

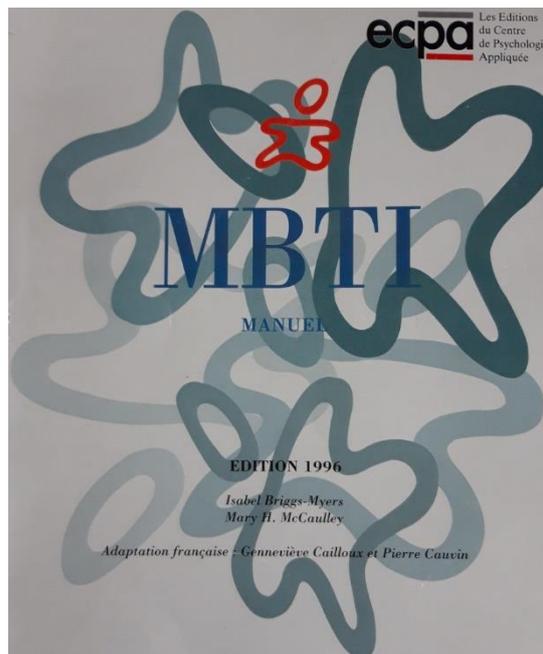
FICHE TECHNIQUE

MBTI analyses psychométriques

V1.0

15/05/2024

Fabien BELTRAME, Ph.D.



Si vous avez téléchargé ce document depuis une autre source, assurez-vous d'avoir la dernière version en visitant la page d'origine : <http://deporientation.free.fr/Ressources/Ressources.html>

Mise à disposition, en version numérisée, d'un rapport d'analyses psychométriques indépendantes de la version G du MBTI

1 Contexte

Plusieurs éléments sont à l'origine de la rédaction de cette fiche technique :

1. Une confusion entre la théorie typologique de Carl Gustave Jung et le test de personnalité MBTI
2. L'absence de transparence sur les qualités psychométriques du MBTI
3. Une polarisation des tenants et détracteurs du MBTI souvent sans arguments scientifiques vérifiés

Je dispose dans mes archives, d'un rapport intitulé « Analyse critique de quelques outils utilisés dans l'évaluation des compétences : considérations psychométriques, théoriques et pratiques ». Il s'agit d'un rapport de recherche réalisé pour le compte de l'Union Européenne (programme LEONARDO, projet CLEVER). Ce rapport est proposé par Paul Dickes, Jean-Luc Kop et Jocelyne Tournois du Groupe d'Analyse Psychométrique des Conduites (GraPCo - Laboratoire de Psychologie - EA-2337 - Université Nancy 2). Datant de 1997, ce document est imprimé au format papier et sa version « en ligne » n'existe pas.

Les test étudiés dans ce document sont principalement des outils utilisés en bilan de compétences : MBTI, SOSIE, 16 PF 5, etc ... (liste exhaustive dans l'introduction).

Ce rapport est maintenant assez ancien. Néanmoins, il conserve un intérêt pour sa partie consacrée au MBTI. En effet, ce test de personnalité s'appuie en grande partie sur la proposition de typologie de Carl Gustave Jung et nous ne disposons pas d'analyses psychométriques permettant de valider ou invalider ce modèle de la personnalité.

Les conclusions, apportées dans ce rapport, nous éclairent sur ce qui est valide et ce qui ne l'est pas. Ainsi, les praticiens peuvent mieux adapter leur utilisation de cet outil en connaissant ses limites démontrées par une étude indépendante. Même s'il existe de nouvelles versions du MBTI développées depuis 1997, ces études reposent sur le fondement théorique de base de l'outil : la typologie de Jung légèrement augmentée par les auteurs du MBTI.

Dans cette fiche technique, vous trouvez le chapitre 1 que j'ai donc numérisé et mis en forme à l'identique du document papier.

2 La confusion entre la typologie de Jung et le MBTI

Sans entrer dans un détail hyper technique, Jung n'était pas psychométricien. Il ne lui serait donc jamais venu à l'esprit de proposer un test pour identifier un profil psychologique selon le modèle qu'il propose. Par ailleurs, il diffuse sa typologie en 1921. Autant dire qu'à cette époque les tests psychométriques en sont à leur balbutiement.

Le MBTI a été conçu par Katharine Cook Briggs et Isabel Briggs Myers. Mais elles n'étaient pas spécialistes de la psychométrie. Par contre, elles ont énormément étudié la typologie de Jung. Comme expliqué sur le site du MBTI « Le développement de l'indicateur MBTI s'est déroulé sur plusieurs années. Il a été initié par Katharine Briggs sur la base de la théorie de Carl Jung, développé par Isabel Myers, et à sa mort, poursuivi par plusieurs équipes de psychologues, y compris l'équipe de recherche de The Myers-Briggs Company (anciennement CPP, Inc.). »¹

Le MBTI, c'est 4 dimensions :

- ✓ Extraversion–Introversion (E - I)
- ✓ Sensation–Intuition (S - N)
- ✓ Pensée–Sentiment (T - F)
- ✓ Jugement–Perception (J - P)

Les 3 premières dimensions sont directement issues de la typologie de Jung. Et la 4ème (J - P) est une proposition des auteurs Katharine Cook Briggs et Isabel Briggs Myers. Le MBTI n'est donc pas une opérationnalisation, stricto sensu, de la théorie des types psychologiques, mais bien un « test dont les soubassements théoriques s'appuient sur le concept de préférence de Jung ».

Par ailleurs Jung parle de préférence, de fonction dominante et inférieure, il ne formalise pas 16 types distincts composés de toutes les combinaisons des 8 lettres. Jung présente plutôt 4 types extravertis et 4 types introvertis :

Sensation extravertie
Intuition extravertie
Pensée extravertie
Sentiment extraverti

Sensation introvertie
Intuition introvertie
Pensée introvertie
Sentiment introverti

Il n'existe donc pas de parallélisme strict entre la typologie jungienne et les types MBTI

3 L'absence de transparence sur les qualités psychométriques du MBTI

Les données de validation d'un test psychométrique sont généralement présentées dans le manuel technique du test. Mais ce manuel technique est rarement disponible gratuitement. Il est généralement commercialisé par l'éditeur du test.

¹ <https://eu.themyersbriggs.com/fr-FR/Knowledge-centre/MBTI-Facts>

C'est exactement le cas pour le MBTI. Et pour accéder au manuel technique il faut suivre le schéma suivant : participer à une formation (payante) à l'utilisation du test, puis après obtention de la certification délivrée par l'éditeur du test, le certifié peut acheter le matériel du test et donc le fameux manuel technique.

Pour le MBTI, nous pouvons trouver, sur le site de l'éditeur, quelques documents, en accès libre, qui présentent des chiffres. Mais les analyses restent à un niveau très pauvre. Par exemple, dans le document « MBTI step I instrument – European Data Supplement – december 2016 »² nous allons trouver des statistiques descriptives de la population de répondants ; les alpha de consistance interne ; les intercorrélations entre les dimensions et les effectifs de répondants dans les 16 types. Autant dire que ces données, relativement basiques, ne permettent pas d'attester de la validité scientifique du MBTI.

Et même le manuel technique ne contient pas toujours l'exhaustivité des analyses psychométriques nécessaires à la connaissance suffisante pour valider psychométriquement l'instrument. Par exemple, dans le manuel technique du MBTI pour la version G (1996) qui sert de référence aux auteurs du rapport qui suit, sur 83 pages, seules 8 présentent des données quantitatives françaises. Le reste des pages présentent les dimensions du MBTI, les types, le concept de préférence, etc ...

Cependant, il ne faut pas non plus stigmatiser les éditeurs de tests. Les analyses psychométriques d'un test sont une arme à double tranchant : elles attestent de la qualité de l'outil, mais elles peuvent aussi révéler des failles ou même des secrets de fabrication. Et on peut tout à fait comprendre que l'éditeur ne souhaite pas les mettre sur la place publique au risque de voir des individus peu scrupuleux se les approprier pour fabriquer une contrefaçon du test. Tout est donc question de mesure.

4 La polarisation des tenants et détracteurs du MBTI

Le monde ne se divise pas en deux catégories mais en au moins quatre :

1. Les admirateurs de la typologie de Jung et de ce qu'ils pensent être son outil : le MBTI
2. Les adeptes inconditionnels du MBTI
3. Les détracteurs du MBTI souvent sans arguments solides
4. Les adorateurs experts de Jung, souvent psychologues, et dénigreur du MBTI

En parcourant les réseaux sociaux, les forums et ouvrages en librairie sur le thème de Jung et le MBTI, vous avez pu constater qu'il n'y a pas unanimité.

Par exemple, sur « Le forum français du MBTI - MBTI, Jung, Myers Briggs, psychologie, sciences humaines, tests, individuation »³ vous trouvez toutes sortes de personnes en quête d'une meilleure compréhension d'elles-mêmes. Profanes de la psychologie et souvent de Jung, ils échangent sur leur type et leurs comportements pour s'assurer, d'une part qu'ils ont bien identifié leur type et d'autre part, de leur bonne compréhension des caractéristiques de l'assemblage de leurs 4 lettres.

D'un autre côté, sur le réseau Reddit, on échange sur le caractère de pseudoscience⁴ du MBTI. Ou encore, dans cet article, l'auteur dénigre le MBTI en expliquant que le critère de fidélité est invalide en fournissant comme argument ... son propre cas individuel⁵. Cela va sans dire, mais cela va mieux en le disant : il faut une population de référence d'au moins 3 fois le nombre d'items du test pour

² https://eu.themyersbriggs.com/fr-FR/-/media/Files/PDFs/Technical-information/mbti_step_i_european_data_supplement.pdf

³ <https://mbti.forumactif.fr/>

⁴ https://www.reddit.com/r/changemyview/comments/68xs9v/cmvm_bti_is_pseudoscience_and_deserves_the_same/?tl=fr

⁵ <https://www.psychologytoday.com/intl/blog/give-and-take/201309/goodbye-to-mbti-the-fad-that-wont-die>

estimer la fidélité. Donc pour le MBTI il faut le calculer à partir de 89 fois 3 égal 267 personnes ... et pas généraliser à partir d'un seul cas. Spoiler : dans le rapport qui suit, le critère de fidélité est validé.

Du côté de la littérature, vous trouvez le « Deviens qui tu es » de Geneviève Cailloux et Pierre Cauvin régulièrement actualisé, qui vous aide à comprendre la personne que vous êtes à la lumière des types du MBTI. Mais vous trouvez également le « Arnaques & Impostures » de Patrick Vareille, qui ne mâche pas ses mots contre le MBTI et autres outils typologiques.

Bref, tous les camps s'expriment lorsqu'il s'agit de point de vue sur le test de personnalité MBTI.

Cette fiche est donc une sorte de réédition des analyses psychométriques, indépendantes de tout éditeur de test, pour apporter des arguments scientifiques à l'usage du MBTI.

5 Synthèse des éléments saillants du rapport

- A. Les analyses factorielles confirment les résultats présentés dans le manuel technique du MBTI : les quatre dimensions extraites des analyses sont validées (E - I ; S - N ; T - F ; J - P)
- B. Ces 4 échelles sont similaires à 4 dimensions du modèle de personnalité en 5 facteurs :
 - a. l'échelle extraversion - introversion : Facteur Extraversion du Big Five
 - b. l'échelle sensation - intuition : Facteur Ouverture du Big Five
 - c. l'échelle pensée - sentiment : Facteur Agréabilité du Big Five
 - d. l'échelle jugement - perception : Facteur Consciencieux du Big Five
- C. Rien ne justifie le système de cotation extrêmement complexe des scores. Une cotation simple et classique d'ajout de point en fonction du seul choix du répondant aboutit aux mêmes résultats pour les analyses psychométriques.
- D. Aucune classification en termes de typologie n'est validée. Les auteurs ont analysé la classification en 16 types proposée par Isabel Myers et Katharine Briggs, mais également la classification en huit types de Jung, en quatre tempéraments de Keirsey et Bates : aucune de ces typologies n'a reçu de confirmation empirique par les analyses pratiquées.
- E. Considérant que ces typologies ne reposaient que sur des critères a priori sans justification empirique, les auteurs du rapport proposent d'autres systèmes de classification en utilisant une démarche psychométrique exploratoire. Les analyses conduisent à retenir deux classifications possibles : l'une en cinq groupes, l'autre en dix types.

