

Chapitre 9

L'approche différentielle de l'intelligence

Jacques Lautrey

En 1921, les éditeurs du « Journal of Educational Psychology » demandèrent à 17 chercheurs reconnus dans le domaine de la psychologie de donner une définition de l'intelligence. La variété des réponses, dont quelques exemples sont donnés dans l'encadré 9.1, montrait l'absence de consensus sur ce point.

Encadré 9.1

- « La capacité à apprendre ou à profiter de l'expérience » (W.F. Dearborn)
- La capacité à acquérir des connaissances ainsi que l'usage des connaissances acquises (B.A. Henmon).
- L'aptitude à s'adapter de façon adéquate aux situations relativement nouvelles rencontrées dans la vie (R. Pintner)
- Un individu est intelligent en proportion de sa capacité à élaborer une pensée abstraite (L.M. Terman)
- (a) La capacité à inhiber une adaptation instinctive, (b) la capacité à redéfinir l'adaptation inhibée à la lumière des essais et erreurs d'expériences imaginaires (c) La capacité à traduire de façon délibérée cette modification de l'adaptation dans un comportement qui soit à l'avantage de l'individu en tant qu'animal social » (L.L. Thurstone).

Les réponses à la même question, posée 65 ans plus tard à 24 chercheurs considérés comme experts en psychologie l'intelligence, ont montré qu'il n'y avait guère plus de consensus en 1986 qu'en 1921 (Sternberg et Detterman, 1986).

Néanmoins, lorsque les attributs du concept d'intelligence mentionnés à ces deux époques sont comparés, comme l'ont fait Sternberg et Berg (1986), il ressort que quelques uns d'entre eux sont cités de façon récurrente, notamment la capacité à s'adapter à son environnement, à résoudre des problèmes nouveaux, à apprendre, à abstraire. Il y a aussi une certaine permanence dans les points de désaccord, notamment sur la question de savoir s'il y a une intelligence ou s'il faut en distinguer plusieurs. Enfin, certains des attributs cités en 1986 ne l'étaient jamais en 1921, notamment la capacité à contrôler ses propres processus cognitifs (les processus par lesquels un individu contrôle ses propres processus cognitifs sont appelés métacognitifs, ou exécutifs). Dans la période récente, l'intelligence est aussi plus souvent définie comme relative à un contexte culturel : « the purposive adaptation to, selection of, and shaping of real-world environments relevant to one's life and abilities » (Sternberg, 1988, p. 65). Dans cette définition, difficile à traduire de façon exacte en français, il faut relever l'extension de la notion d'adaptation : celle-ci n'est en effet pas forcément l'adaptation de l'individu à un environnement qui serait considéré comme un donné intangible, elle peut aussi consister, pour cet individu, à sélectionner un autre environnement qui convienne mieux à ses dispositions, ou encore à transformer son environnement pour l'adapter à ses propres capacités, valeurs, ou intérêts. Il faut aussi relever l'introduction de la notion de pertinence attachée à l'environnement dans lequel vit un individu : l'intelligence ne peut être appréciée qu'au regard de ce qui est pertinent, de ce qui a du sens, du point de vue de la vie de chacun dans le monde réel auquel il est confronté. Autrement dit, même si une définition universelle de l'intelligence peut être recherchée, il faut avoir constamment à l'esprit que ses manifestations sont toujours relatives au contexte culturel dans lequel les individus se développent (voir la partie consacrée aux théories socio-culturelles du développement cognitif dans le chapitre 2). En conséquence, il faut préciser d'entrée de jeu que les tests d'intelligence dont il sera question dans ce chapitre ont été conçus pour l'évaluation du développement et du fonctionnement cognitifs de personnes qui ont été élevées dans le contexte culturel de la société occidentale et n'ont de sens que lorsqu'ils sont utilisés dans ce contexte.

Ceci étant précisé, que conclure de la difficulté à parvenir à un consensus sur la définition de l'intelligence ? Une conclusion possible est que cette notion est trop vague, trop générale, pour donner prise à une approche scientifique. Une autre conclusion possible – la nôtre - est que l'intelligence est une fonction dont la finalité peut être atteinte de multiples façons. La finalité est l'adaptation du comportement aux situations nouvelles et cette finalité est atteinte grâce à une propriété essentielle du système cognitif, celle de s'auto-modifier en fonction de

l'expérience. Cette capacité est sous-tendue par la plasticité du cerveau, mais elle se manifeste à tous les niveaux d'organisation du système cognitif - celui des connexions neuronales, celui des représentations, et celui des comportements. C'est cette propriété d'auto-modification qui donne son unité à la notion d'intelligence. La multiplicité des mécanismes par lesquels elle peut s'opérer est par contre à l'origine de la multiplicité des facettes de l'intelligence.

La recherche des lois générales auxquelles obéit l'intelligence peut s'appuyer sur les différentes sortes de variation du comportement qui peuvent être observées dans les tâches intellectuelles : ses variations en fonction de la situation (celles que provoque la psychologie expérimentale pour étudier la résolution de problèmes), en fonction de l'âge (celles qu'étudie la psychologie développementale pour comprendre l'ontogenèse* de l'intelligence) ou en fonction des individus (celles qu'étudie la psychologie différentielle pour dégager les dimensions de l'intelligence). C'est cette dernière approche de l'intelligence, celle de la psychologie différentielle, qui fait l'objet du présent chapitre.

L'analyse des différences individuelles dans les performances à des tâches intellectuelles a été utilisée pour mettre au point des instruments d'évaluation de l'intelligence, les tests, mais aussi pour mieux comprendre la structure de l'intelligence. Curieusement, ces deux courants de recherche, le premier plus tourné vers les applications et le second plus tourné vers la recherche fondamentale, sont restés assez longtemps cloisonnés. Toutefois, dans l'un et l'autre, et de façon relativement indépendante, on est passé, au cours du siècle qui nous sépare de l'invention du premier test d'intelligence, de modèles unidimensionnels à des modèles multidimensionnels de l'intelligence. L'analyse de cette évolution est la ligne directrice autour de laquelle ce chapitre est organisé.

1. Les modèles unidimensionnels de l'intelligence : Binet et Spearman

Les premières tentatives d'évaluation objective de l'intelligence, celle Binet et celle de Spearman, s'appuyaient sur des méthodes fort différentes mais opérationnalisèrent toutes deux une conception unidimensionnelle de l'intelligence.

1.1. Le Binet-Simon

La construction de l'échelle métrique imaginée par Alfred Binet (1857-1911) pour « mesurer » l'intelligence s'appuyait sur quelques idées directrices qui ont permis à Binet de

sortir les recherches sur les tests mentaux de l'impasse où Francis Galton et James McKeen Cattell les avaient engagées (voir le chapitre sur l'histoire de la psychologie différentielle).

Contrairement à Galton et Cattell, qui pensaient pouvoir évaluer l'intelligence en mesurant l'efficacité des processus sensoriels – les seuls processus psychologiques que l'on savait mesurer à l'époque – Binet était persuadé que la mesure devait porter prioritairement sur les activités mentales supérieures. Il a varié dans la liste de ces activités mentales mais toutes se situaient à un niveau élevé d'intégration des informations: la mémoire, l'imagination, l'attention, la faculté de comprendre (Binet et Henri, 1895) ; le bon sens, le sens pratique, l'initiative, la faculté de s'adapter (Binet et Simon, 1905) ; la compréhension, l'invention, la direction, la censure (Binet, 1911a). Mais si le choix des processus supérieurs résolvait un problème, il en créait aussitôt un autre car si on savait, à l'époque, mesurer les seuils sensoriels, on ne voyait par contre pas du tout comment mesurer objectivement des processus mentaux aussi généraux que ceux cités plus haut.

C'est une autre idée nouvelle qui a permis à Binet de surmonter cet obstacle, celle de prendre appui sur les différences entre les individus pour fonder la mesure. Ce sont d'abord les comparaisons entre des individus considérés comme normalement intelligents et des individus considérés comme retardés mentaux, qui ont permis de définir les niveaux d'intelligence distingués dans la première version de l'échelle métrique (Binet et Simon, 1905). Ce sont ensuite les comparaisons entre enfants d'âges différents qui ont permis d'affiner ces niveaux d'intelligence dans la seconde version de l'échelle (Binet et Simon, 1908). L'échelle des âges auxquels les différents problèmes ou items étaient normalement réussis permettait de définir le « niveau mental » de ceux-ci. Par exemple, Binet avait observé qu'à 3 ans, la plupart des enfants peuvent répéter deux chiffres, qu'à 7 ans ils peuvent en répéter cinq et que vers 12 ans ils peuvent en répéter 7. Répéter cinq chiffres est donc une tâche caractéristique de 7 ans ; en répéter 7 est caractéristique de 12 ans. Plus précisément, pour être caractéristique d'un âge, par exemple 12 ans, un item doit être encore échoué par une majorité des enfants de 11 ans, il doit être réussi par la plupart des enfants de 13 ans et par environ 50% des enfants de 12 ans. Le travail de Binet et de Simon a donc consisté à trouver, pour chacun des processus supérieurs jugés pertinents (jugement, raisonnement, mémoire, etc.), des items caractéristiques de chaque âge. C'est ainsi qu'ils ont élaboré l'échelle de 1908, dont le contenu a été enrichi dans la troisième et dernière version de cette échelle (Binet, 1911b). En cherchant à varier autant que possible les processus auxquels les items faisaient appel, ils ont retenu pour chacun des âges de 3 à 16 ans, quatre ou cinq items caractéristiques de cet âge (une liste des items qui constituaient l'échelle de 1908 est donnée

dans le chapitre sur l'histoire de la psychologie différentielle). La trouvaille consistait à utiliser l'échelle des âges pour définir un ordre développemental des items et établir, du même coup, l'ordre de complexité cognitive de ces items : un item caractéristique de 12 ans correspond à une étape plus avancée du développement de l'intelligence qu'un item caractéristique de 10 ans et, de cela, on peut inférer qu'il se situe à un niveau de complexité cognitive plus élevé, même sans savoir exactement à quoi tient cette différence de complexité cognitive. L'ordre des niveaux de complexité des items permettait à son tour d'ordonner les sujets, y compris ceux de même âge chronologique, en fonction du « niveau mental » des items réussis. Ce sont donc les différences entre individus qui fournissaient le principe de la « mesure » de l'intelligence et Binet était parfaitement conscient que cette mesure se situait au niveau ordinal (pour la définition des niveaux de mesure, voir le chapitre sur la méthode des tests) : « Le mot mesure n'est pas pris ici au sens mathématique : il n'indique pas le nombre de fois qu'une quantité est contenue dans une autre. L'idée de mesure se ramène pour nous à celle de classement hiérarchique » (Binet, 1911a, p. 135).

La troisième idée qui a joué un rôle important dans la genèse de l'échelle métrique est celle qui consistait à considérer l'intelligence comme la résultante des différentes activités mentales évaluées dans l'échelle. C'est ce qui, dans l'esprit de Binet, justifiait l'addition des mois d'âge mental attribués aux différents items : « un test particulier, isolé de tout le reste, ne vaut pas grand chose...ce qui donne une force démonstrative, c'est un faisceau de tests, un ensemble dont on conserve la physionomie moyenne ». Chaque item, quel qu'il soit, était donc crédité d'un certain nombre de mois (s'il y avait, par exemple, 4 items caractéristiques d'un âge donné, chacun rapportait 12 mois / 4 = 3 mois) et l'âge mental était calculé en ajoutant tous les mois correspondant aux items réussis. Le calcul de l'âge mental se faisait donc en effectuant la somme non pondérée des items réussis, ou plus exactement des nombres de mois dont chacun était crédité.

