp. 721-739.
- Kintsch W. (1998)**, *Comprehension : A Paradigm for Cognition*, Cambridge University Press, New York.
- Kintsch W. et van Dijk T.A. (1978)**, « Toward a Model of Text Comprehension and Production », *Psychological Review*, vol. 85, n° 5, pp. 363-394.
- Kruley P., Sciama S.C. et Glenberg A.M. (1994)**, « On-Line Processing of Textual Illustrations in the Visuospatial Sketchpad : Evidence From Dual-Task Studies », *Memory and Cognition*, vol. 22, n° 3, pp. 261-272.
- Landi N. et Perfetti C.A. (2007)**, « An Electrophysiological Investigation of Semantic and Phonological Processing in Skilled and Less-Skilled Comprehenders », *Brain and Language*, vol. 102, n° 1, pp. 30-45.
- Laveault D. et Grégoire J. (2002)**, *Introduction aux théories des tests en éducation et en psychologie*, 2e édition, De Boeck-Université, Bruxelles.
- Le Ny J.-F. (2005)**, *Comment l'esprit produit du sens*, Odile Jacob, Paris.
- Linn M. et Petersen A. (1985)**, « Emergence and Characterization of Sex Differences in Spatial Ability : A Metaanalysis », *Child Development*, vol. 56, n° 6, pp. 1479-1498.
- Logie R.H. (1995)**, *Visuo-Spatial Working Memory*, Hove, UK : Lawrence Erlbaum Associates.
- Lord F.M. et Novick M.R. (1968)**, *Statistical Theories of Mental Test Scores*, Addison-Wesley Publishing Company, Reading MA.
- Mantel N. et Haenszel W. (1959)**, « Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease », *Journal of the National Cancer Institute*, vol. 22, n° 4, pp. 719-748.