Implication pour les utilisateurs du MBTI :

- 1. L'utilisation du MBTI est justifiée et validée
- 2. L'évaluation des préférences, au sens de Jung, est validée
- 3. Une utilisation indépendante des 4 dimensions est pertinente
- 4. La référence à un des 16 types pour une personne est hasardeuse

Pour aller plus loin :

<https://eu.themyersbriggs.com/fr-FR>

<https://www.capt.org/>

Université Nancy 2
Groupe d'Analyse Psychométrique des Conduites
Laboratoire de Psychologie (EA-2337)

Analyse critique de quelques outils
utilisés dans l'évaluation des compétences.
Considérations psychométriques,
théoriques et pratiques.

Paul Dickes, Jean-Luc Kop et Jocelyne Tournois

Décembre 1997

Rapport de recherche réalisé pour le compte de l'Union
Européenne (programme LEONARDO, projet CLEVER)

Avant-propos

L'étude présentée dans ce document fait partie du projet CLEVER (Concept et Logique d'Evaluation des compétences : Etude et Recherche). Le projet, cofinancé par le programme LEONARDO de l'Union Européenne et piloté par l'INFPC (Institut National pour le développement de la Formation Professionnelle Continue de Luxembourg) a pour principal objectif l'évaluation des différentes méthodes d'analyse des compétences et du potentiel du capital humain. Le GRAPCO (Groupe d'Analyse Psychométrique des Conduites) du Laboratoire de Psychologie de l'Université Nancy 2 est l'un des partenaires de ce projet.

Ce rapport présente les analyses psychométriques d'un certain nombre d'outils psychologiques, employés couramment dans les bilans de compétences. Il s'agit de quatre tests d'aptitudes et de trois questionnaires de personnalité.

Les données sur lesquelles s'appuient ces analyses ont été mises à notre disposition par les Editions du Centre de Psychologie Appliquée (ECPA) auxquelles nous exprimons nos plus profonds remerciements.

Les critiques que l'on trouvera au fil du texte relatives à tel ou tel outil ne doivent pas faire oublier que la mise au point d'un instrument psychologique est une entreprise longue et difficile. Par la confiance qu'elles nous ont accordée, les ECPA ont montré tout l'intérêt qu'elles portent à l'amélioration de leurs outils. Les auteurs de ce rapport se réjouissent d'une telle attitude et espèrent, modestement, y contribuer.

dont les données expérimentales nous ont été aimablement fournies par les ECPA (Les Editions du Centre de Psychologie Appliquée). Sans leur concours, les analyses secondaires auxquelles on a procédé n'auraient pas été possibles et nous tenons à présenter aux Editions du Centre de Psychologie Appliquée nos plus vifs remerciements.

Introduction

L'objectif de ce rapport est de procéder à des analyses psychométriques et de tests et questionnaires utilisés comme outils d'évaluation dans le cadre des bilans de compétence.

Il n'est pas question d'envisager l'analyse exhaustive des très nombreux outils disponibles sur le marché. Seuls quelques outils prototypiques seront analysés dans un cadre de référence psychométrique. Cette évaluation suppose notamment la vérification des informations psychométriques fournies par l'éditeur de l'outil et par des analyses complémentaires permettant de préciser la signification de ce qui est mesuré par chacun des outils et/ou d'apporter à l'utilisateur de nouvelles aides à l'interprétation.

1. Cadre de référence théorique

Les bilans de compétence ont recours à des techniques et à des outils hétérogènes s'inscrivant dans des cadres théoriques et méthodologiques diversifiés. Ainsi, Joras (1995) distingue quatre grandes catégories d'outils : les études sur dossiers, les entretiens, les tests ou questionnaires et les essais ou simulations. Dans chacune de ces catégories, on peut encore distinguer des sous-catégories et certains préfèrent élaborer une typologie plus fine. C'est le cas par exemple de la typologie proposée par l'INFPC (1996)¹ dans laquelle on trouve huit catégories qui peuvent sans grande difficulté trouver leur place dans les catégories élaborées par Joras.

L'une des questions fondamentales concerne *la fiabilité des techniques d'évaluation utilisées*. Le cadre législatif élaboré lors de la mise en place des bilans de compétence insiste tout particulièrement sur ce point (cf. article R. 900-4 ; c'est nous qui soulignons) : "[les méthodes et techniques utilisées]... doivent avoir fait la preuve de leur pertinence parce qu'élaborées à partir de théories validées par des pratiques professionnelles, ou par l'intermédiaire de méthodes scientifiques d'étalonnage". Selon Gérard Lyon-Caen (1991), les techniques utilisées doivent faire preuve des quatre qualités suivantes :

- fidélité (les résultats doivent être indépendants de celui qui met en œuvre la technique et présenter une constance temporelle substantielle) ;
- pouvoir discriminant (aptitude à classer) ;
- pertinence ;
- validité scientifique (prédictivité éprouvée).

¹ Les huit catégories distinguées par HNFPC sont : l'entretien, les itinéraires autobiographiques, les mises en situation (essais professionnels, stages d'observation), les techniques informatisées, les techniques de groupe, le portefeuille de Compétences, l'analyse de l'environnement.

Si la question de la fiabilité des techniques est une question fondamentale, il reste à trouver un cadre général à l'intérieur duquel une évaluation des techniques utilisées dans les bilans de compétence peut prendre place. L'un de ces cadres généraux est, nous semble-t-il, celui de la psychométrie. Dans un ouvrage publié récemment (Dickes, Tournois, Flieller & Kop, 1994) nous avons défini la psychométrie comme l'ensemble des théories et des méthodes liées à la mesure en psychologie. Les bilans de compétences utilisent des outils dont l'objectif est de fournir une mesure. Le cadre de référence psychométrique s'impose donc naturellement.

Notre conception de la psychométrie est beaucoup plus générale que le rôle instrumental qu'on lui assigne trop souvent. Ce rôle instrumental se limite à la construction et à l'étude des tests. Mais l'objet de la psychométrie est bien plus large, car les tests ne sont qu'une classe d'instruments de mesure parmi d'autres. De plus, la démarche psychométrique est fondamentalement une démarche théorique, ce qui est trop souvent oublié par les utilisateurs d'instruments de mesure. Lorsqu'on réalise une mesure, on teste un modèle théorique. L'enjeu de la psychométrie est dès lors *d'expliciter les hypothèses qui sous-tendent la théorie et de proposer des méthodes permettant de tester ces hypothèses.*

En termes psychométriques, la question de la qualité ou de la fiabilité des instruments de mesure s'exprime en termes de *validité*. On dit classiquement, dans une définition peu opérationnelle il est vrai, qu'un instrument est valide dès lors qu'il mesure ce qu'il est censé mesurer. Pour rendre cette définition opérationnelle, différentes stratégies sont utilisées. Ces stratégies, dénommées un temps *types de validité*, contribuent toutes à assurer la validation d'une mesure. A titre d'exemple, on peut citer la validité critérielle. On s'intéresse, dans ce cas, à la *valeur pronostique* (ou prédictive) d'un instrument de mesure. En ce qui concerne les bilans de compétence, la validité critérielle d'un instrument sera assurée si celui-ci permet de prédire la réussite (ultérieure) du projet professionnel du salarié. C'est cette idée que l'on retrouve dans la notion de "validité scientifique" évoquée par Gérard Lyon-Caen.

Mais *la validité critérielle n'est pas suffisante* pour s'assurer de la qualité d'une mesure. D'autres considérations doivent être prises en compte. Ainsi, imaginons que la couleur des yeux permette de prédire la réussite du projet professionnel. On devrait conclure que cette mesure est valide sur le plan critique. Mais est-ce pour autant une bonne mesure ? Certes non ! et notamment parce qu'elle serait mal acceptée par les utilisateurs. Mais il y a aussi une raison plus fondamentale qui fait que cette mesure n'est pas une mesure valide. Cette raison fait référence au modèle théorique sous-jacent à l'élaboration et à l'utilisation de toute mesure (on n'a aucune théorie, du moins à notre connaissance, qui prédit que la couleur des yeux est liée à la réussite du projet professionnel).

Afin de résoudre les difficultés d'une évaluation de la validité en termes critiques et de synthétiser les différentes stratégies proposées dans la littérature, nous avons proposé, dans l'ouvrage cité plus haut (Dickes et al., 1994), une *démarche générale de la validation* des mesures psychologiques.

Nous considérons que la démarche de validation est un processus continu qui adopte les mêmes règles que la démarche scientifique en général. Les stratégies utilisées dans le processus US de validation peuvent être de nature extrêmement variées, à condition toutefois qu'elles aient pour objectif la *vérification empirique d'hypothèses théoriques* concernant le concept que l'on cherche à mesurer (par exemple, connaissances dans une certaine matière, aptitude à la négociation, niveau de culture générale, etc.). Ainsi, tout processus de validation, comme toute démarche scientifique, repose sur un modèle théorique plus ou moins formalisé.

L'originalité de notre conception de la validation consiste à distinguer *deux types de modèles théoriques* qui correspondent à *deux formes de validation* : la *validation intraconcept* et la *validation interconcepts*. La première porte uniquement sur le concept que l'on cherche à mesurer la seconde porte sur les relations entre le concept mesuré et d'autres concepts. L'exemple de la validité critérielle renvoie à une validation interconcepts : en vérifiant par exemple que l'aptitude à la négociation (concept A) permet de prédire la réussite du projet professionnel (concept B), on met bien en relation deux concepts différents².

Outre la validité critérielle, l'autre grande stratégie utilisée dans la validation interconcepts est la *validité divergente*. Il s'agit de montrer que deux instruments censés mesurer deux concepts différents ne fournissent pas les mêmes résultats, ni quand on les confronte directement, ni quand on les confronte indirectement par l'intermédiaire d'une troisième variable. Autrement dit, pour montrer, par exemple, la validité divergente d'une mesure de l'aptitude à négocier par rapport à une mesure d'aisance verbale, il faut s'assurer que la corrélation entre les deux mesures est faible et que leurs corrélations avec d'autres variables ne sont pas du même ordre de grandeur (en vérifiant, par exemple, que l'aptitude à la négociation est fortement corrélée avec la réussite du projet professionnel alors que cette dernière n'est que faiblement corrélée avec l'aisance verbale).

Qu'il s'agisse de validité critérielle ou de validité divergente, la logique de la validation interconcepts repose sur le *test d'hypothèses dérivées d'un modèle théorique*. Il ne s'agit pas, bien évidemment, de confronter la mesure d'un concept avec celles d'autres concepts de manière aléatoire ou, plus prosaïquement, en se limitant aux concepts pour lesquels des instruments de mesure sont disponibles sur le marché. Le choix des autres concepts mis en relation avec celui que l'on cherche à mesurer doit être un choix justifié en fonction d'un modèle théorique.

Les considérations générales présentées à propos de la validation interconcepts (possibilité d'avoir recours à différentes stratégies, importance du modèle théorique sous-jacent) se retrouvent dans le cas de la *validation intraconcept*. A la différence de la précédente, celle-ci ne met pas en relation des concepts différents. On se focalise alors sur le dispositif utilisé pour mesurer le concept et en particulier, lorsque l'instrument de mesure est un test ou un questionnaire, sur les items qui le composent et sur les relations entre ces items.

La stratégie de validation intraconcept la plus importante est désignée habituellement sous le nom de validité de contenu. L'objectif est simple à comprendre. Il s'agit de vérifier que les items qui composent l'instrument de mesure sont pertinents et représentatifs du domaine couvert par le concept à mesurer. Malgré la simplicité apparente de cet objectif, celui-ci s'avère très difficile à atteindre. Comment montrer en effet que les items d'un questionnaire destiné à mesurer les aptitudes à négocier sont tous pertinents par rapport à ce concept et, surtout, qu'ils couvrent toutes les facettes de ce concept ? Certains proposent de faire appel à des collègues d'experts qui évaluent chacun des items d'un instrument en fonction de sa pertinence et évalue si des facettes importantes du concept n'ont pas été omises. D'autres préfèrent une démarche plus empirique consistant à utiliser des méthodes statistiques qui reposent sur la logique suivante : on calcule les corrélations entre chaque item et l'ensemble des autres items. Lorsque les corrélations sont faibles ou négatives, on décide que les items en question ne sont pas pertinents par rapport au concept mesuré.

² Cet exemple simple illustre, au passage, le fait que dans la validité critérielle, il faut disposer d'une mesure du critère (ici, la réussite du projet professionnel) et que la démarche de validation dans son ensemble doit s'appliquer aussi à la mesure du critère.