C'est William Stern, un psychologue allemand, qui suggéra en 1912 de pondérer l'âge mental par l'âge chronologique en faisant le quotient du premier par le second. Ainsi, un enfant qui a un âge mental de 12 ans et un âge chronologique de 10 ans a un quotient de $12 / 10 = 1,2$. L'avantage de ce quotient est de pondérer l'écart (ici une avance de 2 ans) par l'âge chronologique (ici 10 ans). Cette pondération fait qu'un écart donné n'a pas la même signification aux différents âges chronologiques : ce même écart de 2 ans chez un enfant de 4 ans se traduirait par un quotient de $6 / 4 = 1,5$. Calculé ainsi, le quotient intellectuel est un indicateur de la vitesse du développement intellectuel chez l'enfant. Lorsqu'il publia en 1916 l'adaptation américaine du Binet-Simon qu'il avait réalisée à l'Université de Stanford,

Terman multiplia ce quotient par 100 pour éviter les décimales. Un QI de 1,2 devint donc un QI de 120, prenant ainsi l'ordre de grandeur que nous lui connaissons actuellement.

En sommant indistinctement les réussites à tous les types d'items de l'échelle, quel que soit le domaine dont ceux-ci relèvent, la procédure de calcul de l'âge mental définissait une dimension unique de développement de l'intelligence sur laquelle un indicateur chiffré unique, le QI , ordonnait les individus de façon univoque.

1.2. Le facteur g de Spearman

Un an avant que Binet et Simon publient la première version de leur échelle métrique de l'intelligence, un psychologue anglais, Charles Spearman (1863-1945) publie un article intitulé « L'intelligence générale objectivement déterminée et mesurée » (Spearman, 1904). Le problème qui préoccupe Spearman est celui de la relation entre les activités psychiques élémentaires, comme la discrimination sensorielle qu'il a appris à mesurer dans le laboratoire de psychologie expérimentale de Wundt, et les activités psychiques complexes qui sont en jeu hors du laboratoire, à l'école ou dans la vie professionnelle (voir le chapitre sur l'histoire de la psychologie différentielle). Contrairement à Binet, Spearman pense que ces deux sortes d'activité psychique font appel à l'intelligence. Si l'on a échoué à le montrer c'est, pense-t-il, parce que l'intelligence se manifeste probablement à des degrés différents dans ces différents types d'activité psychique. Pour apprécier le degré de relation entre ces activités, il propose de calculer leurs corrélations, une méthode de quantification des relations entre deux séries de mesures, dont le principe a été inventé par Galton quelques années auparavant. Si on trouve une corrélation entre les performances dans des tâches de discrimination sensorielle et les performances dans des tâches intellectuelles plus complexes, pense-t-il, c'est que les variations des performances observées dans ces deux domaines obéissent à une cause commune. Selon lui, cette cause commune pourrait être l'intelligence générale. Spearman a baptisé cette approche « psychologie corrélationnelle ». Elle consiste à s'appuyer sur les différences individuelles pour analyser, par le calcul de corrélations, la structure des relations entre les variables.

Mettant ces idées en pratique, Spearman a donc recueilli les notes scolaires d'élèves des écoles environnantes, leur a fait passer des épreuves de discrimination sensorielle, et a calculé les corrélations entre les scores des élèves à toutes ces épreuves. Trouvant des corrélations positives, mais d'intensité variable entre ces différentes épreuves, il a mis au point une méthode d'analyse de la table de corrélations permettant de séparer, dans ces corrélations, la

part de variance commune à toutes les épreuves, de celle qui est spécifique à chaque épreuve. C'est ainsi qu'il a extrait le facteur commun de variation sous-jacent à ces corrélations et qu'il a pu évaluer le degré auquel chacune des variables était affectée par ce facteur commun de variation (sa saturation dans le facteur). La première méthode d'analyse factorielle était née.

Spearman a interprété ce facteur commun, qu'il a désigné par la lettre g , comme étant un facteur d'intelligence générale intervenant, avec un poids variable, dans toutes les tâches intellectuelles, ce qu'il formulait ainsi : « toutes les branches de l'activité intellectuelle ont en commun une fonction fondamentale » (1904, p. 284). Il appelait sa théorie « bifactorielle » car elle expliquait la variance des scores dans chaque tâche intellectuelle par deux facteurs : le facteur g et un facteur de variation spécifique à la tâche. Dans le cadre de ce modèle bifactoriel, le score d'un sujet dans une tâche intellectuelle x est une fonction linéaire du score dans le facteur général et du score dans le facteur spécifique à la tâche, ce qui, du point de vue mathématique, s'écrit de la manière suivante :

$$x = \alpha g + \beta s + e$$

où x est le score dans la tâche, α le coefficient de saturation de la tâche en g (le poids du facteur général dans cette tâche, exprimé par la corrélation entre la tâche et le facteur qui varie donc de -1 à $+1$), g le score en facteur général, β le coefficient de saturation de la tâche dans le facteur qui est spécifique à cette tâche, s le score en facteur spécifique, et e l'erreur de mesure.

Spearman a plus tard interprété le facteur général comme correspondant à l'énergie mentale, une métaphore empruntée à la physique. Il en a aussi proposé une interprétation en termes de processus en observant que les tâches les plus saturées en g étaient celles qui exigeaient « l'appréhension de relations » (on dirait aujourd'hui le codage*), « l'éduction de relations » (on dirait aujourd'hui l'inférence : étant donnés deux termes, par exemple, gant et chaussure, trouver la relation qui les lie) et « l'éduction de corrélats » (on dirait aujourd'hui l'application : étant donné un terme et une relation, par exemple le terme « gant » et la relation « protège une partie du corps », trouver un second terme qui soit dans cette relation avec le premier). Comme on le verra plus loin, l'interprétation de g est encore problématique aujourd'hui mais la contribution essentielle de Spearman est d'avoir montré comment l'analyse mathématique de la matrice des corrélations observées dans un ensemble de variables manifestes peut permettre d'identifier une dimension latente (un facteur théorique) susceptible d'être la source commune des différences mesurées sur cet ensemble de variables

(on appelle variables manifestes celles qui sont observées, les notes dans les tests par exemple, et variables latentes celles qui sont inférées, théoriques, et que l'on suppose être à la source des variations observées dans les variables manifestes). La limite de la contribution de Spearman est, comme nous allons le voir, d'avoir réduit à une seule dimension, le facteur g, l'explication des différences individuelles observées dans les tâches intellectuelles (voir figure 9.1a la représentation graphique de la structure factorielle correspondant à ce modèle unidimensionnel).

 Insérer ici la figure 9.1

2. Des modèles unidimensionnels aux modèles multidimensionnels de l'intelligence

Binet et Spearman ont posé les bases de deux approches différentes de l'évaluation de l'intelligence. La première, issue des travaux de Binet est la construction d'échelles d'intelligence donnant lieu au calcul d'un QI, la seconde, issue des travaux de Spearman, est l'élaboration de batteries de tests dits « factoriels ». Cette dernière approche ne s'est cependant pas limitée à la mise au point de tests, elle est devenue en même temps une méthode de recherche fondamentale sur la structure de l'intelligence.

Au cours du siècle qui nous sépare des travaux de Binet et de Spearman, ces deux conceptions sont longtemps restées cloisonnées. Il est d'autant plus frappant d'observer que, par des voies différentes, l'une et l'autre ont progressivement abandonné le modèle unidimensionnel de l'intelligence qui sous-tendait les travaux de Binet et de Spearman pour aller vers des modèles multidimensionnels. Nous commencerons par retracer les grandes lignes de l'évolution des modèles factoriels pour en venir, dans un second temps, à l'évolution des échelles de QI.

2.1. L'évolution des modèles factoriels

2.1.1. Thurstone et les aptitudes primaires

Plusieurs auteurs ont critiqué la théorie bifactorielle de Spearman en cherchant à montrer que le facteur général n'épuisait pas la variance commune s'exprimant dans les corrélations, notamment Thorndike dès 1909, Thomson, dès 1916 et Kelley en 1928, mais Spearman résistait à l'idée qu'entre le facteur général et les facteurs spécifiques, il y ait place pour d'autres facteurs. Le modèle de l'intelligence unidimensionnelle a commencé à être

sérieusement ébranlé lorsqu'un psychologue américain, Louis Leon Thurstone (1887-1955), a proposé une nouvelle méthode d'analyse factorielle, dite multiple, dont l'objectif était d'aboutir à une structure factorielle aussi « simple » que possible. Dans son langage, une structure simple était une structure dans laquelle chaque groupe de variables plus corrélées entre elles qu'avec les autres était représenté par son propre facteur commun. La procédure qu'il utilisait rendait ces différents facteurs indépendants et permettait que chacun de ces groupes de variables soit saturé par un seul facteur (et ait donc des saturations nulles sur les autres facteurs). En appliquant cette procédure à l'analyse factorielle des corrélations entre les scores d'un échantillon important de sujets (240 étudiants volontaires) qui avaient passé un grand nombre de tests mentaux (56 tests variés, dont beaucoup mis au point pour l'occasion), Thurstone ne trouvait pas de facteur général d'intelligence, mais une douzaine de facteurs indépendants dont sept étaient interprétables : « Pour autant que nous puissions le déterminer pour l'instant, les tests qui ont été supposés être saturés par le facteur général divisent leur variance entre des facteurs primaires qui ne sont pas présents dans tous les tests. Nous ne pouvons rapporter aucun facteur commun général dans la présente étude (Thurstone, 1938, p. ix).

L'interprétation d'un facteur se fait en examinant le contenu des tests qui sont saturés par ce facteur et en cherchant ce qu'ils ont en commun. Trois des sept facteurs interprétables pouvaient être indexés en référence au contenu des tests qu'ils saturaient (Verbal, Numérique, ou Spatial), trois autres en référence aux fonctions mises en œuvre (Mémoire, Induction et Dédution¹), et le dernier en référence à la fois à un contenu et à un processus (Fluidité Verbale²). Thurstone considérait que ces 7 facteurs correspondaient à des aptitudes primaires (« Primary Mental Abilities ») et selon lui, la structure simple à laquelle il aboutissait démontrait que ces différentes aptitudes étaient indépendantes. En conséquence, il recommandait que chaque individu soit décrit par son profil d'aptitudes plutôt que par un unique indicateur de l'intelligence : voir figure 9.1b la représentation graphique de ce premier modèle multidimensionnel de l'intelligence.

La controverse entre Spearman et Thurstone sur le caractère unitaire ou multiple de l'intelligence a marqué le début d'un débat qui se prolonge encore sous d'autres formes : existe-t-il une forme unique d'intelligence, à l'œuvre dans toutes les tâches cognitives, ou

¹ Thurstone a dénommé ce facteur « Relations Verbales », mais l'inspection des tests les plus saturés conduit aujourd'hui à l'interpréter comme un facteur commun aux tests de raisonnement axés sur la déduction.

² Dans les tests de fluidité verbale, le sujet doit trouver en un temps limité, par exemple 2 minutes, le maximum de mots commençant par une lettre donnée, par exemple commençant par un S

plusieurs formes d'intelligence indépendantes dont chacune est spécifique à un domaine de compétence ? Il a fallu quelques temps à Spearman et Thurstone, ainsi qu'à leurs partisans respectifs, pour réaliser, puis admettre, que l'existence d'un facteur général et celle de facteurs plus spécifiques à des domaines n'étaient pas contradictoires.

2.1.2. Les modèles factoriels hiérarchiques

La solution de la contradiction apparente a été trouvée en élaborant des modèles factoriels hiérarchiques. Contrairement à ce que pensait Thurstone, du moins au début de ses travaux, les facteurs correspondant aux aptitudes primaires sont en général corrélés entre eux. Par conséquent, il est possible de faire une analyse factorielle de ces facteurs primaires (on appelle cela une analyse factorielle de second ordre) et d'extraire un ou plusieurs facteurs de second ordre rendant compte de la variance commune aux facteurs de premier ordre. Dans ce modèle factoriel hiérarchique, la variance des scores dans les tests peut donc être fractionnée en plusieurs parts : par exemple, une part de variance correspondant à un facteur général, des parts de variance correspondant à des facteurs larges correspondant à de grands domaines, et des parts de variance spécifiques à chacune des épreuves (correspondant aux facteurs s de Spearman).