- Megherbi H. et Ehrlich M.-F. (2005)**, « Language Impairment in Less Skilled Comprehenders : The On-line Processing of Anaphoric Pronouns in a Listening Situation », *Reading and Writing*, vol. 18, n° 7-9, pp. 715-753.
- Megherbi H., Seigneuric A. et Ehrlich M.-F. (2006)**, « Reading Comprehension in French 1st and 2nd Grade Children : Contribution of Decoding and Language Comprehension », *European Journal of Psychology of Education*, vol. 21, n° 2, pp. 135-147.
- Millsap R.E. et Everson H.T. (1993)**, « Methodology Review : Statistical Approaches for Assessing Measurement Bias », *Applied Psychological Measurement*, vol. 17, n° 4, pp. 297-334.
- Nandakumar R. (1994)**, « Assessing Dimensionality of a Set of Item Responses-Comparison of Different Approaches », *Journal of Educational Measurement*, vol. 31, n° 1, pp. 17-35.
- Perfetti C.A. (1994)**, « Psycholinguistics and Reading Ability », dans M.A. Gernsbacher (éd.), *Handbook of Psycholinguistics*, Academic Press, San Diego, pp. 849-894.
- Perfetti C.A. et Hart L. (2001)**, « The Lexical Bases of Comprehension Skill », in D.S. Gorfein (éd.), *On the Consequences of Meaning Selection : Perspectives on Resolving Lexical Ambiguity. Decade of Behavior*, American Psychological Association, Washington, pp. 67-86.
- Perfetti C.A. (2007)**, « Reading Ability : Lexical Quality to Comprehension », *Scientific Studies of Reading*, vol. 11, n° 4, pp. 357-383.
- Rahman Q. et Wilson G.D. (2003)**, « Large Sexual-Orientation-Related Differences in Performance on Mental Rotation and Judgments of Line Orientation Tasks », *Neuropsychology*, vol. 17, n° 1, pp. 25-31.
- Rocher T. (2003)**, « La méthodologie des évaluations internationales de compétences », *Psychologie et Psychométrie*, vol. 24, n° 2/3, Éditions EAP, Paris.
- Roussos L.A. et Stout W.F. (1996)**, « Simulation Studies of the Effects of Small Sample Size and Studied Item Parameters on SIBTEST and Mantel-Haenszel Type I Error Performance », *Journal of Educational Measurement*, vol. 33, n° 2, pp. 215-230.
- Shealy R. et Stout W.F. (1993)**, « A Model-Based Standardization Approach that Separates True Bias/DIF from Group Ability Differences and Detects Test Bias/DTF as Well as Item Bias/DIF », *Psychometrika*, vol. 58, n° 2, pp. 159-194.
- Stout W. (1987)**, « A Nonparametric Approach for Assessing the Latent Trait Unidimensionality », *Psychometrika*, vol. 52, n° 4, pp. 589-617.
- Stout W., Froelich A.G. et Gao F. (2001)**, « Using Resampling Methods to Produce an Improved DIMTEST Procedure », dans A. Boomsma, M.A.J. van Duijn et T.A.B. Snijders (éds.), *Essays on Item Response Theory*, Springer-Verlag, New York, pp. 357-375.
- Stumpf H. et Eliot J. (1995)**, « Gender-Related Differences in Spatial Ability and the k Factor of General Spatial Ability in a Population of Academically Talented Students », *Personality and Individual Differences*, vol. 19, n° 1, pp. 33-45.
- Stumpf H. et Eliot J. (1999)**, « A Structural Analysis of Visual Spatial Ability in Academically Talented Students », *Learning and Individual Differences*, vol. 11, n° 2, pp. 137-151.
- Tate R. (2003)**, « A Comparison of Selected Empirical Methods for Assessing the Structure of Responses to Test Items », *Applied Psychological Measurement*, vol. 27, n° 3, pp. 159-203.
- Traxler M. et Gernsbacher M.A. (2006)**, *Handbook of Psycholinguistics*, 2e édition, Elsevier, New York.
- Voyer D, Nolan C. et Voyer S. (2000)**, « The Relation Between Experience and Spatial Performance in Men and Women », *Sex Roles*, vol. 43, n° 11-12, pp. 891-915.
- Voyer D., Voyer S. et Bryden M. (1995)**, « Magnitude of Sex Differences in Spatial Abilities : A Meta-Analysis and Consideration of Critical Variables », *Psychological Bulletin*, vol. 117, n° 2, pp. 250-270.
- Vrignaud P. (2002)**, « Les biais de mesure : savoir les identifier pour y remédier », *Bulletin de Psychologie*, vol. 55, n° 6, pp. 625-634.
- Yuill N. et Oakhill J.V. (1991)**, *Children's Problems in Text Comprehension*, Cambridge University Press, Cambridge.

**TABLEAU DE L'ANALYSE FACTORIELLE AVEC ROTATION OBLIQUE (PROMAX)
SUR DEUX FACTEURS**

A - Corrélations semi-partielles entre les items et les facteurs d'une analyse factorielle en rotation oblique (Promax) avec deux facteurs

	Items	Facteur 1	Facteur 2
Récit	h1	0,43	0,15
	h2	0,20	0,09
	h3	0,35	0,07
Graphique	h4	0,17	0,28
	h5	0,19	0,31
	h6	0,22	0,09
Texte expositif	h8	0,36	0,03
	h9	0,42	0,08
	h10	0,33	0,00
	h11	0,35	0,00
	h12	0,54	- 0,05
Texte théorique	h13	0,42	0,03
	h14	0,31	0,12
	h15	0,13	0,10
Cartes spatiales	h16	0,21	0,09
	h17	0,01	0,42
	h18	- 0,03	0,41
	h19	0,12	0,53
	h20	0,03	0,45
	h21	0,12	0,39
h22	- 0,07	0,65	

B - Corrélations entre les facteurs

	Facteur 2
Facteur 1	0,61

*Lecture : les corrélations supérieures à 0,20 sont en gras.
Le facteur 2 comprend les items portant sur les cartes spatiales et les graphiques (sauf h6). Il est corrélé à 0,61 avec le facteur 1.
Champ : individus ayant passé le module Haut de l'enquête IVQ 2004.
Source : calculs à partir des données de l'enquête Information et Vie Quotidienne 2004, Insee.*