Ces deux démarches sont utiles mais elles restent insatisfaisantes. A nouveau, nous proposons d'adopter une démarche plus théorique. En simplifiant, disons qu'il s'agit de proposer, a priori, *une définition du concept mesuré qui soit formulée de manière à pouvoir être testée* (et donc éventuellement rejetée). Pour construire une telle définition, une analyse théorique du concept est indispensable. Celle-ci doit notamment préciser :

- les *différentes facettes du concept* (on peut supposer par exemple que l'aptitude à la négociation doit distinguer les domaines sur lesquelles portent la négociation) ;
- le niveau de *constance temporelle et situationnelle du concept* (l'aptitude à la négociation est-elle stable dans le temps et selon les situations, par exemple selon le nombre de personnes qui interviennent dans la négociation ?) ;
- les *processus cognitifs* qui sous-tendent les réponses aux items.

A travers les quelques exemples ci-dessus, on comprend que les définitions doivent être plus qu'un simple énoncé verbal vague et ambigu. En spécifiant les définitions de la manière que nous venons d'évoquer, *il est possible de les vérifier* comme n'importe quelle autre proposition théorique. Ce faisant, on contribue à assurer la validité de contenu des instruments de mesure.

Outre la validité de contenu, la validation intraconcept s'appuie aussi sur une stratégie cherchant à assurer *la convergence des mesures*. Ainsi, deux mesures d'un même concept utilisant des dispositifs différents (par exemple, changement d'examineur, changement d'items, etc.) doivent fournir des résultats équivalents.

La démarche de validation des instruments de mesure que nous venons d'esquisser correspond à la volonté d'évaluer la fiabilité des outils utilisés dans le cadre des bilans de compétences. Cette démarche intègre sans difficulté les différents critères de qualité listés dans la littérature (validité scientifique, pertinence, fidélité) Elle présente en outre l'avantage d'offrir un cadre plus général et plus systématique.

2. Démarche générale

Pour effectuer des analyses psychométriques d'outils standardisés, il est généralement nécessaire de disposer de bases de données importantes. Celles-ci sont très coûteuses à construire, à la fois en termes de temps et d'argent. Le projet CLEVER ne permettait pas d'envisager la constitution de telles bases de données.

Il a donc été envisagé, dans un premier temps, de demander le concours de centres de bilans de compétences. Comme ces derniers utilisent assez largement ces outils, la mise en commun des données provenant de plusieurs centres aurait pu permettre de constituer des fichiers suffisamment importants. Cette solution présentait de nombreux avantages. En particulier, elle garantissait la pertinence des outils analysés puisqu'ils auraient été, de fait, utilisés dans l'évaluation des compétences. Surtout, grâce au suivi des bénéficiaires, elle aurait fourni des données concernant la validité prédictive des outils analysés. Malheureusement, cette solution s'est très vite avérée impraticable pour des raisons légales. En effet, le texte de loi

précisant la mise en place du dispositif de bilan de compétences spécifie la stricte confidentialité des données recueillies et la destruction, par l'organisme, de ses archives³.

Devant ces difficultés, nous avons contacté les Editions du Centre de Psychologie Appliquée (ECPA) qui diffusent la majorité des outils standardisés utilisés dans l'évaluation des compétences. Les ECPA ont accepté de mettre à notre disposition les fichiers ayant servi à l'élaboration de quelques outils largement utilisés par les centres de bilan. Ces outils seront détaillés *infra*. Nous nous contenterons ici de donner les grandes lignes de la démarche méthodologique, sachant que des démarches spécifiques devront être adoptées pour l'analyse de chaque outil. Ces spécificités seront présentées en détail dans les chapitres propres à chacun des outils.

Les fichiers de données fournies par les ECPA ne permettent pas d'étudier de manière globale la validité des outils. Il manque, en particulier, des informations sur la valeur prédictive, que seules des enquêtes extrêmement lourdes permettraient d'obtenir. En termes psychométriques, c'est donc essentiellement la validité intraconcept des mesures qui sera étudiée. Cette limitation est certes dommageable mais elle ne manque pas d'intérêt.

Pour ne citer qu'un exemple, Beaufils (1994) insiste sur la nécessité d'effectuer des analyses secondaires pour évaluer la validité d'un outil au-delà des résultats publiés par l'éditeur du test ou du questionnaire qui se contente parfois d'un point de vue partial.

Au-delà d'un score général fourni par un test ou un questionnaire, il s'agit de s'interroger sur les différentes facettes mesurées par tel ou tel outil, sur les processus cognitifs mis en œuvre par les sujets pour répondre, sur les hiérarchies de compétences pouvant exister, sur l'adéquation entre ce que l'on croit que mesure tel outil et sur ce qu'il mesure effectivement, etc.

Outre les méthodes classiques (analyse de fidélité, analyse d'items, analyse factorielle) déjà mises en œuvre, la plupart du temps, par ceux qui ont conçu ou adapté ces outils, nous aurons recours à des méthodes sinon plus sophistiquées du moins plus spécialisées afin de répondre aux questions esquissées ci-dessus. Ces méthodes sont principalement au nombre de trois :

- Les modèles d'équations structurales de type LISREL (LInear Structural RELationships) ces modèles sont une synthèse des modèles d'analyses factorielles, des modèles en pistes causales, des modèles de régression linéaire et des modèles linéaires canoniques. La logique fondamentale qui sous-tend ces modèles est de tester des hypothèses inférées sur la base modèle théorique. De ce fait, ces modèles sont particulièrement utiles pour vérifier si le schéma théorique postulé pour la mesure de telle ou telle caractéristique est compatible avec ce qui est observé empiriquement. Cette particularité des modèles d'équations structurales fait que l'on considère qu'ils procèdent d'une logique

³ En fait, le décret d'application indique que tous les documents élaborés pour la réalisation du bilan doivent être aussitôt détruits, sauf demande écrite du bénéficiaire. Dans ce cas, ils ne peuvent être conservés que dans un délai maximum d'un an. Les situations dans lesquelles les bénéficiaires demandent expressément la conservation des documents sont extrêmement rares ce qui fait que, même si nous avons pu obtenir une autorisation exceptionnelle pour analyser ces données, les effectifs disponibles auraient été insuffisants.

confirmatoire par opposition à d'autres modèles plus classiques (comme les analyses factorielles) qui procèdent d'une logique exploratoire.

- Les modèles de réponses à l'item (MRI ou IRT : Item Response Theory). Dans ces modèles, on cherche à prédire la probabilité qu'un individu donne une réponse d'un certain type (e.g. une réponse correcte) à un item. Ces modèles se différencient selon le nombre de paramètres qui sont pris en compte. Les plus simples ne comportent qu'un seul paramètre : en simplifiant, disons que ces modèles supposent que la probabilité d'une réponse correcte à un item ne dépend que de la compétence de l'individu et de la difficulté des items ; les plus complexes comptent jusqu'à trois paramètres. Par rapport aux questions évoquées ci-dessus, ces modèles vont principalement nous permettre d'analyser l'existence d'éventuelles hiérarchies de compétences.
- Les modèles d'échelonnement multidimensionnel (EMD ou MDS : MultiDimensional Scaling). Les modèles d'échelonnement multidimensionnel ont pour principale finalité de représenter les relations entre des variables dans un espace à plusieurs dimensions en faisant en sorte que les variables soient d'autant plus proches qu'elles se ressemblent le plus. Comme avec les modèles d'équations structurales, il s'agit principalement d'étudier si le schéma théorique postulé pour la mesure d'une certaine caractéristique est compatible avec les observations. Techniquement, ces deux familles de modèles se distinguent surtout par le fait que les exigences mathématiques des EMD sont plus souples que celles des modèles d'équations structurales. Elles complètent donc utilement les informations fournies par ces dernières en autorisant la prise en compte d'autres formes de relations entre les variables.

L'ensemble de ces analyses permettront d'enrichir de façon notable les informations traditionnellement dispensées aux utilisateurs d'outils psychologiques par les maisons d'éditions. Ces informations complémentaires devraient permettre aux praticiens de l'évaluation d'opérer des choix plus éclairés parmi les outils qui sont à leur disposition.

3. Outils analysés

Les analyses psychométriques de tests et questionnaires utilisés dans le cadre de bilans de compétences constituent des analyses secondaires. La caractéristique principale de ce type d'analyse est de porter sur des données déjà recueillies. En l'occurrence, les populations concernées ont été définies par les éditeurs des outils. On trouvera ci-dessous, en même temps qu'une rapide description des instruments analysés, le descriptif des échantillons.

L'examen des commandes passées par les centres de bilans de compétences auprès des Editions du Centre de Psychologie Appliquée a permis d'arrêter une liste de 7 instruments devant faire l'objet d'analyses psychométriques complémentaires⁴. Ces instruments couvrent à la fois le champ conatif (personnalité, valeurs) et le champ cognitif.

⁴ Les ECPA nous ont également fourni les données correspondant au California Personality Inventory (version révisée). Malheureusement, le temps nous a manqué pour effectuer l'analyse de ce questionnaire.

Nous avons demandé aux ECPA de nous transmettre les fichiers avant toute manipulation (recodage des données manquantes, etc.). Ce souhait a été parfaitement exaucé. Transmis sous forme de fichiers texte accompagnés de leur grille de codification, ceux-ci ont été ensuite transformés par nos propres soins en vue des analyses psychométriques.

L'Indicateur Typologique de Myers-Briggs (MBTI, forme G, version révisée) est un questionnaire de personnalité de 89 items. Elaboré selon la théorie des types psychologiques de Jung, sa première version date, aux Etats-Unis, de 1942. En 1996, c'est la troisième adaptation française de ce questionnaire très utilisé dans le milieu professionnel qui est mise sur le marché.

L'objectif du questionnaire est de mettre en relief les préférences individuelles sur quatre échelles définies chacune par deux pôles opposés : extraversion - introversion ; sensation intuition ; pensée - sentiment ; jugement - perception.

Une typologie théorique est dégagée sur la base d'une dichotomisation de chacune des échelles. L'examen de la pertinence de cette typologie théorique constitue, là aussi, l'une des premières investigations psychométriques à effectuer.

Le **SOSIE** est un instrument dans la popularité est grandissante dans les centres de bilan. Il s'agit en fait de la réunion de trois inventaires élaborés par L.V. Gordon et destinés à mesurer : 9 traits de personnalité ; 6 valeurs interpersonnelles et 6 valeurs personnelles. Les items sont au nombre de 98 et présentent la particularité d'offrir aux sujets des choix entre des propositions équilibrées en termes de désirabilité sociale.

La lecture des résultats s'effectue à un double niveau : un niveau théorique distinguant quatre axes et un niveau empirique s'appuyant sur les dimensions dégagées par analyse factorielle. La mise en correspondance entre ces deux niveaux de lecture ne semble pas évidente. De nouvelles analyses permettront peut-être d'éclaircir cet aspect.

Le **16 PF 5** est la cinquième édition internationale du très célèbre " 16 Personality Factors" de Cattell qui comporte 185 items. Les principales innovations de cette nouvelle version, mise sur le marché en 1995, sont la possibilité de calculer des scores à cinq facteurs de second ordre et des indices de validité des réponses (désirabilité sociale, tendance à l'acquiescement, réponses rares).

Le **test d'évaluation de la pensée critique de Watson-Glaser** (C.T.A.) est composé de 80 items répartis en 5 subtests (inférence, reconnaissance d'hypothèse, déduction, interprétation, évaluation d'arguments).

Relativement peu connu, ce test, dont les ancrages théoriques sont flous, mérite probablement une analyse psychométrique critique afin d'en apprécier les qualités et les défauts.

Le **test des syllogismes** est directement issu des célèbres travaux de Guilford sur la structure de l'intelligence et s'inspire des travaux de Thurstone. Cet instrument est censé mesurer la capacité à l'évaluation logique à l'aide de 40 items présentés sous la forme de syllogismes.

Le **Test d'Efficiences des Représentations Mentales (TERM)** est, comme le précédent, un instrument élaboré en collaboration avec Guilford. Composé de 48 items rédigés selon une structure théorique parfaitement explicite, il mesure le raisonnement général (capacité à définir, comprendre, structurer un problème dans le but de sa résolution).

Les **Tests d'Evaluation des Aptitudes Intellectuelles (TEAI)**, dont la mise sur le marché français a lieu à la fin de l'année 1996, constitue une nouveauté importante dans le domaine des batteries d'intelligence.

Adapté d'un instrument anglais, les TEAI comportent trois épreuves destinées à mesurer les aptitudes verbales (60 items), numériques (33 items) et abstraites (115 items) et sont destinés à des personnes ayant un niveau d'éducation élevé.