Les plus influents de ces modèles hiérarchiques ont été celui de Burt-Vernon et celui de Cattell-Horn. Burt était anglais, disciple de Spearman et donc convaincu de l'importance du facteur général. Il a formulé le principe d'une analyse hiérarchique qui consisterait à extraire d'abord la variance commune à l'ensemble des tests, correspondant donc au facteur g , puis à analyser la matrice des corrélations résiduelles pour en extraire, à un niveau inférieur, les facteurs larges communs à des groupes d'épreuves seulement, puis à extraire, à un niveau encore inférieur, la variance spécifique à chacune des épreuves. Philip E. Vernon (1905-1987), disciple de Burt, a appliqué cette méthode d'analyse à une batterie de tests cognitifs (Vernon, 1950). Il a effectivement trouvé un facteur général d'intelligence puis, une fois la variance de celui-ci extraite, il a trouvé deux grands facteurs de groupe. L'un, dénommé verbal-éducatif (V-Ed) saturait surtout les tests du domaine verbal et les tests de connaissances scolaires, l'autre, dénommé kinesthésique-moteur (K-M), saturait surtout les tests des domaines pratique, mécanique, spatial et physique. Une fois la variance de ces deux facteurs larges extraite, l'analyse des corrélations résiduelles permettait d'extraire des facteurs plus étroits analogues aux aptitudes primaires de Thurstone.

Raymond B. Cattell (1905-1998), à ne pas confondre avec James Mc Keen Cattell dont il a été question dans le chapitre sur l'histoire de la psychologie différentielle, était lui aussi d'origine anglaise, il a soutenu sa thèse sous la direction de Spearman en 1929, mais il a émigré aux Etats-Unis en 1937 pour rejoindre le laboratoire de Thorndike, dont on se souvient qu'il était un des premiers opposants au facteur général cher à Spearman. Quelques années après, Cattell a défendu la thèse selon laquelle le facteur g ne suffisait pas et a proposé de distinguer deux facteurs généraux au lieu d'un seul, un facteur général d'intelligence fluide (désigné par le sigle Gf et correspondant à la capacité de raisonnement dans les situations nouvelles ne faisant que très peu appel à des connaissances antérieures) et un facteur général d'intelligence cristallisée (désigné par le sigle Gc et correspondant à la capacité à acquérir de nouvelles connaissances en s'appuyant sur les connaissances anciennes et des stratégies familières). Il a par la suite précisé cette théorie, longtemps appelée théorie Gf-Gc, et l'a étayée en faisant l'analyse factorielle hiérarchique de plusieurs batteries de tests (Cattell, 1963, 1971). La procédure d'analyse hiérarchique qu'utilisait Cattell était différente de celle de Vernon. Elle s'inspirait de celle que Thurstone avait le premier employée et –sachant que pour Thurstone, c'étaient les facteurs primaires qui étaient jugés les plus importants- elle procédait de bas en haut. La technique de Thurstone était d'abord utilisée pour extraire les facteurs primaires, puis, dans la mesure où ces facteurs primaires étaient corrélés, une analyse factorielle de second ordre était entreprise pour extraire le ou les facteurs communs à ces facteurs primaire. Par cette méthode, Cattell trouvait au premier niveau les facteurs correspondant aux aptitudes primaires de Thurstone, à ceci près qu'ils étaient plus nombreux, et au second niveau, deux facteurs généraux correspondant à l'intelligence fluide et à l'intelligence cristallisée. En enrichissant les batteries de tests, Horn, un disciple de Cattell, a par la suite, d'abord en collaboration avec Cattell, puis seul, étendu le nombre des facteurs primaires et le nombre des facteurs généraux de second ordre. Dans les années 60, il a ajouté aux deux facteurs généraux de second ordre initiaux Gf et Gc des facteurs généraux de visualisation (Gv), de Récupération en mémoire à long terme Gr, et de vitesse cognitive (Gs). Dans les années 90, Horn a porté à 9 le nombre de ces facteurs de second ordre (Horn, 1994). Ni Cattell ni Horn n'ont jamais véritablement admis l'existence d'un facteur général de troisième ordre rendant compte des corrélations entre les facteurs de second ordre (voir la figure 9.1c pour une représentation graphique de ce modèle).

Si le modèle factoriel hiérarchique de l'intelligence a permis une première forme de consensus dans la controverse sur le nombre des intelligences, une ou plusieurs, des divergences subsistaient, on le voit, quant à la structure précise de ce modèle hiérarchique. Le

modèle de Burt-Vernon comportait trois étages et un facteur général, celui de Cattell-Horn seulement deux étages et plusieurs facteurs généraux. Pourquoi ces discordances entre les modèles factoriels de l'intelligence ?

2.1.3. Les raisons des discordances

Pour comprendre les nombreuses controverses qui ont émaillé l'histoire des modèles factoriels de l'intelligence, il faut garder à l'esprit que dans une analyse factorielle, les facteurs que l'on trouve sont toujours relatifs. Relatifs à quoi ?

D'abord relatifs à l'échantillonnage des tests sur lesquels porte l'analyse factorielle. Il est clair qu'aucun facteur spatial ne pourra apparaître si l'on n'a pas mis d'épreuves de résolution de problèmes spatiaux dans la batterie de tests ou si l'on n'en a mis qu'un seul (pour trouver une part de variance commune il faut au moins deux épreuves du même domaine). Mais comment savoir si l'on a bien mis toutes les tâches qui sont susceptibles de faire apparaître un facteur de l'intelligence ? Cela supposerait que l'échantillonnage des tâches intellectuelles soit guidé par une théorie aboutie de l'intelligence. En l'absence d'une telle théorie, la constitution des batteries de tests sur lesquelles des analyses factorielles ont été effectuées a toujours été assez empirique. La liste des tests mentaux et celle des facteurs se sont allongées avec le temps et les recherches, mais comment savoir si des variables essentielles n'ont pas été oubliées, et comment savoir si leur introduction ne modifierait pas la structure qui fait maintenant consensus ?

Les facteurs sont aussi relatifs à l'échantillonnage des personnes à qui est administré l'échantillon de tests retenus. Un test qui est discriminant avec un échantillon de personnes peut ne pas l'être avec un autre échantillon. Pour prendre un exemple trivial, un test de raisonnement logique qui est discriminant avec un échantillon de sujets tout-venant risque de ne pas l'être avec un échantillon de logiciens. S'il n'est pas discriminant ou s'il l'est trop peu, c'est à dire si tous les sujets ont à peu près le même score, il ne pourra corrélérer avec aucune autre variable et la structure factorielle en sera changée.

Les facteurs sont enfin relatifs à la méthode d'analyse factorielle adoptée. De nombreuses options sont possibles et les chercheurs n'ont compris que progressivement les relations entre les différentes méthodes possibles et les différentes options possibles au sein de chaque méthode. C'était évidemment une des raisons du désaccord entre Spearman et Thurstone.

C'était aussi une des raisons du désaccord entre les modèles hiérarchiques de Burt-Vernon et de Cattell-Horn. Les uns et les autres faisaient certes une analyse factorielle hiérarchique mais les premiers la faisaient en procédant de haut en bas (extraction du facteur g d'abord) tandis que les seconds la faisaient de bas en haut (extraction des facteurs primaires d'abord). Pour des raisons qui seront expliquées un peu plus loin, cette différence de méthode peut expliquer que les uns aient trouvé trois étages et les autres seulement deux.

Toutes les analyses factorielles qui ont été évoquées jusqu'ici étaient des analyses dites « exploratoires ». Dans ce type d'analyse, on recherche de façon empirique une structure qui résume les données au mieux. L'inconvénient est que dans cette situation, ni le nombre de facteurs, ni les relations entre ces facteurs (sont-ils ou non corrélés ?), ni les saturations de ces facteurs dans les différentes épreuves, ne peuvent être spécifiés au départ. Cela fait beaucoup d'inconnues, beaucoup trop pour que l'on puisse arriver à une solution unique, complètement déterminée. On aboutit une famille de solutions possibles, qui peuvent être équivalentes du point de vue mathématique mais ne le sont pas nécessairement du point de vue psychologique. Une partie des incohérences entre les différents modèles factoriels de l'intelligence vient donc de l'indétermination inhérente à l'analyse factorielle exploratoire.

Le développement, dans les dernières décennies, de modèles structuraux comme, par exemple, le modèle LISREL³ (Jöreskog & Sörbom, 1993), a doté les chercheurs de méthodes d'analyse factorielle dites « confirmatoires » permettant de dépasser ces limitations. Dans l'approche confirmatoire, le chercheur peut mettre un modèle hypothétique de structure factorielle à l'épreuve. Ses hypothèses de départ lui permettent de fixer certains des paramètres de l'analyse (par exemple le nombre de facteurs, les relations entre les facteurs, les épreuves saturées par chacun des facteurs, etc.) ce qui réduit d'autant le nombre des inconnues et permet de déterminer une solution unique. Une fois que cette solution est déterminée, le modèle est testé en calculant l'écart entre la matrice de corrélations de départ (celle qui a été observée sur les données) et la matrice de corrélations d'arrivée (celle qui a pu être recomposée en appliquant ce modèle hypothétique des relations entre les variables latentes et manifestes). Cette procédure permet d'accepter ou de rejeter le modèle selon que l'écart entre les deux matrices de corrélations est ou non significatif, elle permet aussi de comparer des modèles concurrents et de les départager en fonction de la qualité de leur

³ Linear Structural RELationships

adéquation aux données (pour une présentation synthétique des deux logiques de l'analyse factorielle –exploratoire versus confirmatoire- voir Dickes, 1996).

2.1.4. Vers un consensus : le modèle Cattell- Horn-Carroll (CHC)

Malgré les difficultés signalées plus haut, un consensus assez général s'est fait, dans la période récente, autour du modèle proposé par John B. Carroll (1916-2003). En ré-analysant les données de 460 analyses factorielles de tests d'intelligence disponibles dans la littérature, Carroll (1993) a montré que toutes étaient compatibles avec une même structure factorielle générale. En appliquant la même méthode à toutes ces données, qui intégraient toute la variété des échantillons de tests et de sujets observés jusqu'alors, on peut penser que Carroll a neutralisé une partie des problèmes de méthode et d'échantillonnage signalés plus haut (une partie seulement car même si cette approche lui a permis d'élargir considérablement l'échantillon des tests pris en compte rien ne permet de dire que ceux-ci forment un échantillon représentatif de l'univers des tests d'intelligence possibles).

 Insérer ici la figure 9.2

En gardant ces limites en tête, venons en maintenant à la description de la structure factorielle dans laquelle Carroll a pu intégrer toutes les données qu'il a ré-analysées. Il s'agit d'une structure hiérarchique composée de trois strates (cf. figure 9.1d). La première strate, qui constitue la base de la pyramide, est composée d'une quarantaine de facteurs primaires, qui correspondent à des domaines trop spécifiques pour être détaillés ici (mais voir figure 9.2). Comme ces facteurs primaires entretiennent des corrélations – variables - entre eux, une analyse de second ordre permet d'extraire une seconde strate de huit facteurs larges qui rendent compte des parts de variance communes à des groupes de facteurs primaires. Puis, comme ces huit facteurs larges d'ordre 2 sont aussi corrélés, une analyse de troisième ordre permet d'extraire, au niveau de la troisième strate, un unique facteur général. Après orthogonalisation des facteurs extraits dans ces trois strates de l'analyse (c'est à dire en les rendant indépendants, de façon que chacun n'explique que la part de variance qui lui est propre), la variance observée dans les scores à un test d'intelligence peut être fractionnée en quatre parties distinctes : une partie qui est spécifique à ce test, et n'est donc pas très intéressante pour notre propos, une partie qui est commune à ce test et à un petit groupe d'autres tests couvrant le même champ étroit, par exemple le petit groupe des tests de rotation mentale (strate I), une partie commune à ce test et à un ensemble plus large de tests qui

couvrent le même domaine de la cognition, par exemple, tous les tests d'aptitude spatiale (strate II). Enfin, une partie commune à tous les tests d'intelligence, du moins à tous ceux qui ont été inclus dans les études ré-analysées par Carroll (strate III).