CHAPITRE 1

LE MBTI : UN INDICATEUR TYPOLOGIQUE ?

1. Introduction

Présentation sommaire du MBTI

Le MBTI (indicateur typologique de Myers-Briggs) est un questionnaire de 89 items dichotomiques destiné à opérationnaliser la théorie des types psychologiques de Jung. Ce questionnaire est construit autour de quatre échelles qui reflètent les quatre préférences fondamentales qui structurent, selon Jung, la perception et le jugement. Le MBTI est très utilisé, tant aux Etats-Unis qu'en France : ses domaines d'utilisation concernent à la fois l'éducation, le conseil, l'orientation professionnelle mais aussi la gestion d'équipe et la communication.

Les 89 items de la forme G révisée publiée en 1996 sont regroupés en quatre échelles bipolaires :

- * l'échelle E-I (extraversion - introversion) : les extravertis sont préférentiellement orientés vers le monde extérieur et se concentrent sur les choses et les gens, alors que les introvertis privilégient le monde intérieur en se concentrant sur les concepts et les idées.
- * l'échelle S-N (sensation - intuition) : le pôle sensation de cette échelle renvoie au fait d'appréhender les faits et les choses à travers les sens ; le pôle intuition, quant à lui, serait caractérisé par le fait de saisir les significations, les relations, les possibilités de façon instinctive, sans intervention de l'esprit humain.
- * l'échelle T-F (pensée - sentiment) oppose deux critères de jugement : la pensée, qui privilégie les raisonnements logiques dans la prise de décision et le sentiment, qui privilégie les considérations personnelles ou sociales.
- * l'échelle J-P (jugement - perception) décrit le processus par lequel un individu aborde le monde extérieur : le jugement fait référence à la pensée ou au sentiment ; la perception fait référence à la sensation ou à l'intuition.

L'administration du questionnaire permet de calculer 8 scores (un score pour chacun des deux pôles de chaque échelle). Les auteurs du questionnaire privilégient une interprétation basée sur des préférences. Un individu aura une préférence pour l'extraversion si son score E est supérieur à son score I. Il en est de même pour les autres échelles. La combinaison des quatre préférences permet de construire 16 types psychologiques (ex : ISTJ, ENFP, ENTP, etc.) qui sont largement décrits dans le manuel du test.

Ainsi qu'on peut le constater, l'utilisation et l'interprétation du MBTI accordent une très faible place au caractère continu des échelles. Ainsi, un individu qui a un score de 10 à E et de 9 à I sera classé comme extraverti de la même manière qu'une autre personne ayant un score de 20 à E et de 1 à I. Autrement dit, il est fait très peu de cas de l'intensité avec laquelle s'exprime une préférence. Le manuel laisse à l'appréciation du professionnel l'utilisation de cette information, en proposant le calcul éventuel des quatre scores qui expriment "la clarté de la préférence" (p.56). Le manuel propose aussi, pour certaines études de corrélations qui le nécessiteraient, d'établir des scores continus aux quatre échelles. Le calcul de ces scores paraît un peu complexe. Pour chacun d'eux, il se fonde sur la différence des deux sous-scores aux

deux pôles de l'échelles, différence à laquelle est ajouté ou retranché un point (selon que l'un des pôles l'emporte sur l'autre), et qui est suivie d'une transformation linéaire. On est conduit à se demander ce qui nécessite une telle complexité.

Dans le même ordre d'idées, le système de codification des items (i.e. le mode de correction) présenté dans le manuel semble assez complexe. Les réponses sont codées 0, 1 ou 2 points selon les items et selon le sexe de l'individu. Il en découle que les hommes et les femmes ne sont pas toujours évalués sur la base des mêmes items. On peut se demander si la complexité d'un tel système est réellement justifiée sur le plan empirique, d'autant que les explications avancées pour expliquer ces différences de pondération dans le manuel ne paraissent pas très convaincantes.

Intérêts du MBTI

Cette présentation sommaire suffit à préciser les intérêts majeurs du MBTI. Ceux-ci résident dans sa spécificité. L'instrument est original sur plusieurs points :

- Il vise incontestablement à établir des "types" plutôt qu'à situer les personnes selon des dimensions. L'indicateur "ne mesure pas les gens", précise le manuel, "mais les trie en groupes auxquels, selon la théorie, ils appartiennent déjà" (sic, p.65).
- Il vise à appréhender des dimensions qui renvoient à des manières de faire, des façons de procéder. C'est "la manière de recueillir l'information et de prendre des décisions" dont on vise la description (p.65). On vise donc plus à appréhender "comment" le sujet se comporte qu'à décrire "ce" qu'il fait. Les dimensions appréhendées relèvent ainsi plus de "styles" que de "traits".
- Il repose sur l'hypothèse selon laquelle "les préférences naturelles existent réellement" (p.65) et s'intéresse donc davantage au fait qu'une personne préfère l'extraversion à l'introversion (ou l'inverse), plutôt qu'au fait qu'elle soit faiblement, moyennement ou fortement extravertie (ou l'inverse).

Autrement dit, ce qui caractérise essentiellement le MBTI, par rapport aux autres grands instruments de mesure de la personnalité (comme le 16PF de Cattell ou l'EPI d'Eysenck ou les instruments de mesure du modèle en cinq facteurs), c'est le fait qu'il exploite seulement des "préférences" sur des modes de perception ou de décision en vue de révéler des types de personnes.

Ces caractéristiques majeures du MBTI guideront la majorité des analyses complémentaires que nous allons entreprendre.

Plan des analyses

Les analyses portent sur les données ayant servi à élaborer la version française révisée du MBTI : 89 items renseignés par 731 sujets. Les caractéristiques de cet échantillon de sujets seront d'abord brièvement examinées.

Les analyses suivantes seront guidées par deux préoccupations : la première concerne la complexité du système de notation et la non moindre complexité du calcul éventuel des "scores continus" ; la deuxième concerne la validité de la typologie proposée, qui fait l'originalité majeure du MBTI. Ces questions sont en fait fortement liées, puisque la typologie s'appuie, il va de soi, sur les scores calculés : ces scores sont établis par un système de correction complexe, qui débouche sur des dimensions continues qui sont ensuite dichotomisées, les préférences pour l'un ou l'autre pôle de l'échelle fondant l'établissement de la typologie. Ainsi, nous nous interrogerons successivement sur les points suivants :

- les quatre échelles de base du MBTI sont-elles valides ?
- ces quatre échelles mesurent-elles plutôt des styles ou des traits ?
- la complexité de la notation se justifie-t-elle ?
- est-il justifié de séparer les scores aux quatre échelles en sous-scores ?
- les types du MBTI ont-ils une réalité empirique ?

2. L'échantillon de sujets

L'échantillon de sujets qui a renseigné les 89 items conservés dans la version définitive du questionnaire est présenté à la page 75 du manuel. Il y est décrit par quatre caractéristiques : le sexe, l'âge, le niveau d'études et la catégorie socio-professionnelle.

Le sexe

Les hommes (59,1%) sont plus fortement représentés, dans cet échantillon, que les femmes. Comme les analyses, sont, pour la plupart, effectuées séparément sur chacun des sexes, ce déséquilibre ne pose guère problème.

L'âge

L'âge moyen de l'échantillon (38 ans) est inférieur à celui la population des adultes vivant en France. Ceci reflète surtout une sous-représentation des personnes âgées : seuls, 3,9% des sujets de l'échantillon ont 55 ans et plus ; en leur sein, seulement 5 (soit 0,7%) ont 60 ans ou plus, et un seul a plus de 70 ans.

Les analyses complémentaires effectuées sur cet échantillon montrent, de plus, une liaison non négligeable entre l'âge et le sexe. Elle s'exprime aussi bien sur la distribution complète des âges (de 13 à 71 ans) que sur une classification des âges en catégories. Le tableau 1.1 présente cette liaison sur la base de la classification de l'échantillon en cinq catégories d'âge, classification adoptée dans le manuel.

	13-24 ans	25-34 ans	35-44 ans	45-54 ans	55 et plus
hommes	13	98	160	118	24
femmes	42	118	74	49	3
ensemble	55	216	234	167	17

Tableau 1.1 : Répartition de l'échantillon de sujets selon l'âge et le sexe
(Chi² à 4 dd1 = 72,92 ; p < .00001 ; phi = 0,32)

Dans ce tableau, on observe une majorité de femmes de moins de 35 ans et une majorité d'hommes de plus de 35 ans. Autrement dit, les femmes sont plus jeunes que les hommes. Leurs moyennes d'âge sont respectivement de 34 ans 1/2 et de 40 ans.

Le niveau d'études

Le niveau d'études n'est connu que pour 491 des sujets de l'échantillon (soit 47,2%). La répartition du niveau d'études, présentée dans le tableau 2, montre à l'évidence que cet échantillon est fortement éduqué : les personnes ayant effectué cinq années d'études supérieures ou plus, après le baccalauréat, constituent à elles-seules plus du tiers de l'échantillon.

	CAP / BEP	Bac ou niveau bac	Bac +1, bac +2	Bac + 3, Bac + 4	Bac + 5 et plus
effectif	67	80	83	85	176
pourcentage	13,8	16,3	16,9	17,3	35,8

Tableau 1.2 : Répartition du niveau d'études dans l'échantillon (effectif = 491 ; 240 réponses non indiquées)

La catégorie socio-professionnelle

Compte tenu des informations précédentes relatives au niveau d'études, d'une part, à l'âge, d'autre part, on peut s'attendre à ce que la distribution des catégories socio-professionnelles présente des spécificités : en particulier, que les catégories correspondant à un statut élevé soient fortement représentées et que les retraités soient à peine représentés. Il en va bien ainsi, comme l'atteste le tableau suivant, établi sur les 609 sujets pour lesquels cette information est connue.

	Artisans, commerçants	cadres supérieurs, prof. intellectuelles	prof. intermédiaires	employés	ouvriers	retraités	sans activité
effectif	19	306	161	74	13	1	35
pourcentage	3,12	50,25	26,44	12,15	2,13	0,16	5,75

Tableau 1.3 : Répartition des catégories socio-professionnelles dans l'échantillon

On y observe que l'échantillon est constitué pour moitié de cadres supérieurs et professions intellectuelles, alors que les ouvriers sont très faiblement représentés. Les retraités en sont pour ainsi dire absents.

En bref, l'échantillon qui a servi à mettre au point le MBTI est assez spécifique : les hommes y sont, en moyenne, plus âgés que les femmes, et les personnes âgées en sont pratiquement exclues. Cet échantillon est, surtout, fortement éduqué et les catégories socioprofessionnelles à haut statut y sont fortement sur-représentées.

3. Les quatre échelles

La conception de Jung, opérationnalisée dans le MBTI, considère les quatre oppositions suivantes

- l'attitude *extravertie* s'oppose à l'attitude *introvertie* ;
- le processus de perception des informations s'effectue par *sensation* ou par *intuition*.
- le processus de jugement privilégie la *pensée* ou le *sentiment* ;
- le style d'approche du monde extérieur oppose l'utilisation principale du *jugement* à l'utilisation principale de la *pensée*.

Cette dernière opposition relève de l'interprétation qu'ont faite Myers et Briggs de la conception de Jung.

Plusieurs questions se posent à propos des quatre échelles du MBTI, la première étant celle de la validité des quatre échelles. Celle-ci est examinée par analyses factorielles. Les résultats seront complétés par l'examen de la fidélité de ces échelles. Enfin, nous nous intéresserons aux contenus de ces échelles.

Analyses factorielles en quatre facteurs

Le manuel rapporte que le développement de la forme G révisée du MBTI s'est largement basé sur une stratégie factorielle visant à sélectionner, au fur et à mesure, les meilleurs items pour chaque échelle. Toutefois, les informations techniques fournies dans ce même manuel ne sont pas suffisantes pour répliquer exactement les analyses : la méthode d'analyse factorielle utilisée n'est pas précisée (on parle seulement d'une "analyse factorielle avec rotations obliques") et la manière dont ont été gérées les données manquantes n'est pas non plus indiquée⁵.

Afin d'essayer de rester le plus proche possible de la démarche suivie dans le manuel, et en fonction des informations dont nous disposons, nous avons effectué des analyses factorielles séparées pour les hommes et les femmes selon la méthode des axes principaux suivie d'une rotation oblique (algorithme OBLIMIN).

Le manuel ne propose pas les résultats des analyses factorielles réalisées sur la version définitive du questionnaire. Seuls des résultats partiels de la dernière version préliminaire, portant sur 94 items, sont présentés. Ces résultats ont entraîné la suppression de 5 items, conduisant à retenir les 89 items qui constituent la version définitive. Il ne nous est donc pas

⁵ Ces non-réponses sont assez peu nombreuses, de l'ordre de 10 (sur 731) par item (le maximum étant de 17). Elles ne peuvent donc qu'avoir une influence minimale sur les résultats.

possible de comparer directement nos résultats, portant sur 89 items, aux résultats de l'analyse de 94 items.

Nous pouvons toutefois essayer de valider la structure théorique du questionnaire en vérifiant que les facteurs correspondent aux quatre échelles théoriques.

Par ailleurs, ces analyses factorielles peuvent être réalisées, soit à partir des réponses directes des sujets au questionnaire, soit à partir d'une codification plus complexe de ces réponses, accordant une pondération différente aux items, telle qu'elle est décrite dans le manuel. Cette codification repose sur le calcul d'un "ratio prédictif" qui "donne la probabilité pour qu'une réponse soit donnée en accord avec la note totale à l'échelle" (p.77). Ce "ratio prédictif" est, en fait, un indicateur de discrimination, qui permet de s'assurer de la consistance interne des items d'une échelle : une pondération de 2 est accordée aux items qui sont le plus liés au score total à l'échelle, une pondération de 1 est donnée aux items qui sont encore jugés suffisamment liés au score total ; les items qui sont trop peu ou inversement liés au score total sont éliminés de la notation, en leur accordant une pondération nulle. Il découle de ce système de notation des items, que si un item a une pondération nulle sur chacun des pôles de l'échelle, il n'intervient aucunement dans le calcul des scores. Au total, 9 items sont dans ce cas. Mais comme les pondérations sont différenciées pour les hommes et les femmes et que cette exclusion d'items ne s'applique qu'à l'un des deux sous-échantillons, ils ne sont pas supprimés du questionnaire : ils contribuent seulement à différencier le questionnaire selon qu'un homme ou une femme y répond. Le tableau 1.4 précise, pour chaque échelle, les items qui n'entrent pas dans le calcul des scores masculins ou féminins.

	E - I	S - N	T - F	J - P
	nombre de questions posées			
	21	24	22	22
	questions n'intervenant pas dans le calcul du score			
échantillon masculin			n° 83, 86, 88	
échantillon féminin	n° 57	n° 7, 69	n° 10, 70	n° 75

Tableau 1.4 : Items n'entrant pas dans le calcul des scores, pour chaque échelle, selon les sous-échantillons masculin et féminin

On constate ainsi que, selon que le sujet est un homme ou une femme, son score E-I est calculé sur 21 ou 20 items, son score S-N, sur 24 ou 22 items, son score T-F, sur 19 ou 20 items et son score J-P, sur 22 ou 21 items.

Nos analyses factorielles ont été effectuées selon ces deux possibilités :

- d'une part, avec une simple notation dichotomique de l'ensemble des 89 items : pour chaque item, la première modalité de réponse a été codée "0" et la seconde a été codée "1". Les non-réponses ont été recodées aléatoirement à l'une ou l'autre de ces deux valeurs.
- d'autre part, en adoptant le système de pondération différentielle selon le sexe, qui correspond au système de notation du MBTI, et qui conduit à éliminer certains items de la note totale à l'échelle. Ce système de notation revient à situer les réponses dichotomiques sur une échelle bipolaire en 5 points (de -2 à +2).

Comme ces analyses factorielles produisent des résultats très similaires selon le système de cotation des réponses adopté, nous ne présentons dans le détail que ceux qui portent sur la cotation dichotomique la plus simple. Les éventuelles différences seront brièvement commentées. Dans tous les cas, nous avons retenu, sur la base du scree-test de Cattell, des solutions à quatre facteurs pour les analyses portant sur les échantillons masculin et féminin.

	Facteur 1 (J - P)	Facteur 2 (E - I)	Facteur 3 (T - F)	Facteur 4 (S - N)		Facteur 1 (J - P)	Facteur 2 (E - I)	Facteur 3 (T - F)	Facteur 4 (S - N)
IT01	-.04606	.66106 *	-.01 166	.05231	IT46	-.22292	.43722	-.01690	.09374
IT02	-.09656	-.02517	.09095	38081 *	IT47	<u>.30652</u>	-.20974	-.24687	-.40193 *
IT03	-.25253	.10736	.36937 *	.06851	IT48	.04670	.15835	-.31843 *	-.21725
IT04	.504-46 *	.04663	.16277	.11760	IT49	<u>.27031 *</u>	-.05196	-.25005	-.09830
IT05	.0051 1	.41392 *	-.05067	.05547	IT50	-.27190	.21055	.12398	.38142 *
IT06	.53690 *	.01583	-.06802	10689	IT51	.01212	-.13095	.43036 *	-.02316
IT07	.04103	.13492	-.02004	.39913 *	IT52	-.10631	<u>.25337 *</u>	.05832	-.00321
IT08	-.52528 *	.12640	-.01455	<u>.30400</u>	IT53	-.01603	.39114 *	-.11768	.09724
IT09	-.09767	-.46310 *	.00124	.02568	IT54	<u>-.34216</u>	.26331	.03518	.51179 *
IT10	.00575	-.21954	-.33659 *	.04877	IT55	-.22478	.20806	.32488 *	.25068
IT11	.04803	.34654 *	-.07462	.03070	IT56	.37237 *	-.07071	-.07775	-.21126
IT12	<u>-.41651</u>	.09843	.12867	54252 *	IT57	.03327	<u>.27499 *</u>	-.09011	.01434
IT13	<u>-.31008</u>	.14208	.38149 *	.29512	IT58	.46680 *	.00600	-.01976	-.09866
IT14	.41761 *	.06232	-.03746	.21540	IT59	.19338	-.10622	.05515	.41160 *
IT15	-.01903	.45275 *	.01081	.08846	IT60	<u>-.27974 *</u>	10198	-.03308	.13119
IT16	*42211 *	-.01472	-.15837	.01425	IT61	15101	.42417 *	.05012	.21 154
IT17	.22500	.18360	-.05405	.45462 *	IT62	-.21425	.04564	.06904	.44905 *
IT18	-.51902 *	.02166	.06478	.25576	IT63	-.09818	-.05493	.34338 *	.04958
IT19	-.02340	<u>-.26588 *</u>	.02206	.06897	IT64	-.03549	<u>.24807 *</u>	-.19002	.07624
IT20	-.12495	.32397 *	.04947	.02389	IT65	.49746 *	.00666	-.13500	-.16448
IT21	-.24880	.22276	.06554	.52989 *	IT66	.14271	-.13157	.08299	-.40567 *
IT22	-.19308	.04199	.35465 *	.06694	IT67	-.33213 *	.01583	.01153	.01769
IT23	.53838 *	.02760	-.13229	-.13763	IT68	.10742	.36798 *	.08588	.00494
IT24	.02410	.47285 *	.09020	.08557	IT69	-.10533	.21915	-.13188	.33978 *
IT25	.41740 *	.17107	-.09010	-.17050	IT70	-.14200	4000	.44551 *	.12038
IT26	.16246	-.04298	-.01222	.45938 *	IT71	.38651 *	.05206	-.06587	-.05332
IT27	-.53771 *	.11244	.04416	17411	IT72	-.07480	.51310 *	-.00014	.07910
IT28	.01982	-.32245 *	.06719	-.06295	IT73	18286	.13245	-.12583	4.46194 *
IT29	.01943	.13814	-.16019	-.41 192 *	IT74	.13291	-.05420	-.48311 *	-.05980
IT30	.03312	-.04362	-.36213 *	.19785	IT75	<u>2682.8 *</u>	-.26271	-.17355	-.19481
IT31	.67861 *	.01327	-.05458	<u>-.30635</u>	IT76	-.14348	-.05536	-.01290	.44052 *
IT32	-.15844	-.01556	.14145	.37780 *	IT77	.07656	.09256	<u>.2054 1 *</u>	.07417
IT33	-.04959	-.12966	.49078 *	-.01619	IT78	-.09123	.48814 *	.08652	.14310
IT34	.05542	.47255 *	.19787	-.00397	IT79	.04213	-.05302	.05745	-.44763 *
IT35	-.08688	.15015	-.06457	-.33253 *	IT80	-.00138	.15061	-.37596 *	.06783
IT36	.08382	.02406	-.38483 *	-.11903	IT81	.33844 *	-.19776	-.06734	-.19193
IT37	.32206 *	.02863	-.09684	-.16699	IT82	-.16243	.12230	-.05230	.32929 *
IT38	-.23464	.01851	.18877	<u>.24633 *</u>	IT83	.05121	.00279	<u>.15726 *</u>	-.10327
IT39	-.15071	.11176	.41090 *	.04057	IT84	.00704	<u>.27787 *</u>	.01715	-.05642
IT40	-.16971	.52406 *	-.04269	.07899	IT85	<u>.32451</u>	-.12005	.00184	-.55216 *
IT41	.09773	-.13802	.04273	-.42303 *	IT86	.03950	-.17610	<u>-.13485 *</u>	-.20339
IT42	.07063	.10531	<u>-.24210 *</u>	-.05570	IT87	.45961 *	-.15585	-.19215	-.22241
IT43	.30863 *	-.11031	-.00211	-.26917	IT88	.02406	-.16589	<u>.08429 *</u>	-.04394
IT44	-.28938	.08098	.10369	.57233 *	IT89	-.02827	-.01608	-.24217	.03270
IT45	-.27457	-.06601	.34259 *	.22109					

Tableau 1.5 : Analyse factorielle selon la méthode des axes principaux avec rotation oblique sur les hommes (n = 432 ; 89 items dichotomiques). Les astérisques indiquent l'appartenance théorique de l'item ; les items soulignés peuvent être considérés comme problématiques (cf. texte) (les items mentionnés en gras n'entrent pas dans le calcul du score lorsque le système de cotation pondérée est adopté