Ce modèle est de plus en plus souvent dénommé CHC (McGrew, 2005) car il fait la synthèse du modèle de Cattell- Horn (figure 9.1c) et de celui de Carroll (figure 9.1d). Les facteurs trouvés par Carroll sont en effet, pour l'essentiel, les mêmes que ceux du modèle de Cattell-Horn, la seule différence importante étant que Carroll trouve, dans toutes les études qu'il a ré-analysées, un facteur général qui constitue la strate III de son modèle. Horn et Cattell n'ont jamais extrait ce facteur car ils ont toujours été réticents sur son interprétation, mais ils ont toujours reconnu les corrélations existant entre leurs différents facteurs généraux (cf. figure 9.1c).

Le détail des facteurs trouvés dans les trois strates du modèle (CHC) est donné dans la figure 9.2. Le commentaire se limitera ici aux huit facteurs « larges » de la strate II. Dans la figure 9.2, ils sont placés dans l'ordre d'importance de leur saturation en facteur général, telle qu'elle a été estimée par Carroll en prenant en compte l'ensemble des données qu'il a ré-analysées. Les saturations les plus fortes sont celles des facteurs d'intelligence fluide (Gf) et d'intelligence cristallisée (Gc). Comme on peut le voir, sur la figure 9.2, chacun des facteurs de la strate II sature à son tour un certain nombre de facteurs de la strate I (aptitudes primaires, dont la liste s'est considérablement allongée depuis les travaux de Thurstone).

Une présentation brève de l'interprétation des différents facteurs de la strate II du modèle de Carroll est donnée dans l'encadré 9.2, en indiquant, pour chacun, le sigle par lequel il est le plus souvent désigné. Des exemples d'items de tests saturés par ces différents facteurs pourront être trouvés ailleurs (Huteau et Lautrey, 2003, p. 151-161).

Encadré 9.2

- *Intelligence fluide* (Gf) : sature les tests qui font appel au raisonnement et, plus généralement, à des opérations mentales contrôlées, pour résoudre des problèmes nouveaux faisant peu appel aux connaissances.
- *Intelligence cristallisée* (Gc) : sature les tests faisant appel à la connaissance du langage, de l'information et des concepts spécifiques à une culture et / ou à l'application de cette connaissance.

- *Mémoire et apprentissage* (Gm): correspond à l'efficacité de la mémoire à court terme et de la mémoire de travail, mais le contenu de ce facteur est assez flou dans la mesure où ces différents aspects de la mémoire n'étaient pas clairement distingués dans les études ré-analysées par Carroll.
- *Représentation visuo-spatiale* (Gv) : sature les tests faisant appel à la capacité à générer, retenir, retrouver et transformer des images visuelles.
- *Représentation auditive* (Ga) : sature les tests faisant appel à la capacité à analyser, manipuler, comprendre et synthétiser des éléments sonores, des groupes de sons ou des patterns sonores.
- *Récupération en mémoire à long terme* (Gr) : capacité à stocker l'information nouvelle en mémoire à long terme et à la retrouver plus tard de façon flexible. Ce facteur sature notamment les tests de fluidité verbale et de créativité.
- *Rapidité cognitive* (Gs): Capacité à effectuer de façon automatique et rapide des tâches relativement faciles ou surprises.
- *Vitesse de traitement* (Gt) : capacité à réagir ou à décider rapidement en réponse à des stimuli simples, comme c'est le cas dans les tâches évaluant le temps de réaction ou le temps d'inspection.

2.1.5.L'interprétation du facteur général

S'il y a un relatif consensus sur l'interprétation des facteurs de la strate II, l'interprétation et l'existence même du facteur général qui figure dans la strate III restent problématiques. Le désaccord ne porte pas sur la possibilité d'extraire un facteur général de la matrice des corrélations observées dans une batterie de tests cognitifs. Du fait que l'on observe toujours des corrélations positives entre les performances observées dans un ensemble de tests cognitifs, il est techniquement possible d'extraire un facteur général rendant compte de la variance commune à tous les tests, qui rend correspond à cette réalité statistique. C'est l'interprétation psychologique de cette réalité statistique que Cattell et Horn ont toujours contestée. Selon eux, cette part de variance commune à tous les tests peut tenir à ce qu'il n'existe pas de tâche cognitive complexe qui soit pure, au sens où elle ne ferait intervenir qu'un seul des processus ciblés par les facteurs de la strate II ; toutes font intervenir, à des degrés divers, plusieurs de ces processus (par ex. inférence, langage, mémoire, vitesse de traitement etc.). Auquel cas, la variance commune à toutes les épreuves pourrait être due au chevauchement des différents processus déjà pris en compte par les facteurs de la strate II. Pour démontrer la spécificité psychologique du facteur g, il faudrait démontrer qu'il rend

compte d'une variance qui lui est spécifique, qu'il ne partage donc avec aucun des autres facteurs, et que l'on puisse interpréter en termes de processus psychologique. Selon Horn et Blankson (2005), ces exigences ne sont pas remplies.

Allant dans le même sens, certains auteurs avancent que la variance expliquée par g serait en fait la même que celle expliquée par G_f . Cette hypothèse est notamment étayée par les résultats d'une étude de Gustaffson (1984). Cet auteur a testé par une méthode d'analyse factorielle confirmatoire un modèle factoriel hiérarchique comportant, dans la strate I, 10 facteurs primaires corrélés entre eux, dans la strate II, trois facteurs de second ordre également corrélés entre eux (les facteurs G_c , G_f , et G_v), et un facteur général coiffant l'ensemble dans la strate III (cf. figure 9.1d). Les données sur lesquelles ce modèle a été testé ont été obtenues en faisant passer 16 tests (2 par facteur primaire) à 1000 élèves du sixième degré en Suède. Il s'est avéré que le meilleur modèle factoriel pour recomposer les corrélations observées était bien le modèle hiérarchique en trois strates de Carroll (figure 9.1d). Mais un résultat très intéressant de cette étude est que ce modèle hiérarchique n'a pu être identifié qu'en spécifiant une saturation de 1 (c'est à dire une corrélation parfaite) entre g et G_f , autrement dit en posant que la variance expliquée par g est la même que celle expliquée par G_f . L'identité de g et G_f laisse alors le choix entre un modèle à trois strates dans lequel il existe un facteur g dans la strate III et pas de facteur G_f dans la strate II, ou un modèle à deux strates, dans lequel il n'existe pas de facteur g dans la strate III mais un facteur G_f dans la strate II, ce second modèle étant en fait celui de Cattell-Horn (cf. figure 9.1c).

Ouvrons une parenthèse pour revenir au désaccord signalé plus haut entre les modèles de Burt-Vernon et de Cattell-Horn. On se souvient que Vernon trouvait une structure factorielle à trois strates avec un facteur général dans la strate III et, dans la strate II, deux facteurs larges, Verbal-Educationnel et Kinesthésique-Moteur, dont les contenus sont en fait très proches, respectivement, des facteurs G_c et G_v du modèle de Cattell-Horn. Par contre, Vernon ne trouvait pas, dans la strate II, de facteur analogue au facteur G_f de Cattell-Horn (qui, eux, ne trouvaient pas de facteur général). A la lumière de l'étude de Gustaffson, on peut faire l'hypothèse que le désaccord entre ces deux modèles était dû à la différence de méthode dans l'analyse factorielle hiérarchique. Si dans ces études également la variance expliquée par g et par G_f était la même, ceci pourrait expliquer qu'en procédant de haut en bas, c'est à dire en commençant par extraire le facteur g , Vernon ne trouvait pas de variance correspondant au facteur G_f dans la strate suivante. Réciproquement, il n'est pas surprenant qu'en procédant de

bas en haut Cattell et Horn n'aient pas trouvé de variance spécifique au facteur g après avoir extrait le facteur Gf dans la strate II.

L'assimilation de g à Gf ne fait cependant pas l'unanimité. Dans les données qu'il a ré-analysées, Carroll trouve que Gf est certes le facteur de la strate II le plus fortement saturé par g mais qu'il n'épuise pas pour autant toute la variance de g et c'est la raison pour laquelle Carroll maintient g distinct de Gf et donc la strate III de son modèle.

Si l'interprétation de g reste problématique (voir Sternberg et Grigorenko (2002) pour une vue d'ensemble des interprétations en compétition et du débat à ce sujet), celle de Gf a davantage progressé dans la période récente. Pour beaucoup d'auteurs, le facteur Gf correspond à l'efficacité de la mémoire de travail (sur la mémoire de travail, voir le volume de psychologie cognitive, chapitre mémoire). La première recherche à avoir montré l'importance de ce lien entre l'intelligence fluide et la mémoire de travail est celle de Kyllonen et Christal (1990). Ces auteurs ont réalisé plusieurs expériences dans lesquelles de larges échantillons de sujets ont passé des épreuves variées de mémoire de travail et des tests variés de raisonnement (dont on connaît la forte saturation par le facteur Gf). Les analyses factorielles confirmatoires effectuées sur ces données ont permis, dans chacune de ces expériences, d'extraire un facteur commun aux épreuves de raisonnement (Gf) et un facteur commun aux épreuves de mémoire de travail. La corrélation entre ces deux facteurs variait entre .80 et .88 selon les études. Kyllonen et Christal en concluaient que l'aptitude au raisonnement pouvait, à peu de choses près, être assimilée à la capacité de la mémoire de travail. Engle, Tuholski et al. (1999) ont montré une relation forte entre le facteur général des épreuves de mémoire de travail et le facteur d'intelligence fluide. En incluant des épreuves de mémoire à court terme et des épreuves de vitesse cognitive dans leurs batteries, Conway, Cowan et al (2002) ont précisé ce résultat en montrant que la relation forte entre la capacité de la mémoire de travail et le facteur Gf ne pouvait être expliquée ni par l'empan de la mémoire à court terme (correspondant à la seule capacité de stockage) ni par la vitesse de traitement.

Compte tenu du fort recouvrement entre le facteur Gf et le facteur g, peut-on expliquer Gf, et donc une part importante de g, par l'efficacité de la mémoire de travail et plus généralement par l'efficacité des processus exécutifs ? Beaucoup de recherches vont actuellement dans ce sens (voir aussi le chapitre sur les théories néo-piagétienne dans la partie développementale de ce volume). Il faut cependant avoir conscience que le concept de mémoire de travail est

lui-même très général et qu'avant de comprendre comment l'efficacité de la mémoire de travail influence la performance dans les tests d'intelligence il faudra mieux comprendre comment les différents processus cognitifs intervenant dans la mémoire de travail s'orchestrent au cours de l'exécution d'une tâche cognitive complexe (Lautrey 2005a ; Süß et al, 2002).

2.2. L'évolution des échelles d'intelligence : l'exemple des échelles de Wechsler

La première adaptation américaine du Binet-Simon, celle que Terman a effectuée à l'Université de Stanford et publiée en 1916, était fidèle au principe de construction adopté par Binet et au mode de calcul du QI préconisé par Stern (quotient de l'âge mental sur l'âge chronologique). Les premiers changements dans le principe de construction des échelles d'intelligence sont dus à Wechsler. Les échelles de Wechsler étant à l'heure actuelle celles qui sont les plus utilisées en France et dans le monde pour l'évaluation individuelle, ce sont elles que nous prendrons comme exemple pour illustrer l'évolution des tests de QI.