	Facteur 1 (J - P)	Facteur 2 (E - I)	Facteur 3 (T - F)	Facteur 4 (S - N)		Facteur 1 (J - P)	Facteur 2 (E - I)	Facteur 3 (T - F)	Facteur 4 (S - N)
IT01	.08619	.64477 *	-.02920	.03260	IT46	.29179	.43136 *	-.03034	.16961
IT02	.06378	.04148	-.11205	.32909	IT47	-.25827	-.22550	<u>-.42930</u>	-.43460 *
IT03	.17289	.414218	.33856	.00378	IT48	.11252	.09229	<u>-.26847 *</u>	-.11832
IT04	-.52308 *	-.04725	-.00687	.8419	IT49	.30641 *	-.12572	<u>-.31944</u>	-.13297
IT05	.08179	.47308 *	.03416	.11490	IT50	<u>.43613</u>	.19905	.10295	.38943 *
IT06	-.57548 *	-.07541	.01791	.14899	IT51	-.09393	-.05727	.31887 *	.04828
IT07	-.03504	.12002	.08412	<u>-.19314 *</u>	IT52	.14804	<u>.28298 *</u>	.02562	-.16053
IT08	.59792 *	.18339	-.07249	.033852	IT53	.22980	.42731 *	-.11707	.12156
IT09	.00317	-.42385 *	-.03815	.25833	IT54	.23243	.27332	.06990	.43778 *
IT10	.01196	-.16876	<u>.18554 *</u>	-.03711	IT55	.08093	.08036	.34494 *	.28024
IT11	.10386	.47901 *	-.00397	.00235	IT56	-.32674 *	-.11724	-.02634	-.26426
IT12	.25165	.05362	.16603	.63826 *	IT57	.04636	<u>.18416 *</u>	.00370	-.00345
IT13	.17453	.07131	.47735 *	.06246	IT58	-.53439 *	-.08088	.07851	-.16288
IT14	-.54029 *	-.01033	-.02237	.17509	IT59	.19522	-.14779	.17464	<u>-.24509 *</u>
IT15	.20297	.38718 *	.10059	.05388	IT60	.32965 *	.19667	-.15507	.19103
IT16	-.48806 *	-.06832	-.12291	.00471	IT61	-.01842	.35381 *	.04483	-.21342
IT17	-.10922	-.23510	-.08276	-.33251 *	IT62	.23278	.14118	.12111	.54565 *
IT18	.45798 *	.13065	.04419	.17229	IT63	.03219	.01399	.36166 *	.02033
IT19	-.03079	-.33113 *	.01779	-.00600	IT64	.15848	<u>.24980 *</u>	<u>-.37609</u>	.09680
IT20	.09710	.38475 *	.03223	.06743	IT65	-.34235 *	-.14951	.07619	-.17318
IT21	.22862	.09564	.18460	.45503 *	IT66	-.19279	.01634	.26755	<u>-.26531 *</u>
IT22	.06593	.06424	.30878 *	.12829	IT67	.34498 *	.08728	-.01235	-.02163
IT23	-.52002 *	-.07729	-.09425	-.10330	IT68	.07303	.35415 *	.01519	-.00987
IT24	.15649	.45252 *	.12794	-.02894	IT69	.19337	.10860	-.16471	<u>.25295 *</u>
IT25	-.39392 *	.03162	.05812	-.22679	IT70	.12739	.14981	.30902 *	.09016
IT26	-.23586	.05639	.06280	-.41463 *	IT71	-.42915 *	-.16668	.02098	-.17190
IT27	.41329 *	.15958	-.04053	.11535	IT72	.12285	.45442 *	-.07605	.03088
IT28	.04524	-.35006 *	-.10275	-.05190	IT73	-.09804	.13971	-.07426	-.38961 *
IT29	-.05318	.21163	-.03613	-.28635 *	IT74	.09567	-.13300	-.41414 *	.04289
IT30	.00054	.10460	-.32954 *	.25066	IT75	<u>-.17345 *</u>	-.23545	-.15051	-.12147
IT31	-.62470 *	-.11742	-.07460	-.27438	IT76	.15917	.03443	.05046	.37270 *
IT32	.20988	-.08014	.07030	.46868 *	IT77	.01306	-.04991	<u>.18799 *</u>	.17368
IT33	.01407	.02888	.42918 *	.02035	IT78	.07062	.49825 *	.16039	.11750
IT34	.10512	.35605 *	.15704	.09053	IT79	-.18516	.03730	.03095	-.40404 *
IT35	-.16166	.10278	-.03953	-.40003 *	IT80	.01862	.06400	-.39917 *	.11675
IT36	.04443	.10265	-.40449 *	-.10911	IT81	-.33629 *	-.21362	-.15424	<u>-.31189</u>
IT37	-.25488 *	-.10474	.00265	.16123	IT82	<u>.37036</u>	.11228	-.15384	.44399 *
IT38	.14507	-.10224	.04158	<u>.28317 *</u>	IT83	-.08684	-.04543	.42453 *	-.05977
IT39	.11388	.15632	.47117 *	-.03217	IT84	.09234	<u>.26839 *</u>	.07838	-.14572
IT40	.10150	.40041 *	.10463	.06160	IT85	<u>-.41778</u>	-.16427	.12763	-.44649 *
IT41	-.00776	-.10601	-.10315	-.41584 *	IT86	-.03625	-.13696	-.36168 *	-.16491
IT42	-.15853	.03012	<u>-.22455 *</u>	-.18919	IT87	-.33623 *	.06395	-.11232	<u>-.32337</u>
IT43	-.41366 *	-.16612	-.15217	-.22850	IT88	-.01267	.08237	<u>.27957 *</u>	-.02366
IT44	.22564	.16506	.27823	.58465 *	IT89	.04436	-.12096	<u>-.26067</u>	-.01868
IT45	.16530	.02570	.35402 *	.15817					

Tableau 1.6 : Analyse factorielle selon la méthode des axes principaux avec rotation oblique sur les femmes (n = 299 ; 89 items dichotomiques). Les astérisques indiquent l'appartenance théorique de l'item ; les items soulignés peuvent être considérés comme problématiques (cf. texte) ; (les items mentionnés en gras n'entrent pas dans le calcul du score lorsque le système de cotation pondérée est adopté).

Ainsi que l'attestent les tableaux 1.5 (pour les hommes) et 1.6 (pour les femmes), les saturations les plus élevées sur chaque facteur sont observées, en général, en fonction de

l'appartenance théorique des items à chaque échelle. La validation de la structure théorique en quatre facteurs bipolaires peut être considérée comme satisfaisante. Il subsiste toutefois quelques exceptions qui auraient probablement justifié quelques aménagements supplémentaires. Si on retient, comme le font les auteurs dans le manuel, le seuil arbitraire de 0.30 pour décider qu'un item est saturé sur un facteur, on peut dégager deux ensembles d'items problématiques. Ce sont les items qui sont soulignés dans les tableaux .

- les items qui ont une saturation inférieure à .30 sur leur facteur théorique ;
- les items qui ont une saturation supérieure à .30 sur un autre facteur que leur facteur théorique d'appartenance.

Ces items se retrouvent identiquement dans les analyses exploitant une codification pondérée des items. Le nombre d'items problématiques est synthétisé dans le tableau qui suit.

échantillon	codification	E - I	S - N	T - F	J - P	total
hommes	dichotomique	5	3	6	8	22
	pondérée	5	3	3	8	19
femmes	dichotomique	4	8	9	4	25
	pondérée	3	6	7	3	19

Tableau 1.7 : Nombre d'items "problématiques" selon les analyses (homme - femme), selon l'échelle d'appartenance théorique et selon la cotation des items (dichotomique ou pondérée)

Il est vrai que le système de pondération adopté contribue à éliminer certains des items problématiques, puisqu'il supprime 3 items chez les hommes et 6 items chez les femmes, items dont les saturations sont faibles sur leur facteur théorique d'appartenance. Mais ce système de pondération laisse encore un nombre non négligeable d'items défectueux, puisque ces items représentent plus de 20 % de l'ensemble des items.

La proximité des résultats obtenus selon que les analyses portent sur des données simplement dichotomiques ou sur des données pondérées selon un système complexe de notation, différenciant les réponses des hommes et des femmes, conduit à se demander s'il y a un véritable bénéfice à opter pour un système complexe. Un gain paraît manifeste par le fait que l'élimination de certains items, qui découle de la pondération nulle, supprime effectivement des items défectueux. Mais cette suppression indirecte d'items reste insuffisante (des items problématiques sont préservés) et elle entraîne des différences entre hommes et femmes qui risquent de compromettre les comparaisons ultérieures de leurs scores, puisque ceux-ci ne sont pas établis sur un même nombre d'items. L'amélioration des résultats pourrait être effectuée de façon beaucoup plus simple et plus radicale, par élimination directe des items défectueux du questionnaire.

S'il fallait résumer en quelques mots les résultats de ces premières analyses, on pourrait dire que les analyses factorielles que nous avons effectuées confirment approximativement la structure théorique du questionnaire. Il reste que la validation de cette structure aurait encore pu être améliorée car certains items ne présentent pas la structure factorielle attendue.

Au titre de la validation intraconcept, ce premier examen de la structure du MBTI est complété par l'examen de la fidélité de chacune des échelles.

La fidélité des échelles

La fidélité est une qualité essentielle de toute mesure et on est quelque peu surpris de ne trouver dans le manuel que des informations indirectes sur cette question. En effet, les coefficients de Cronbach⁶ correspondant aux quatre échelles, présentés dans le manuel, correspondent à une version antérieure du questionnaire, différente de la version finale ! De plus comme ces coefficients sont présentés en référence aux résultats d'analyses factorielles, la méthodologie utilisée n'apparaît pas clairement. On ne sait pas, par exemple, quelles règles de codification ont été utilisées pour les items ni comment ont été traitées les non-réponses.

Nous avons donc vérifié nous-mêmes la fidélité des quatre échelles sur la version finale du questionnaire en prenant les options suivantes : les non-réponses ont été recodées aléatoirement comme précédemment ; les deux systèmes de pondération ont été utilisés pour permettre les comparaisons : d'une part, le système simple de codification dichotomique des items (1 point pour les réponses allant dans le sens de la dimension et 0 point pour les réponses inverses) ; d'autre part, le système de codification pondérée (0, 1, 2), adopté dans le manuel, différencié selon le sexe, et qui revient à coder les réponses sur une échelle en 5 points allant de -2 à +2, selon le pôle de la modalité de réponse sur chaque échelle bipolaire. Les résultats sont résumés dans le tableau 1.8.

	hommes			femmes			hommes + femmes
	B	P	M	B	P	M	B
E - I α	.80	.80	.80	.79	.79	.79	.80
Nombre d'items	21	21		21	20		
S - N α	.85	.84	.83	.83	.83	.81	.84
Nombre d'items	24	24		24	22		
T - F α	.73	.72	.72	.75	.73	.74	.77
Nombre d'items	22	19		22	20		
J - P α	.84	.84	.83	.83	.83	.83	.84
Nombre d'items	22	22		22	21		

Tableau 1.8 : Fidélité en termes de consistance interne des échelles (coefficients oc de Cronbach) pour les quatre échelles selon l'échantillon (B : calcul sur codes dichotomiques ; P : calcul sur codes pondérés ; M : valeurs fournies dans le manuel)

Au total, les résultats montrent une fidélité satisfaisante des échelles. Les résultats sont peu différenciés pour les hommes et les femmes et nos résultats, qu'ils soient établis sur la base d'une codification dichotomique ou pondérée, sont proches de ceux qu'indique le manuel (qui ne correspondent pas, rappelons-le, à la version finale du questionnaire). La fidélité de l'échelle T-F est un peu moins bonne. On peut penser que si les modifications évoquées à propos des résultats de l'analyse factorielle avaient été effectuées, la fidélité des échelles aurait pu encore être améliorée

⁶ Ces coefficients estiment la fidélité au moyen de la méthode de la consistance interne.

Le coefficient de Cronbach exprime la consistance interne d'une échelle, autrement dit son homogénéité. Il va de soi que l'homogénéité d'une échelle sera d'autant plus importante que les items seront fortement redondants, en termes de contenu : à l'extrême, si l'on posait toujours la même question aux sujets, on obtiendrait par nécessité une consistance interne très élevée, mais cette homogénéité serait factice. La valeur d'un coefficient α ne prend véritablement son sens que si les items ne sont pas redondants. Cet examen relève d'une analyse du contenu des items. Nous l'entreprenons brièvement, une telle analyse permettant aussi de répondre à une autre question que nous nous sommes posée au regard du MBTI : celle de l'originalité de ses contenus.

Le contenu des quatre échelles

L'échelle **extraversion - introversion** est très certainement la moins originale du MBTI. Cette opposition est introduite dans la plupart des grands inventaires de personnalité : on la trouve, par exemple, dans l'EPI d'Eysenck, dans le modèle de personnalité à cinq facteurs, modèle de la personnalité le plus consensuel actuellement, ou, sous forme de facteur secondaire, dans le modèle de Cattell en 16 facteurs primaires. Dans le MBTI, elle est conçue comme une attitude.

Cette dimension bipolaire est appréhendée (dans son pôle extraverti) par des items⁷ renvoyant à :

- une liaison facile avec autrui (1, 34) une communication facile (11) et rapide (20) avec autrui (40) , au fait de parler (52), de prendre part à la conversation (5) dans un groupe, d'y faire les présentations (53) ; au fait d'exprimer (24) ou de révéler (24) facilement ses sentiments (24), d'être facile à connaître (61), d'avoir beaucoup d'amis (19), d'avoir des informations sur tout le monde (28) ;
- être cordial (78) ;
- s' amuser dans les réceptions (9) et y jouer le boute-en train (72), faire la fête (84), être animé (46), utiliser des plaisanteries (64), être plutôt enthousiaste (15, 11) ;
- s'intéresser à la mode (57).

Elle contient donc une relative redondance dans les questions posées, bien que divers thèmes soient abordés (facilité de contact et de communication, goût pour le contact social, cordialité, enthousiasme).

Les trois autres échelles nous intéressent davantage car elles sont plus spécifiques au MBTI. Elles visent à appréhender des processus ou des fonctions, des modes d'approche et, en cela, se distinguent nettement des traits classiques.

L'échelle **sensation – intuition** est appréhendée par des items renvoyant à :

- la préférence pour la simplicité directe (2), le réalisme (12), le concret (73), le caractère pratique (21, 7, 26), le bon sens (62), les faits (29), plutôt que la nouveauté (21), l'ingéniosité (26), l'imagination (12, 62), l'aspect fascinant (47), l'abstrait (73), la théorie (7, 29) ;
- préférer le caractère conventionnel (17, 54), la conservation (82), les procédures connues (59), le connu (85), les méthodes éprouvées (66), plutôt que le changement (82), l'originalité (2, 17, 54), une méthode personnelle (59), l'inconnu (85), la nouveauté (66) ;

⁷ Le numéro de l'item est présenté entre parenthèses.

- la fabrication (41), la production (79), la construction (76), l'affirmation (35), plutôt que la création (41) la conception (79, 35), l'invention (76) ,
- préférer l'aspect logique à l'aspect fascinant (47)
- la préférence pour une occasion conduisant à de plus grandes choses, plutôt qu'une expérience qu'on est sûr d'aimer (69)

En dehors de l'item 69 (item peu saturé sur le facteur considéré pour les hommes, item supprimé de la codification pour les femmes), ces items opposent le concret à l'abstrait, le pratique au théorique, le conventionnel à l'original, à la nouveauté, le connu, le routinier à l'inconnu.

En dehors d'une certaine redondance qui se dégage de ces contenus, notre interrogation porte sur l'interprétation de l'échelle. L'opposition entre sensation et intuition, en tant que processus ou fonction influençant la perception, telle que l'échelle est définie, ne paraît guère immédiate. L'intuition est appréhendée par une préférence pour la nouveauté, l'inconnu, l'abstraction, la conception, la création ; elle ne l'est pas par le fait de saisir "les significations, les relations, les possibilités, de façon instinctive, sans intervention de l'esprit humain" (p.4). La sensation est appréhendée par une préférence pour le concret, le conventionnel, le connu, le routinier, le bon sens ; elle ne l'est pas par des items renvoyant à l' "appréhension des faits et des choses à travers les sens" (p.4).

Le contenu de ces items rappelle fortement le contenu des items d'ouverture à l'expérience, qui s'opposent au conventionnalisme, tels qu'ils sont exprimés dans les instruments mesures du modèle en cinq facteurs.

L'échelle **pensée - sentiment** est appréhendée par des items qui renvoient à une opposition entre :

- la tête (3), la logique (13, 22, 70), la raison (55), la pensée (30), l'analyse (39), être convaincant (36) l'absence de chaleur (10), la justice (42), juger (77), argumenter (83), les droits (63) et le cœur (3, 55), les sentiments (13, 22, 30, 63, 70), la sympathie (39), le surcroît de chaleur (10), être touchant (36), l'indulgence (42), concilier (77), se mettre d'accord (83) ;
- supporter, et pardonner (89) ;
- les profits (48), la fermeté (33), la fermeté d'esprit (74), la détermination (80), la dureté (51), la prévision (45), la critique (88) et les bienfaits (48), la gentillesse (33), la chaleur (74), le dévouement (80), la douceur (51), la compassion (45), l'absence de critique (88).

Outre la forte redondance des items qui varient sur le thème de l'opposition cœur-raison et des comportements qui en dérivent dans les relations interpersonnelles, ces items ne sont pas sans rappeler aussi une autre dimension du modèle en cinq facteurs, la dimension d'agrabilité, ou la bienveillance, facteur secondaire du modèle de Cattell.

L'échelle **jugement - perception** est appréhendée par des items qui renvoient à une opposition entre :

- la planification (6, 18, 71), la préparation à l'avance (6) en commençant le plus tôt possible (65). la programmation (14, 31, 58), l'organisation soigneuse du travail (23) ou des choses à faire (16), le suivi d'un plan prévu (25), d'un plan soigneusement préparé (8), un emploi du temps régulier (56), un programme régulier (60), sans faire les choses à la dernière minute (27), systématique (43, 75), par décision (49) et l'absence de planification (6, 31, 71 de préparation (6), de programmes (14, 58), d'emploi du temps régulier (56), qui laisse découvrir au fur et à mesure (23), de façon informelle (43), spontanée (16, 75), permettant d'affronter l'inattendu (8), quitte à travailler sous pression (18), à la dernière minute (27, 67), par impulsion (49) ;
- ordonné (87), ponctuel (37) et décontracté (87), pas pressé (37) ;
- permanent (81) et changeant (81, 60).

Ces contenus paraissent fortement redondants : ils opposent un goût de l'organisation méthodique et consciencieuse à son contraire. Ils rappellent ainsi très fortement la dimension "caractère consciencieux" du modèle en cinq facteurs. A nouveau, l'interprétation de cette échelle en termes de jugement et de perception ne paraît pas immédiate : une personne privilégiant le jugement est sensée utiliser "principalement une fonction de jugement (la Pensée ou le Sentiment) pour traiter avec le monde extérieur" (p.4). Si l'on peut comprendre que la planification requiert le jugement, on voit mal où se trouve représenté, dans les items, le sentiment. Une personne privilégiant la perception est, quant à elle, sensée utiliser "principalement une fonction de perception (la Sensation ou l'Intuition) pour traiter avec le monde extérieur" (p.4). On voit mal aussi en quoi l'absence de goût pour la prévision et l'organisation renverrait nécessairement à un mode de perception fondé sur les sens ou fondé sur l'intuition. Toutefois, on peut relever une proximité de contenu entre la préférence pour l'imprévu, l'inattendu du pôle "perception" de cette échelle et la préférence pour la nouveauté, le changement du pôle "intuition" de l'échelle sensation - intuition.

De l'ensemble de ces constats, il nous semble que les 3 dimensions du MBTI qui entendent mesurer des processus (ou fonctions) ou des styles d'approche, perdent beaucoup de leur attrait en termes d'originalité

- l'opposition entre les processus de perception "sensation - intuition" n'est pas sans rappeler la dimension d'ouverture du modèle en cinq facteurs elle oppose essentiellement le goût pour la nouveauté au conventionnalisme ,
- l'opposition entre les processus de jugement "pensée - sentiment", qui utilise bien des contenus renvoyant à ces deux thèmes, rappelle la dimension d'agréabilité du modèle en cinq facteurs : les personnes qui privilégient les sentiments accordent leur préférence à la gentillesse, la chaleur; la douceur, le dévouement,-etc. ; elles s'opposent aux personnes qui privilégient l'esprit critique, la dureté, etc. ;
- l'opposition dans le style d'approche du monde extérieur "jugement - perception" renvoie nettement, dans ses contenus, à la dimension "caractère consciencieux" du modèle en cinq facteurs. Il est difficile d'y voir autre chose, le contenu des items étant fortement redondant.

Certes, cette analyse de contenu relève de notre subjectivité interprétative. Mais, à notre sens, elle devrait être facilement étayée par des corrélations calculées entre les dimensions respectives du MBTI et du modèle en cinq facteurs. Aucune étude de validation convergente n'étant présentée dans le manuel, nous pouvons juste trouver une confirmation indirecte de notre interprétation dans les résultats d'études américaines (McCrae & Costa, 1989) ou italiennes (Saggino & Kline, 1996), qui notent que les échelles du MBTI sont liées à quatre des échelles du modèle en cinq facteurs : l'ouverture à l'expérience, l'extraversion, l'agréabilité et le caractère consciencieux.

Notre examen rapide du contenu des items révèle aussi une relation probable de l'échelle J-P avec l'échelle S-N, due à des contenus renvoyant dans les deux cas au goût pour la nouveauté l'imprévu. Il est probable que ce soit cette signification commune qui s'exprime dans la corrélation assez élevée (.45) entre ces échelles, que présente le manuel. Il s'agit, de loin, de la plus forte corrélation entre les quatre échelles de l'instrument.

Enfin, cette analyse sommaire des contenus révèle aussi une certaine redondance dans les items utilisés pour chacune des échelles. Cela s'explique en partie par le fait que les items présentent soit un choix entre des propositions détaillées, soit un choix entre de simples concepts. Les propositions détaillées ne sont, le plus souvent, que des illustrations très directes des concepts. Par exemple :

N° 12 Vous entendez-vous mieux

- avec les personnes imaginatives

ou

- avec les personnes réalistes

N° 44 - imaginatif

- réaliste

Une telle duplication quasi systématique des contenus (le questionnaire rassemble 48 oppositions détaillées et 51 oppositions sous forme de concepts) apporte nécessairement une redondance qui ne paraît pas vraiment justifiée. Aussi la consistance interne des échelles, par ailleurs très élevée, s'explique certainement pour partie par une telle redondance.

4. La complexité du système de codification des items

Dans les pages qui précèdent, nous avons déjà été confronté à la complexité du système de codification des items présenté dans le manuel. Les réponses sont codées 0, 1 ou 2 points selon les items et selon le sexe de l'individu. La similarité des résultats obtenus précédemment sur la base de cette codification complexe et sur la base d'une codification classique simplement dichotomique ne nous incite pas à considérer qu'une telle complexité est réellement justifiée. Dans cette section, nous allons poursuivre cette interrogation de façon plus systématique, car ce n'est pas la seule forme de complexité présente dans le manuel. Dans un premier temps, nous confronterons les scores que nous appellerons désormais "pondérés" (qui résultent du système officiel de correction) aux scores appelés "bruts" (qui sont calculés à partir d'un système beaucoup plus simple). Dans un deuxième temps, nous examinerons le mode de calcul des scores "continus" proposé dans le manuel.

Scores bruts et scores pondérés

Avant d'effectuer cette comparaison entre scores bruts et pondérés, signalons que nous avons pris la peine de vérifier qu'en calculant les scores pondérés, nous obtenions exactement les mêmes résultats que ceux présentés dans le manuel (à propos de la fréquence des différents types notamment). Comme il n'y a pas de divergence entre nos résultats et ceux du manuel, nous ne ferons pas état de ces analyses plus avant.

Précisons encore que les réponses des sujets conduisent au calcul de 8 scores, un par pôle de chaque échelle, et non de 4, comme on aurait pu s'y attendre (nous reviendrons sur cet aspect ultérieurement). Rappelons enfin que les scores bruts ont été obtenus simplement en attribuant 1 point aux items allant dans le sens de l'échelle considérée et 0 point aux réponses inverses. Cette règle de correction est identique pour les hommes et les femmes.

La confrontation directe entre les scores bruts et les scores pondérés est naturellement la première étape de l'évaluation de la pertinence du système complexe de correction. Les corrélations entre ces scores selon le sexe et selon les échelles sont présentées dans le tableau 1.9.

échelles	hommes	femmes	hommes + femmes
E	.96	.96	.96
I	.98	.97	.97
S	.98	.97	.98
N	.98	.96	.97
T	.92	.93	.93
F	.92	.93	.92
J	.98	.98	.98
P	.98	.98	.98

Tableau 1.9 : Corrélations entre les scores bruts et les scores pondérés selon le sexe et selon les échelles

Toutes les corrélations sont très élevées. Elles sont un peu plus basses pour les échelles T et F, mais cela peut s'expliquer facilement par la moindre fidélité de cette dimension (cf. supra). Ces résultats ne plaident pas en faveur de l'intérêt d'adopter un système complexe de codification.

Une autre manière d'envisager la confrontation entre scores bruts et scores pondérés est d'étudier si les corrélations entre les dimensions changent selon le système de correction utilisée. Comme les deux pôles de chaque échelle expriment quasi la même information, nous ne reportons dans le tableau 1.10 que les corrélations en fonction de l'un des pôles de chaque dimension.

Les corrélations entre les scores changent très peu lorsque l'on passe d'un système de codification à un autre. Sur le plan pratique, on peut estimer que ces différences n'engendreraient aucune conséquence si on adoptait le système simplifié de cotation.

	hommes			femmes			hommes + femmes		
	E br. pd.	S br. pd.	T br. pd.	E br. pd.	S br. pd.	T br. pd.	E br. pd.	S br. pd.	T br. pd.
S brut pondéré	-.12 -.13			-.09 -.10			-.11 -.12		
T brut pondéré	-.03 -.01	.21 .14		-.05 -.06	.13 .15		-.04 -.03	.22 .16	
J brut pondéré	-.05 -.06	.44 .42	.26 .23	-.25 -.19	.50 .49	.14 .12	-.13 -.11	.47 .45	.24 .19

Tableau 1.10 : Corrélations entre les scores selon le sexe et selon le système de correction (br. : ; brut ; pd. : pondéré)

Les résultats précédents ont montré que l'adoption de notes brutes à la place de notes pondérées ne changeait que peu de choses sur le plan quantitatif. Qu'en est-il sur un plan plus qualitatif ? On se rappelle que c'est la confrontation entre les scores des deux pôles opposés d'une même dimension qui permet de déterminer le type de l'individu. Ainsi, lorsqu'une personne a un score de 10 à l'échelle E et de 9 à l'échelle I, on considère que sa préférence est l'extraversion (on procède de même pour les autres dimensions). La question qui se pose est donc de savoir si les préférences déterminées à partir des notes pondérées sont stables si l'on passe à des notes brutes. Quelques éléments de réponse à cette question figurent dans le tableau 1.11 .