David Wechsler (1896-1981) était psychologue à la consultation d'adultes de l'hôpital Bellevue, un hôpital psychiatrique new-yorkais. Le Terman-Stanford (devenu ensuite Binet-Stanford) étant mal adapté à l'évaluation de l'intelligence d'adultes, Wechsler a entrepris d'élaborer une nouvelle échelle d'intelligence destinée à cette population. C'est ainsi qu'est née l'échelle dite Wechsler-Bellevue, publiée en 1939 et modifiée ensuite pour devenir en 1955 la Wechsler Adult Intelligence Scale (WAIS). Le succès de l'échelle d'intelligence pour adultes a conduit Wechsler à élaborer sur les mêmes principes une nouvelle échelle adaptée aux enfants de 6 à 16 ans, la Wechsler Intelligence Scale for Children (WISC) qui fut publiée en 1949. Enfin, une échelle adaptée aux enfants d'âge préscolaire, elle aussi bâtie sur les mêmes principes, la Wechsler Preschool and Primary Intelligence Scale (WPPSI) a été éditée nettement plus tard, en 1967. Ces différentes échelles ont été révisées régulièrement : pour la WAIS, en 1981 (WAIS-R) et en 1997 (WAIS III) ; pour la WISC, en 1974 (WISC-R), en 1991 (WISC III) et en 2003 (WISC IV) ; pour la WPPSI, en 1989 (WPPSI-R) et en 2002 (WPPSI III). Chaque révision de ces échelles a été suivie d'une adaptation française de la version américaine.

Wechsler avait une approche pragmatique de l'évaluation de l'intelligence et partageait, dans les grandes lignes, les idées de Binet sur la question. Lui aussi considérait l'intelligence comme une capacité globale faite d'un ensemble d'aptitudes spécifiques auxquelles elle ne se

réduit cependant pas. Il la définissait ainsi: « L'intelligence est déterminée de façon multiple et a des facettes multiples. Ce à quoi elle fait toujours appel n'est pas une aptitude particulière mais une compétence d'ensemble ou capacité globale qui, d'une façon ou d'une autre, permet à un individu sensible de comprendre le monde et de réagir efficacement à ses défis » (Wechsler, 1981, p. 8). Pour évaluer cette capacité globale, Wechsler, comme Binet, s'est attaché à varier autant que possible la nature des subtests de l'échelle et, pour ce faire, il s'est davantage appuyé sur son expérience clinique que sur une théorie précise de l'intelligence. Il s'est par contre écarté des principes de construction de l'échelle de Binet-Simon sur deux points importants. Le premier est d'avoir abandonné la notion d'âge mental pour lui préférer le rang auquel la performance d'un individu situe celui-ci dans son groupe d'âge. Le second est d'avoir amorcé le fractionnement de la notion d'intelligence en permettant d'évaluer séparément un QI verbal et un QI de performance. Nous verrons plus loin que cette évolution des échelles de Wechsler vers une conception multidimensionnelle de l'intelligence s'est encore accentuée plus récemment, notamment avec la publication de la WISC IV, en s'inspirant du modèle factoriel CHC. Cette tendance n'est pas propre aux échelles de Wechsler, elle caractérise l'évolution de toutes les échelles d'intelligence dans la période récente.

2.2.1. L'abandon de la notion d'âge mental et du quotient.

La pratique consistant à évaluer l'intelligence en faisant le quotient de l'âge mental sur l'âge chronologique posait deux problèmes. Le premier tenait à ce que le calcul de ce quotient dans une échelle destinée à l'évaluation de l'intelligence d'adultes conduisait à une absurdité.

L'âge mental plafonne vers la fin de l'adolescence alors que l'âge chronologique augmente tout au long de la vie. En conséquence, à partir de la fin de l'adolescence, un QI calculé ainsi diminuerait régulièrement au fur et à mesure que l'âge chronologique augmente. L'âge chronologique auquel survient ce plafond dans les performances aux tests d'intelligence peut varier selon les individus et la nature des épreuves mais il est en général atteint entre 15 et 25 ans et va de pair avec la fin de la maturation du système nerveux central. Ceci ne signifie évidemment pas que les capacités intellectuelles cessent alors d'évoluer, mais seulement que leur progression ne continue que dans les domaines où l'accroissement de la performance repose sur l'accroissement des connaissances et de l'expertise. Or l'objectif des tests d'intelligence n'est pas d'évaluer les connaissances scolaires ou le degré d'expertise professionnelle (c'est l'objet des tests de connaissances), il est d'évaluer la capacité à apprendre, à s'adapter à des situations nouvelles. Le plafond atteint dans cette capacité vers la

fin de l'adolescence rend l'utilisation de la notion d'âge mental, et donc du quotient de l'âge mental sur l'âge chronologique, inadaptée chez l'adulte.

Le second problème rencontré avec le quotient de l'âge mental sur l'âge chronologique tenait à ce que sa dispersion n'était pas la même aux différents âges chronologiques. L'écart-type du QI pouvait être, par exemple, de 12 à un certain âge et de 18 à un autre, de telle sorte qu'un même QI n'avait pas exactement la même signification selon l'âge auquel il était calculé.

La solution adoptée par Wechsler pour régler ces deux problèmes a consisté à caractériser la performance de la personne examinée par le rang auquel se situe sa note brute dans la distribution des notes brutes observées dans son groupe d'âge (la note brute dans un tests est obtenue en faisant la somme des points obtenus aux items réussis). Lors de l'étalonnage du test, un échantillon de personnes représentatif de la population est examiné pour chacun des groupes d'âge que l'on souhaite constituer. Avec cet étalonnage, la performance de chaque nouvelle personne examinée peut être située par rapport à celles des personnes de son âge. Par exemple, on pourra dire que la note brute obtenue par une personne est dépassée par 16% des notes brutes observées dans l'échantillon d'étalonnage (ou dépasse 84% de ces notes). Wechsler aurait pu en rester à cette caractérisation de la personne, c'est à dire au rang auquel son score la classe dans son groupe d'âge. Il a cependant voulu faire ressembler ce mode de caractérisation à un QI et a pour cela appliqué à ce rang deux transformations.

La première transformation s'appuie sur les propriétés de la distribution normale. Si on normalise la distribution des notes brutes (voir chapitre 8, les étalonnages et la mesure), celles-ci peuvent être transformées en écarts à la moyenne exprimés en prenant pour unité l'écart-type (cf. chapitre 8, figure 8.5b). Ainsi, la note dépassée par 16% des personnes correspond à un écart-type au-dessus de la moyenne ; la note dépassée par seulement 2,5% des personnes correspond à deux écarts-types au-dessus de la moyenne, etc. Cette transformation fait passer d'un niveau de mesure ordinal (le rang auquel sa note brute situe la personne dans l'échantillon) à une échelle d'intervalle dotée d'une origine (le zéro correspond à la moyenne) et d'une unité (l'écart-type). Sur cette nouvelle échelle en écarts-réduits, la note brute dépassée par 16% des individus devient un score de + 1 écart-type, la note brute dépassée par 2,5% des individus devient un score de +2 écarts-types, etc.

Ces scores en écarts réduits n'étant pas très commodes à manier, Wechsler les a transformés en les multipliant par 15 et en ajoutant 100, de telle sorte que la moyenne devienne 100 au lieu de 0 et l'écart-type 15 au lieu de 1. Pourquoi ces nombres ? Parce que cette transformation donne aux scores le même ordre de grandeur que les QI qui, avec les échelles de type Binet-Simon, étaient calculés en faisant le quotient de l'âge mental sur l'âge chronologique. Le QI de 100 correspond au cas où la note brute obtenue est la moyenne du groupe d'âge considéré, ce qui correspond aussi au cas où l'âge mental est égal à l'âge chronologique. L'écart-type de 15 est celui que l'on observait, en moyenne –mais avec les variations signalées plus haut - dans les QI au Binet-Stanford. Wechsler a baptisé QI le score ainsi transformé, de moyenne 100 et d'écart-type 15. Par ce tour de passe-passe, la note brute dépassée par 16% des individus du groupe d'âge considéré devient un QI de 115 (si la moyenne est 100 et l'écart-type 15, un score situé à un écart-type de la moyenne est de $100 + 15$), la note brute dépassée par 2,5% de l'échantillon devient un QI de 130, etc.

Contrairement au quotient de l'âge mental sur l'âge chronologique, ce nouveau mode de calcul du QI peut s'appliquer aux adultes aussi bien qu'aux enfants. D'abord adopté pour l'échelle destinée aux adultes (WAIS) il a été étendu aux échelles pour enfants (WISC et WPPSI). Il résout également le problème que posait la variabilité de l'écart-type du QI de type Stern selon les âges puisque, par construction, l'écart-type est ici de 15 à tous les âges. Wechsler a également appliqué cette procédure de transformation à la note brute obtenue dans chacun des subtests de l'échelle. Dans ce cas la note brute est transformée en une note standard de moyenne 10 et d'écart-type 3. Les notes standard obtenues dans les différents subtests par un même individu deviennent ainsi strictement comparables (une note standard de 7 signifie que la note brute se situe à un écart-type en-dessous de la moyenne, quel que soit le subtest). Ces différents avantages expliquent sans doute que ce nouveau mode de calcul du QI ait été adopté par la suite dans toutes les échelles d'intelligence, y compris dans le Binet-Stanford. Dans ce qui suit, le terme QI employé sans autre spécification renverra donc au QI de type Wechsler tandis que le quotient de l'âge mental sur l'âge chronologique sera dénommé QI de type Stern.

Le QI de type Wechsler n'est donc pas un quotient. Dans les échelles pour enfants (WISC et WPPSI), ce QI peut, comme le QI de type Stern, être interprété comme une évaluation de la vitesse de développement mais celle-ci n'est plus quantifiée de la même manière : c'est ici le rang auquel sa performance situe un enfant dans son groupe d'âge qui renseigne sur le degré

d'avance ou de retard de son développement par rapport à celui des enfants de son âge. Toutefois, pour les raisons expliquées plus haut, ce même indicateur ne peut être interprété comme une évaluation de la vitesse du développement dans les échelles pour adultes.

Les transformations par lesquelles on passe du rang de la note brute à un QI ne doivent pas faire illusion. Elles ne transforment pas l'échelle ordinale dont on est parti (le rang) en une échelle d'intervalle. Comme cela a déjà été précisé à propos de la méthode des tests (voir chapitre 8, les étalonnages) il n'y a pas de sens à parler de la forme de la distribution d'une variable de niveau ordinal. La forme normale de la distribution des notes brutes, sur laquelle on s'appuie pour passer du rang dans le groupe d'âge à un score exprimé en écarts-types, n'est pas une propriété de la mesure de l'intelligence. Elle est purement conventionnelle. Quelle est alors son utilité ? L'adoption de cette convention est généralement justifiée par la volonté de faire des discriminations plus fines aux extrêmes de la distribution que dans la zone moyenne, ou par le fait qu'une distribution normale permet d'utiliser les méthodes statistiques qui postulent la normalité de la distribution. La transformation d'échelle qui permet ensuite d'utiliser un indicateur de moyenne 100 et d'écart-type 15 relève, elle aussi, de la simple commodité. Ces « arrangements » ne doivent pas faire perdre de vue que le niveau de mesure le plus puissant sur lequel les psychologues puissent fonder leurs évaluations de l'intelligence est le niveau ordinal: ils peuvent ordonner les sujets en fonction de leurs performances dans les tests d'intelligence (nombre d'items réussis) et ils peuvent aussi ordonner les items de ces tests en fonction de leur niveau de difficulté (nombre de sujets qui réussissent chaque item). Le QI auquel on aboutit après les transformations qui viennent d'être décrites ne traduit donc rien d'autre que le rang auquel sa note dans le test situe une personne dans son groupe d'âge et il serait souhaitable de s'en tenir à cette information, la seule dont les fondements scientifiques soient solides.

2.2.2 Le fractionnement du concept d'intelligence

Le second point sur lequel Wechsler s'est écarté des principes de construction du Binet-Simon est le fractionnement des subtests de l'échelle en deux sous-échelles, l'une verbale, l'autre de performance. Là encore, ce ne sont pas des considérations théoriques qui ont inspiré ces modifications à Wechsler, mais plutôt son expérience de praticien. Il avait fait partie de la cohorte de jeunes psychologues enrôlés pour faire passer les tests Army alpha et Army beta aux jeunes recrues de l'armée américaine lors de la première guerre mondiale (voir. chapitre 7). Il avait pu observer à cette occasion l'intérêt de disposer d'un test non verbal, l'Army beta

pour l'examen du niveau intellectuel d'adultes qui, pour une raison ou une autre, ont des difficultés dans l'utilisation du langage. Sa pratique clinique ultérieure à l'hôpital Bellevue l'a conforté dans cette idée et il a donc fait en sorte que puissent être évalués deux QI distincts, l'un pour la partie verbale de l'échelle et l'autre pour la partie de performance.

Dans toutes les échelles de Wechsler, la sous-échelle verbale était constituée des sous-tests suivants : information (questions d'information générale de difficulté graduée), mémoire des chiffres (répéter en ordre direct ou en ordre inverse des séries de chiffres de longueur croissante), vocabulaire, arithmétique, compréhension (expliquer des observations de la vie quotidienne ou des proverbes), et similitudes (dire en quoi deux choses différentes se ressemblent, par ex. pomme et prune). La sous échelle de performance était constituée des sous-tests suivants : complètement d'images (trouver la partie qui manque dans des images), arrangement d'images (réordonner des séries d'images de façon à ce que leur séquence corresponde au déroulement de l'histoire qu'elles racontent), cubes (des figures géométriques sont présentées et elles doivent être reproduites avec des cubes dont certaines faces sont unicolores et d'autres bicolores), assemblage d'objets (assembler des morceaux de carton découpé présentés en désordre, de telle sorte qu'ils reconstituent un objet familier), et code (avec un modèle de code dans lequel les chiffres sont remplacés par des signes arbitraires, remplacer les plus grand nombre possible de chiffres par leur code dans un temps limité). Ces différents sous-tests étaient bien entendu de niveaux de difficulté différents dans les échelles pour enfants et dans l'échelle pour adultes.

Dans les subtests de la partie verbale, les réponses étaient données oralement, en temps libre, l'évaluation ne portant que sur leur justesse. Il s'agissait de définitions, d'énoncés, d'explications. Dans les subtests de la partie performance, les réponses étaient non verbales. Il s'agissait d'actions consistant à positionner des cubes, pointer une partie manquante, réarranger des images, etc. Les épreuves de performance étaient par ailleurs chronométrées et dans cette partie de l'échelle, l'efficacité était appréciée à la fois par la justesse de la réponse et par sa vitesse. Par ailleurs, les problèmes à résoudre étaient ici de nature visuo-spatiale, ce qui n'était pas le cas dans la partie verbale de l'échelle.

Autrement dit, si la partition opérée par Wechsler dans son échelle d'intelligence permettait bien de distinguer deux aspects de l'intelligence, il s'agissait d'une distinction assez empirique et encore très globale. Les deux sous-échelles différaient à la fois, on vient de le

voir, par le format de représentation des items proposés (verbal versus spatial), le mode d'expression de la réponse (verbal versus moteur) et l'aspect de la performance pris en compte (justesse versus justesse et vitesse). Il n'était donc pas facile d'interpréter les différences trouvées entre le QI verbal et le QI performance lorsqu'il y en avait. D'autant plus que chacune des sous-échelles mêlait des aspects de l'intelligence correspondant à des dimensions différentes dans le modèle factoriel de l'intelligence (le raisonnement, la représentation visuo-spatiale et la vitesse de traitement dans la sous-échelle de performance ; les connaissances, la compréhension verbale et la mémoire de travail dans la sous-échelle verbale).

Pendant plus d'un demi siècle, les échelles de Wechsler n'ont cependant pas varié. Les révisions successives ont porté sur l'actualisation des contenus et de la difficulté des items, mais pas sur la nature des subtests ni la structure de l'échelle. C'est seulement récemment, déjà avec les révisions qui ont donné lieu à la WISC III et la WAIS III, mais avec la révision qui a donné lieu en 2003 à la WISC IV que la nature des subtests et la structure de l'échelle ont été modifiés pour tenir compte des apports de l'approche factorielle de l'intelligence et de la psychologie cognitive. C'est ainsi un pas de plus vers une évaluation multidimensionnelle de l'intelligence qui a été effectué. Cette dernière évolution sera illustrée avec l'exemple de la WISC IV.

2.2.3. Vers des échelles d'intelligence plus clairement multidimensionnelles: l'exemple de la WISC IV

2.2.3.1. Les principes de construction

Un des objectifs de la révision de la WISC IV (Wechsler, 2005a, 2005b) était de permettre une évaluation plus claire de certaines des dimensions du modèle CHC, en particulier des facteurs d'intelligence cristallisée, d'intelligence fluide, et de vitesse de traitement. Un autre objectif était de permettre l'évaluation de l'efficacité de la mémoire de travail pour prendre en compte les travaux ayant montré les relations entre la capacité de la mémoire de travail et l'intelligence (voir plus haut, et aussi, dans le chapitre 2, les théories néo-piagétienne). Ceci a conduit à fractionner l'échelle en quatre sous-échelles évaluant les performances sur quatre dimensions distinctes du développement cognitif, la compréhension verbale, le raisonnement perceptif, la mémoire de travail, et la vitesse de traitement, chacune donnant lieu au calcul d'un indice factoriel de développement.

L'indice de *compréhension verbale* (ICV) correspond au facteur d'intelligence cristallisée du modèle CHC. Il est évalué par trois des sous-tests qui composaient la sous-échelle verbale des anciennes versions de la WISC : *similitudes*, *vocabulaire* et *compréhension*. Les autres subtests qui faisaient traditionnellement partie de la sous-échelle verbale, information et arithmétique, ont été retirés. Ces deux subtests étaient ceux qui faisaient le plus appel aux connaissances. En les retirant, l'objectif est de centrer l'évaluation sur la compréhension verbale en diminuant autant que faire se peut le poids des acquisitions de type scolaire.

L'indice de *raisonnement perceptif* (IRP) se rapproche du facteur d'intelligence fluide du modèle CHC. Il est évalué par trois subtests : *cubes*, *identification de concepts* et *matrices*. Par rapport aux subtests qui composaient la sous-échelle de performance dans l'ancienne version de la WISC, seul le subtest cubes a été conservé. Deux nouveaux subtests, dont le contenu sera précisé plus loin ont été introduits : d'une part l'épreuve des matrices, conçue sur le même principe que le test des Matrices Progressives de Raven, dont on sait qu'il est fortement saturé par le facteur d'intelligence fluide, d'autre part l'épreuve d'identification de concepts qui fait appel à la catégorisation, l'abstraction et la formation de concepts. Les subtests qui ont été retirés sont par contre ceux qui faisaient fortement appel à la représentation visuo-spatiale (assemblage d'objets, arrangement d'images, complètement d'images) et à la vitesse de traitement (code). Par ailleurs, les réponses ne sont plus chronométrées. En somme, alors que les subtests de la sous-échelle de performance des anciennes versions de la WISC évaluaient indistinctement le raisonnement, la représentation visuo-spatiale et la vitesse de traitement, ceux qui entrent dans le calcul de l'indice de raisonnement perceptif de la WISC IV sont plus nettement focalisés sur le raisonnement, dont on sait qu'il est fortement saturé par le facteur d'intelligence fluide.

L'indice de *vitesse de traitement* (IVT) correspond au facteur de vitesse cognitive dans le modèle CHC. Il est évalué par les subtests *code* et *symbole*. La vitesse de traitement, qui était auparavant évaluée indistinctement dans tous les subtests de la sous-échelle de performance est maintenant évaluée avec deux épreuves spécifiques dont l'une (code) faisait auparavant partie de la sous-échelle de performance tandis que l'autre (symboles) est nouvelle et a été introduite pour renforcer l'évaluation de cette dimension spécifique.

L'indice de mémoire de travail (IMT) évalue la capacité de la mémoire de travail à partir des subtests *mémoire des chiffres* et *séquence lettres-chiffres*. Le subtest de mémoire des chiffres figurait auparavant dans la sous-échelle verbale de la WISC mais n'était pas une bonne évaluation de la mémoire de travail car il mêlait indistinctement des items où la séquence de chiffres énoncés par le psychologue devait être répétée à l'endroit et des items où celle-ci devait être répétée à l'envers. Le premier type d'items, qui ne requiert que le stockage des informations évalue en fait la mémoire à court terme et seul le second, qui requiert simultanément un stockage et un traitement (pour réordonner les chiffres en sens inverse), fait appel à la mémoire de travail.

En résumé, la WISC IV marque l'abandon d'un fractionnement empirique et assez grossier de l'échelle entre une partie verbale et une partie dite de performance, pour se rapprocher d'un modèle multidimensionnel de type CHC intégrant aussi les apports de la psychologie cognitive sur le rôle de la mémoire de travail. Quatre dimensions différentes sont maintenant évaluées, chacune étant plus clairement centrée sur un facteur du modèle CHC (intelligence cristallisée, intelligence fluide, vitesse de traitement) ou un processus (mémoire de travail). Chaque dimension est évaluée avec plusieurs épreuves (3 pour ICV et IRP, 2 pour VT et MT) car il faut au minimum deux épreuves pour extraire un facteur commun.

2.2.3.2. La composition de l'échelle

Compréhension verbale (ICV)

- *Similitudes* (23 items). Une paire de mots est énoncée oralement à l'enfant qui doit trouver la similitude entre les deux objets ou les deux concepts. Exemple : en quoi rouge et bleu se ressemblent ? Les processus ciblés sont le raisonnement verbal et la formation de concepts.
- *Vocabulaire* (36 items). La tâche de l'enfant est de donner la définition des mots énoncés. Exemple : qu'est ce qu'un parapluie ? On évalue ainsi la connaissance du lexique, la formation de concepts verbaux, la structuration de la mémoire à long terme et le développement du langage.
- *Compréhension* (21 items). On demande à l'enfant de dire ce qu'il comprend de situations sociales et de principes généraux. Exemple : pourquoi faut-il se brosser les dents ? Ce subtest évalue plus particulièrement la compréhension et l'expression verbales, la connaissance des conventions comportementales et des normes sociales.

Raisonnement perceptif (IRP)

- *Cubes*. L'enfant utilise des cubes bicolores (blanc / rouge) pour reproduire une construction dont le modèle lui est présenté sur une image. C'est principalement la capacité à analyser et à synthétiser des stimuli abstraits qui est évaluée par ce subtest.

- *Identification de concepts* (28 items). Des images disposées sur deux ou trois rangées (et dont le nombre varie avec le niveau de difficulté des items) sont présentées à l'enfant. Sa tâche est de choisir une image dans chaque rangée pour constituer une catégorie s'articulant autour d'un concept commun. L'exemple d'entraînement est une planche comportant une trompette et un chêne dans la première rangée, et un vase et un sapin dans la seconde rangée. La bonne réponse est le choix du chêne et du sapin. Les processus visés sont le raisonnement catégoriel et le raisonnement abstrait.

- *Matrices* (35 items). L'enfant observe une matrice incomplète et sélectionne la case manquante parmi cinq possibilités de réponse présentées en-dessous. Les processus ciblés sont l'inférence, le raisonnement analogique, qui sont fortement saturés par le facteur d'intelligence fluide et le facteur général d'intelligence.

Mémoire de travail

- *Mémoire des chiffres* (8 items). On lit une série de chiffres que l'enfant doit répéter dans l'ordre direct pour la première partie de l'épreuve et dans l'ordre inverse pour la seconde partie. La longueur de la série de chiffres va de 2 au premier item à 8 au dernier item (chaque item comporte deux essais).

- *Séquence lettres chiffres* (10 items): On lit au rythme d'un élément par seconde une séquence mélangeant des lettres et des chiffres. L'enfant doit en restituer les éléments en commençant par les chiffres en ordre croissant, puis les lettres en ordre alphabétique. Exemple : l'item énoncé 3 – M – 2 – C doit être restitué 2- 3- C- M. Les séquences vont de deux éléments pour les items les plus faciles à 8 éléments pour les plus difficiles. Comme dans la répétition de chiffres à l'envers, cette épreuve requiert la gestion simultanée du stockage et du traitement de l'information (ici le réarrangement des chiffres et des lettres dans leurs ordres respectifs), ce qui est la caractéristique des épreuves de mémoire de travail.

Vitesse de traitement

- *Code*. La tâche consiste à substituer des symboles sans signification à des nombres. L'enfant a à sa disposition un modèle donnant la correspondance entre les nombres et les signes qui les

représentent dans le code. Il doit compléter ensuite la feuille de réponse aussi vite qu'il le peut en mettant sous chaque nombre le signe qui lui correspond.

- *Symboles*. La tâche de l'enfant est de parcourir des lignes de symboles et d'indiquer, pour chaque ligne, si oui ou non un des symboles cibles qui sont placés à gauche se trouve dans la ligne. La consigne est de vérifier ainsi le maximum de lignes en temps limité.

En plus de la vitesse de traitement, ces deux épreuves font notamment appel à l'attention soutenue, à la mémoire à court terme et à l'apprentissage. Le code, qui demande la copie de signes arbitraires appariés à nombres évalue en outre la coordination visuo-motrice.

2.2.3.3. Les principes de cotation

Le manuel précise, pour chaque subtest, le nombre de points accordés pour la réussite de chaque item. Lorsque qu'il s'agit d'items dont les réponses sont ouvertes, comme dans les subtests de vocabulaire ou de compréhension, des exemples de bonnes réponses cotées 2 points, de réponses intermédiaires cotées 1 point et de mauvaises réponses cotées 0 points sont fournies dans le manuel pour limiter la part de subjectivité dans la cotation. Cette cotation des réponses étant faite, la somme des points donne la note brute obtenue dans chacun des dix subtests.

 Insérer ici la figure 9.3

L'exemple de protocole qui figure dans le manuel d'administration de la WISC (cf. figure 9.3) permet de suivre les étapes de la cotation. Le calcul de l'âge de l'enfant est la première opération à faire (partie A). Ceci permet de se reporter à la table dans laquelle figure la distribution des notes brutes observées dans le groupe d'âge de cet enfant lors de l'étalonnage du test. Dans cet exemple, l'enfant fictif Pierre Martin ayant 12ans 3 mois, on se reporte à la table du manuel dans laquelle figurent les distributions des notes brutes aux différents subtests pour la tranche d'âge allant de 12ans à 12 ans 3 mois. Cette table indique, pour chaque score brut, le score standard qui lui correspond dans une distribution normalisée en 19 classes de moyenne 10 et d'écart-type 3 (cf. plus haut la procédure de transformation des scores bruts en scores standards). Dans la partie B du protocole reproduit en figure 9.3, ces notes standard figurent à la droite des notes brutes. Dans les colonnes grisées suivantes, elles sont réparties entre les colonnes correspondant aux quatre indices (compréhension verbale, raisonnement perceptif, mémoire de travail et vitesse de traitement) avec, en bas du tableau, la somme des

notes standards pour chaque indice (les notes obtenues aux deux épreuves optionnelles données dans ce cas, barrage et arithmétique sont mise entre parenthèses car elles n'entrent pas dans le calcul des indices ni dans le calcul du QI total).

La transformation suivante consiste – en se reportant à nouveau à la table appropriée dans le manuel - à passer de ces sommes de notes standard aux notes composites de moyenne 100 et d'écart-type 15. Cette seconde transformation est destinée à rendre ces notes directement comparables (et aussi, comme nous l'avons vu plus haut, à leur donner une échelle de grandeur comparable à celle des anciens QI de type Stern). On trouve dans la première colonne de la partie D du protocole les sommes de notes standard correspondant aux différents indices et au total. Dans la colonne suivante, on trouve les notes composites de moyenne 100 et d'écart-type 15 qui leur correspondent . La somme totale des notes composites est appelée QI total, ce QI total est ici de 101. La troisième colonne du tableau D donne l'équivalence de ces notes composites en percentiles (voir chapitre 8, étalonnages). Cette indication est sans doute celle qui situe le plus clairement le niveau de la performance auquel correspond la note puisqu'elle indique le pourcentage de sujets qui ont eu une note inférieure à celle-ci dans l'échantillon de référence (ici, 82% pour la note de compréhension verbale et 25% pour la note de raisonnement perceptif). La dernière colonne du tableau D indique l'intervalle de confiance de chaque note (cf. chapitre 7, l'erreur-type). Ceci signifie qu'au seuil de risque de 5%, on peut estimer que le score vrai de compréhension verbale devrait se situer quelque part entre 104 et 121.

Les tableaux qui sont à droite permettent de tracer, sur l'échelle des notes allant de 1 à 19, les profils reliant, dans chaque catégorie d'épreuves, les notes standard obtenues dans les différents subtests (partie E du protocole) et, sur une échelle de moyenne 100 et d'écart-type 15, le profil des notes composites obtenues sur les quatre dimensions de l'intelligence distinguées dans l'échelle ainsi que le QI total. Les auteurs du manuel indiquent que le QI total peut être considéré comme une approximation du facteur général d'intelligence. Ce n'en est qu'une approximation grossière, car il est obtenu en faisant la simple somme des notes dans les différents indices factoriels alors que, si un score en facteur général était calculé, chacune des dimensions y interviendrait avec un poids correspondant à sa saturation en facteur général. A titre d'exemple, on sait que la vitesse de traitement contribue peu au facteur général alors que la mémoire de travail y contribue beaucoup. Si l'on calculait un score en facteur général, la note de vitesse de traitement n'y contribuerait donc qu'avec un poids réduit

tandis que celle de mémoire de travail y contribuerait avec un poids important. Par contre, avec la procédure adoptée ici, les notes de vitesse de traitement et celles de mémoire de travail sont ajoutées avec le même poids. Les auteurs de la révision du WISC IV ne sont donc pas allés jusqu'au bout de l'approche multidimensionnelle de l'intelligence : ils ont adopté la logique factorielle pour définir les dimensions correspondant aux facteurs de groupe, mais ils sont restés dans la logique du QI global pour définir le QI total.

Les profils tracés dans les parties E et F de la feuille permettent de visualiser les écarts de performance entre les différents subtests d'une même dimension et entre les dimensions. Ces écarts renseignent sur les forces et les faiblesses relatives, mais ne doivent cependant pas être interprétés sans précautions. Deux informations doivent être prises en compte avant de se lancer dans l'interprétation d'une différence entre deux notes ; d'une part la valeur critique pour laquelle cette différence est significative à un seuil de risque raisonnable, d'autre part la fréquence avec laquelle cette différence a été observée dans l'échantillon de référence. Ces deux informations peuvent être trouvées dans les tables qui figurent en annexe du manuel d'administration. Ainsi, la valeur critique des différences entre indices se situant aux alentours de 15, seules les différences ICV- IRP et ICV-IVT sont significatives au seuil de .05 dans le protocole de la figure 9.3.. Des différences égales ou supérieures à celles-ci sont observées dans l'échantillon de référence dans seulement 5, 7% et de 9, 9% des cas respectivement.

C'est seulement une fois l'ensemble de ces opérations de cotation réalisées que commence le vrai travail du psychologue qui est celui d'interprétation de toutes ces données chiffrées. Il s'appuie alors pour cela sur ses connaissances théoriques, l'expérience clinique accumulée tout au long de sa pratique, et l'ensemble des autres éléments d'information recueillis sur la personne examinée au cours de l'épreuve et dans les autres composantes de l'examen psychologique. Une indication chiffrée seule, qu'il s'agisse d'une note d'indice ou un QI n'a en elle-même aucune signification.

2.2.3.4. Les qualités métriques de l'échelle

Les fidélités.

La consistance interne de l'adaptation française a été évaluée en utilisant la méthode pair-impair (voir chapitre 7, les fidélités). Le coefficient de corrélation pair-impair est de .94 pour l'échelle totale et varie de .84 à .89 pour les quatre notes composites d'indice factoriel. Le

coefficient de stabilité test-retest, avec un intervalle d'environ un mois entre les deux passations, est de .91 pour le QI total et va de .78 à .88 pour les quatre indices factoriels. La fiabilité de l'échelle est donc satisfaisante.

Les validités.

La publication de l'adaptation française étant encore très récente au moment où ces lignes sont écrites, les seuls indices de validité critérielle dont on dispose à son sujet sont les corrélations avec d'autres échelles d'intelligence. Pour ce qui concerne le QI total, la corrélation entre la WISC IV et la WPPSI (estimée avec un échantillon de 60 enfants de 6-7ans qui ont passé les deux échelles) est de .84. La corrélation avec la WAIS III (estimée avec un échantillon de 55 adolescents de 16 ans) est de .83. La corrélation avec la WISC III (la version précédente de la WISC) est de .78. Cette corrélation plus basse s'explique sans doute par les changements opérés dans la composition de l'échelle : l'indice de compréhension verbale conserve une corrélation forte avec le QI verbal de la WISC III (.82), tandis que l'indice de raisonnement perceptif corrèle plus faiblement avec le QI de performance de la WISC III (.62). Les corrélations avec les QI d'une autre échelle d'intelligence pour enfants, le K-ABC se situent aux environs de .70. La WISC IV mesure donc à peu près la même chose que les autres échelles d'intelligence, mais il y aurait un risque évident de circularité à ne valider chaque échelle d'intelligence que par sa corrélation avec les autres échelles d'intelligence.

Ce risque de circularité est moindre en prenant d'autres critères que les tests d'intelligence. La réussite scolaire est un des critères fréquemment utilisés pour les échelles d'intelligence qui sont conçues pour les enfants d'âge scolaire. Avec les versions précédentes de la WISC, les corrélations entre le QI total et la réussite scolaire variaient de .65 à .75 (Zhu et Weiss, 2005). Ces informations ne sont pas encore disponibles pour la WISC IV mais il sera intéressant d'examiner les corrélations entre les différents indices que cette échelle permet de calculer et la réussite scolaire.

La validité théorique ou validité « de construit » (voir chapitre 7, la validité des observations) est ici évaluée par les méthodes d'analyse factorielle. Le test est jugé valide si l'analyse factorielle des subtests donne bien lieu aux quatre facteurs attendus et si chacun des subtests a bien sa saturation la plus forte dans le facteur dont il est censé relever.

Insérer ici le tableau 9.1

L'examen du tableau 9.1, qui donne le résultat de l'analyse factorielle exploratoire de l'adaptation française montre que cette exigence est satisfaite. Il est par contre dommage que le manuel, dans l'état actuel des choses, ne spécifie pas les corrélations entre ces facteurs .

2.2.4. Discussion

L'évolution qui a été ci-dessus illustrée à partir de l'exemple des échelles de Wechsler a pris la même forme pour toutes les échelles d'intelligence. Toutes ont été progressivement modifiées pour se calquer, plus ou moins rapidement et à un degré plus ou moins important, sur le modèle CHC de la structure de l'intelligence. La révision de 1986 du Stanford-Binet avait déjà introduit une évaluation distincte de trois des dimensions du modèle, l'intelligence fluide, l'intelligence cristallisée et la mémoire à court terme. La révision de 2003 a élargi à 5 le nombre de dimensions évaluées à la fois dans le domaine verbal et dans le domaine non verbal (Gf, Gc, Gv, Gq et mémoire de travail). L'échelle de Woodcock-Johnson est celle qui a joué le rôle pionnier dans cette évolution (cf. McGrew, 2005). Sa structure a été entièrement repensée lors de la révision de 89, en la calquant sur celle du modèle CHC (Carroll et Horn ont participé à cette révision): les huit dimensions correspondant aux huit facteurs de la strate II du modèle CHC y étaient évaluées par deux tests chacune. La révision de 2001 a enrichi cette structure multidimensionnelle en incluant des tests correspondant à deux ou trois des facteurs étroits de la strate I pour chacun des facteurs de la strate 2. Il n'existe pas d'adaptation française de ces deux échelles américaines. Par contre, il existe une adaptation française du K-ABC , une échelle initialement construite en s'inspirant de la distinction de Luria entre processus séquentiels et simultanés, dont la dernière révision, le K-ABC II, donne la possibilité d'interpréter les résultats dans le cadre du modèle CHC. Par ailleurs, dans toutes ces échelles comme dans celles de Wechsler des subtests permettant d'évaluer l'efficacité de la mémoire de travail ont été inclus. L'évolution des échelles d'intelligence vers une évaluation multidimensionnelle de l'intelligence est donc bien un mouvement de fond.

2.3. Conclusion.

Les premières échelles de QI, celle de Binet et celles de Wechsler reposaient sur une conception empirique et pragmatique de l'intelligence. En faisant porter l'évaluation sur les processus supérieurs et en situant les performances d'une personne par rapport aux performances observées dans son groupe d'âge, ces pionniers ont mis le doigt sur une procédure efficace de comparaison des productions du système cognitif. Sous réserve que la

personne examinée ait bien baigné dans le milieu culturel pour lequel le test a été conçu, ces productions du système cognitif renseignent, indirectement, sur le fonctionnement global de ce système sans qu'il soit nécessaire pour cela d'en comprendre les mécanismes. Si un enfant a acquis en temps et en heure les connaissances et les habiletés qui sont d'habitude acquises par la plupart des enfants à cet âge là, c'est que son cerveau fonctionne normalement et on peut pronostiquer que, sauf accident, il acquerra en temps et en heure les connaissances et habiletés cognitives plus complexes. Cette forme d'évaluation globale de l'efficacité intellectuelle a rendu de nombreux services. Elle a, par exemple, permis de diagnostiquer les retards ou les avances dans le développement intellectuel ou de pronostiquer les chances de réussite dans tel ou tel type d'apprentissage. Fortes de ces succès, qui reposaient pour une bonne part sur leur approche pragmatique de l'évaluation, les échelles de QI sont restées pendant longtemps inchangées et insensibles aux évolutions théoriques sur l'intelligence. Les théories passaient, mais les tests restaient (Lautrey, 2001). La limite de cette forme d'évaluation globale était cependant de ne fournir que des indications très limitées lorsqu'il s'agissait d'analyser les raisons pour lesquelles le fonctionnement cognitif d'une personne était déficient. L'analyse plus poussée des performances au test devait alors s'appuyer sur l'expérience clinique du psychologue, qui ne reposait pas toujours sur des fondements solides.

L'évolution qui a été décrite dans ce chapitre correspond au passage de cette conception empirique et pragmatique de l'évaluation de l'intelligence à la prise en compte, dans la conception même des échelles d'évaluation, des apports de deux courants théoriques qui ont contribué à faire évoluer les idées sur l'intelligence au cours du siècle dernier : l'approche factorielle et la psychologie cognitive. Le premier de ces deux courants a contribué à mettre à jour les différentes dimensions de l'intelligence et à en catégoriser les principaux aspects, du moins ceux sur lesquels les individus se différencient le plus nettement. Le second a contribué à mettre à jour certains des processus sous-jacents à aux différents facteurs de l'intelligence, la relation trouvée entre l'efficacité de la mémoire de travail et le facteur d'intelligence fluide en est l'exemple le plus net. La poursuite de cette évolution, qui n'en est qu'à ses débuts, devrait doter les psychologues praticiens d'instruments d'évaluation donnant des fondements plus solides à leur analyse clinique des points forts et des points faibles du fonctionnement cognitif des personnes examinées ainsi qu'à la conception des méthodes de remédiation. Le QI, indicateur chiffré unique avec lequel on a cru pouvoir caractériser l'intelligence d'une personne, était une notion qui correspondait à la conception unidimensionnelle de

l'intelligence. Il n'est plus adapté à la conception multidimensionnelle qui prévaut maintenant et devrait donc être remplacé par des indicateurs plus appropriés (voir Lautrey, 2005b).

Lectures conseillées :

Grégoire, J. (2004). *L'examen clinique de l'intelligence de l'adulte*. Sprimont: Mardaga
Huteau, M., & Lautrey, J. (2003). *Evaluer l'intelligence - psychométrie cognitive*. Paris : PUF
Wechsler, D. (2005a). *WISC IV. Manuel d'interprétation*. Paris: Editions du Centre de
Psychologie Appliquée.
Wechsler, D. (2005b). *WISC IV. Manuel d'administration et de cotation*. Paris : Editions du
Centre de Psychologie Appliquée.

Légendes des figures

(Les figures 9.1 et 9.2 sont jointes sur papier et à composer)

Figure 9.1. Principales étapes de l'évolution des modèles de la structure factorielle de l'intelligence : 1a) le modèle du facteur général de Spearman, 1b) le modèle des aptitudes mentales primaires de Thurstone (PMA = Primary Mental Abilities), 1c) le modèle hiérarchique de Cattell et Horn, 1d) le modèle de Carroll

Les cercles représentent les facteurs latents, les carrés représentent les mesures manifestes (tests T1, T2, etc.). Les flèches unidirectionnelles représentent les saturations des facteurs, les flèches bidirectionnelles incurvées représentent les corrélations entre les facteurs

Figure 9.2. Liste des facteurs figurant dans les trois strates de la structure factorielle hiérarchique du modèle de Carroll (adaptée de Carroll, 1993, p. 626).

Figure 9.3. Page de garde du cahier de passation de la WISC IV, où sont reportés et synthétisés les résultats (d'après le Manuel d'administration de la WISC IV, p. 50 ; figure reproduite avec l'autorisation des ECPA).

Tableau 9.1. Saturation factorielle des subtests principaux dans l'échantillon de l'adaptation française (d'après le manuel d'interprétation de la WISC IV, Wechsler, 2005, tableau 5.3, p. 47).

Tous âges (N = 1103)				
Subtests	Facteurs			
	Compréhension Verbale	Raisonnement Perceptif	Mémoire de Travail	Vitesse de Traitement
Similitudes	.71	.11	.01	.00
Vocabulaire	.76	.04	.03	.02
Compréhension	.70	-.05	.00	.01
Cubes	-.03	.54	-.01	.12
Identification de concepts	-.03	.54	-.01	.12
Matrices	-.01	.64	.07	-.04
Mémoire des chiffres	.03	.03	.63	-.05
Séquence Lettres-Chiffres	.01	-.03	.64	.09
Code	.01	-.02	.01	.66
Symboles	.01	.06	.02	.65

WISC-IV

ECHELLE D'INTELLIGENCE DE WECHSLER
POUR ENFANTS — 4^{ème} EDITION

Nom MARTIN Prénom Pierre
 Sexe M Classe 5^{ème} Age 12 ans 3 mois
 Etablissement Collège Jules Ferry
 Psychologue C. Durand

Calcul de l'âge de l'enfant

	Année	Mois	Jour
Date de passation	2004	06	05
Date de naissance	1992	02	18
Age	12	03	24

B Conversion des notes brutes en notes standard

Subtests	Notes brutes	Notes standard			
Cubes	38	9	9	9	9
Similitudes	27	12	12	12	12
Mémoire des chiffres	15	10	10	10	10
Identif. de concepts	18	8	8	8	8
Code	49	9	9	9	9
Vocabulaire	42	13	13	13	13
Séq. Lettres-Chiffres	20	11	11	11	11
Matrices	22	9	9	9	9
Compréhension	26	12	12	12	12
Symboles	22	8	8	8	8
(Complét. d'images)			()	()	()
(Barrage)	84	9	9	9	9
(Information)			()	()	()
(Arithmétique)	24	10	10	10	10
(Raisonnement verbal)			()	()	()

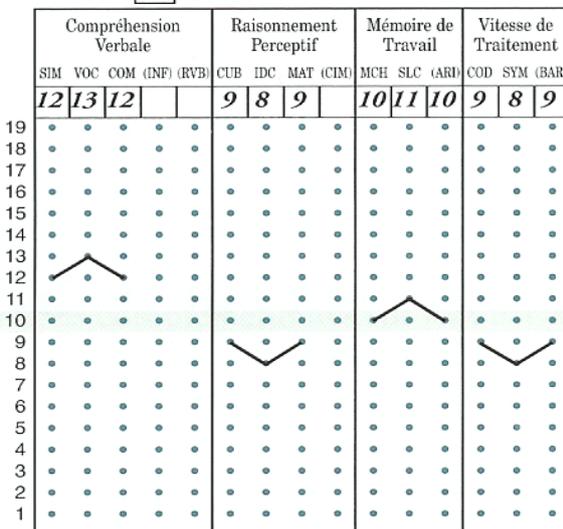
C Sommes des notes standard	37	26	21	17	101
	Comp. Verb.	Rais. Perc.	Mém. Trav.	Vit. Trait.	Totale

D Conversion des sommes des notes standard en notes composites

Echelle	Somme des notes standard	Note composite	Rang percentile	95% Intervalle de confiance
Compréhension Verbale	37	ICV 114	82	104-121
Raisonnement Perceptif	26	IRP 90	25	82-100
Mémoire de Travail	21	IMT 103	58	94-112
Vitesse de Traitement	17	IVT 90	25	82-101
Totale	101	QIT 101	53	94-108

Cahier de passation

E Profil des notes standard



F Profil des notes composites

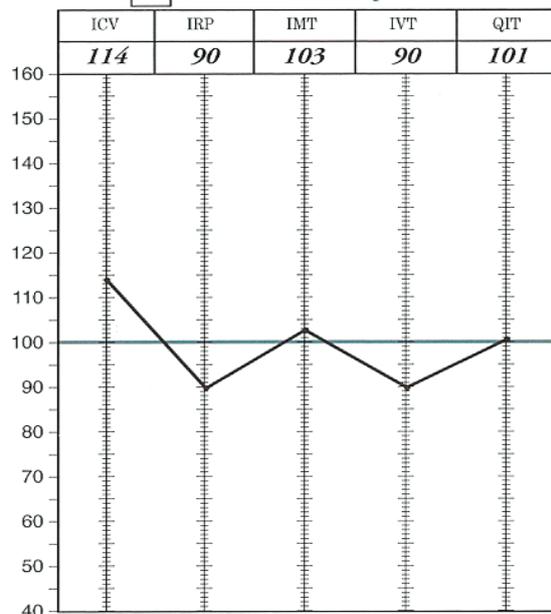


Figure 9.3. Exemple de page de garde du cahier de passation remplie (d'après le Manuel d'administration de la WISC IV, p. 50 ; figure reproduite avec l'autorisation des ECPA).