Brut Pond.	E	I		N	S		F	T		J	P
E	50.1	1.6	N	55.7	0.5	F	63.3	1.0	J	58.8	1.1
I	2.9	45.4	S	5.5	38.3	T	6.0	29.7	P	1.5	38.6

Tableau 1.11 : Pourcentages des préférences entre les pôles selon le système de codification (exemple : pour la première dimension, il y a 4,5% des individus qui changent de type + 1,6%) et 95,6% qui conservent le même type)

A nouveau, ces résultats montrent que le passage d'un système complexe à un système simple de codification n'entraînerait que des modifications minimales. Certes, il existe des différences, mais il est important de souligner que les pourcentages de changement rapportés dans le tableau 1.11 ne sont pas des taux d'erreur imputables à un système de codification simplifié. Les notes pondérées elles-mêmes comportent une part d'erreur et les divergences observées quand on passe d'un système à l'autre n'excèdent pas la marge d'erreur habituelle en fonction de la fidélité des mesures psychologiques.

Les scores continus

Il va de soi que l'option prise, dans le MBTI, de ne considérer que la préférence d'un pôle sur l'autre dans l'établissement des types est une option réductrice : l'information en termes d'intensité est négligée au profit d'une simple présence d'un pôle sur l'autre. Et la personne se

trouve identifiée au pôle pour lequel elle manifeste une préférence, que celle-ci soit forte ou faible. Qu'advient-il lorsque la personne ne manifeste pas de préférence entre l'un et l'autre pôle ? La feuille de réponse stipule qu'elle soit assignée au pôle introversion dans le cas d'E-I, au pôle intuition dans le cas de S-N, au pôle sentiment dans le cas de T-F et au pôle perception dans le cas de J-P.

Toutefois, le manuel propose aussi le calcul éventuel de scores continus, si le besoin s'en fait sentir pour certaines analyses corrélationnelles. Le calcul de ce score continu nous étonne fortement par sa complexité. Le manuel propose la démarche suivante (p.56) :

- déterminer (sur la base de la codification pondérée) "la lettre qui indique la préférence" ;
- "calculer le score qui donne la clarté à la préférence" : "Sur chaque dimension, le score est égal au double de la différence entre le plus grand nombre et le plus petit", il est augmenté de 1 si le pôle ayant le plus de points est I, N, F ou P ; diminué de 1 si le pôle ayant le plus de points est E,S,T ou J ;
- calculer les scores continus : "Les scores continus sont une simple transformation linéaire des préférences, selon la convention suivante :
 - Pour E, S, T et J, le score continu est égal à 100 moins le score tel qu'il est calculé ci-dessus
 - Pour I, N, F et P, le score continu est égal à 100 plus le score".

Cette démarche revient, alors que l'on dispose de 4 dimensions continues (ce sont les dimensions qui résultent des analyses factorielles), à considérer 8 scores correspondant à des sous-dimensions (aux pôles), à confronter ces scores deux à deux et à les soustraire l'un à l'autre, en y ajoutant ou retranchant mystérieusement 1 point, puis à rassembler les deux pôles par addition ou soustraction de 100 points à la différence préalable.

On comprend difficilement pourquoi une telle démarche est appliquée dans la mesure où le score continu, c'est-à-dire le score aux échelles bipolaires, est obtenu directement, comme nous l'avons fait précédemment de façon très classique :

- soit sur la simple base d'une codification dichotomique (0, 1) très simple des réponses : les items sont codés 0 ou 1 de manière à ce que le 1 soit toujours orienté dans le sens du même pôle de l'échelle ,
- soit en adoptant le système de pondération (0, 1, 2) des réponses préconisé dans le manuel : les items sont codés 0, 1 ou 2 quand ils sont orientés dans le sens d'un pôle et -2, -1 et 0 quand ils sont orientés dans le sens opposé.

Une illustration chiffrée sera certainement plus éloquente que de longs commentaires. Le tableau qui suit présente les scores continus obtenus directement et selon le mode de calcul indiqué dans le manuel, pour les premiers sujets de l'échantillon, dans le cas de la dimension "extraversion - introversion". Dans les deux cas de calcul, les scores sont établis sur la codification pondérée.

n° sujet	échelle E-I	pôle E	pôle I	doublé de la différence entre pôles ± 1	transformation linéaire (± 100)	rôle de ± 1
1	<i>23,00</i>	1,00	24,00	47,00	147,00	147,00
2	<i>3,00</i>	8,00	11,00	7,00	107,00	107,00
3	<i>-3,00</i>	11,00	8,00	5,00	95,00	95,00
4	<i>0,00</i>	9,00	9,00	0,00	100,00	101,00
5	<i>10,00</i>	7,00	16,00	19,00	119,00	119,00
6	<i>2,00</i>	8,00	10,00	5,00	105,00	105,00
7	<i>-14,00</i>	19,00	5,00	27,00	73,00	73,00
8	<i>14,00</i>	4,00	18,00	29,00	129,00	129,00
9	<i>4,00</i>	9,00	13,00	9,00	109,00	109,00
10	<i>0,00</i>	11,00	11,00	0,00	100,00	101,00
11	<i>16,00</i>	1,00	17,00	33,00	133,00	133,00
12	<i>-7,00</i>	14,00	7,00	13,00	87,00	87,00
13	<i>12,00</i>	6,00	18,00	25,00	125,00	125,00
14	<i>-3,00</i>	15,00	12,00	5,00	95,00	95,00
15	<i>-2,00</i>	12,00	10,00	3,00	97,00	97,00

Tableau 1.12 : Comparaison des résultats entre un calcul direct du score bipolaire (en italique) et un calcul très indirect selon les instructions du manuel

On constatera aisément que le score direct à l'échelle E - I n'est autre que la différence (orientée dans le sens du pôle considéré, ici I) entre les sous-scores I et E à chacun des pôles. Le score ultime préconisé par le manuel est calculé sur cette base, si ce n'est qu'il ajoute ou retire 1 point.

- Dans notre cas de calcul, un sujet a un score positif s'il est introverti ; il a un score négatif s'il est extraverti.
- Dans le calcul préconisé par le manuel, un sujet a un score supérieur à 100 s'il est introverti, un score inférieur à 100 s'il est extraverti.

Le score préconisé dans le manuel (I_m) est donc une simple transformation linéaire du score continu direct (I) [à la différence d'un point], qui peut s'énoncer très simplement :

$$I_m = 2 I + 100$$

Quel rôle joue l'ajout ou le retrait de 1 point ? Lorsqu'on applique les directives du manuel (p.56), il n'apparaît pas clairement. On ne peut le comprendre que par un exemple de calcul fourni aussi dans le manuel (p.56) pour chaque dimension. Dans le cas de la dimension J - P, on trouve :

$$J = 11 ; P = 11 ; \text{différence} = 0 ; \text{score} = 1 ; \text{score continu JP} = 101$$

Autrement dit, la fonction de l'ajout ou du retrait de 1 point est seulement de remplacer les réponses qui ne manifestent pas de préférence pour l'un ou l'autre pôle par des réponses qui traduisent une préférence pour I, N, F ou P, selon le cas. Les directives du manuel (p.56) doivent donc être complétées ainsi : "(...) il est augmenté de 1 si le pôle ayant le plus de points est I, F ou P" ou si les deux pôles ont le même nombre de points (...); Le résultat de ce dernier calcul est présenté dans la dernière colonne du tableau 1.12.

Il va de soi que les corrélations entre ces deux scores sont très élevées. Elles dépassent, dans tous les cas, .997, la légère différence par rapport à une relation linéaire parfaite s'expliquant par l'ajout ou le retrait de 1 point.

Convenons qu'une telle complexité de calcul ne s'explique guère, si ce n'est que l'ajout arbitraire d'un point à certaines réponses permet de déterminer une préférence chez les sujets qui n'en présentent pas. Une telle pratique nous paraît toutefois contestable. La question qui se pose est de savoir si les sujets qui n'ont pas de préférence sont nombreux ou s'il s'agit d'une rectification tout à fait occasionnelle. Soixante-neuf sujets (soit près de 10 % de l'échantillon) sont affectés par une rectification au moins. Nous laissons le lecteur juge de l'importance de ces rectifications.

Une autre question qui se pose à propos des scores aux échelles est celle de l'incidence de la codification pondérée sur l'établissement des scores. Dans des dimensions bipolaires (telles qu'elles ressortent de l'analyse factorielle), la corrélation entre les pôles d'une même dimension est nécessairement égale à -1. Pour aborder cette question, nous transitons par les scores aux deux pôles. Le tableau 1.13 présente les corrélations entre les deux pôles opposés de chaque dimension, selon que ces scores sont établis sur la base d'une codification dichotomique ou pondérée.

		Hommes	femmes	hommes + femmes
E et I	Pondéré	-.90	-.90	-.89
	brut	-.99	-.996	-.99
S et N	Pondéré	-.93	-.94	-.89
	brut	-.99	-.097	-.99
T et F	Pondéré	-.83	-.82	-.83
	brut	-.98	-.996	-.99
J et P	Pondéré	-.95	-.96	.95
	brut	-.99	-.997	-.995

Tableau 1.13 : Corrélations entre les deux pôles de chaque dimension selon le système de codification

On s'aperçoit que le fait d'adopter un système avec pondération revient à faire baisser quelque peu la valeur des corrélations⁸. Cela se comprend sur le plan empirique, mais ce résultat n'est guère cohérent sur le plan théorique, puisque, dans tout le manuel, on fait référence à des dimensions ayant deux pôles opposés. Ainsi qu'on peut le constater, l'adoption d'un système de codification pondéré conduit à s'écarter de cette stricte unidimensionnalité.

De l'ensemble des confrontations entre les deux systèmes de correction des réponses, il ressort que le système pondéré n'est pas fondamentalement différent d'un système beaucoup plus simple. Par ailleurs, nous venons de voir que, sur le plan théorique, le système pondéré était critiquable car il remettait partiellement en cause l'unidimensionnalité des dimensions. Il existe ; peut-être des raisons importantes qui ont conduit à adopter un système de correction

⁸ On constate à nouveau que c'est dans le cas de la dimension T - F que les différences sont les plus importantes ce qui illustre, indirectement, l'importance qu'il y a de disposer d'échelles les plus fidèles possibles.

pondérée. Comme ces justifications ne sont pas exposées dans le manuel, nous serions plutôt enclins à proposer une simplification du système de correction.

Quant à la complexité adoptée dans le calcul des scores continus aux quatre échelles, nous ne nous expliquons pas sa présence, même si elle permet indirectement d'assigner sur un des pôles de chaque dimension les sujets qui ne présentent pas de préférence .

5. Les types psychologiques

Sur la base d'une dichotomisation des quatre dimensions, on trouve, dans le manuel, une description détaillée des types psychologiques qui constituent l'une des originalités du MBTI. Il importe toutefois de s'interroger sur la pertinence de cette classification.

D'une part, la mise en place de la typologie repose sur une dichotomisation des scores, c'est-à-dire sur une perte d'information. D'autre part, elle est effectuée a priori sans assurance aucune qu'elle corresponde à une réalité empirique. Or il existe des méthodes permettant de réaliser des classifications a posteriori, sur la base des réponses des individus. Dans ces méthodes, désignées sous le terme générique d'analyse en clusters, on cherche à constituer des groupes d'individus homogènes, sur la base de leur ressemblance. Nous examinerons ces deux points successivement.

La dichotomisation des scores aux échelles

Comme nous l'avons vu précédemment, les analyses factorielles conduites sur les réponses des sujets permettent d'assurer assez correctement les quatre dimensions bipolaires du MBTI. Il deviendrait légitime de considérer des sous-scores à ces échelles si ces dimensions présentaient une distribution bimodale, c'est-à-dire, par exemple, si la plupart des sujets étaient soit extravertis soit introvertis, une minorité seulement d'entre eux ayant un score entre ces pôles. La figure 1.1 présente la distribution des scores pour chacune des échelles.

Il importe peu, à ce stade, de différencier les scores selon le système de codification des réponses ou selon le mode de calcul des scores, puisque nous avons vu que ces résultats étaient très fortement similaires.

On peut constater sur ces distributions empiriques de scores qu'aucune ne justifie nettement la répartition en deux sous-catégories de sujets adoptée pour établir la typologie. Outre la perte d'information engendrée par la suppression des scores au profit d'une simple dichotomie, cette répartition en deux catégories ne trouve pas de légitimité empirique.

Il nous reste à examiner si une typologie peut quand même être dégagée des réponses des sujets.

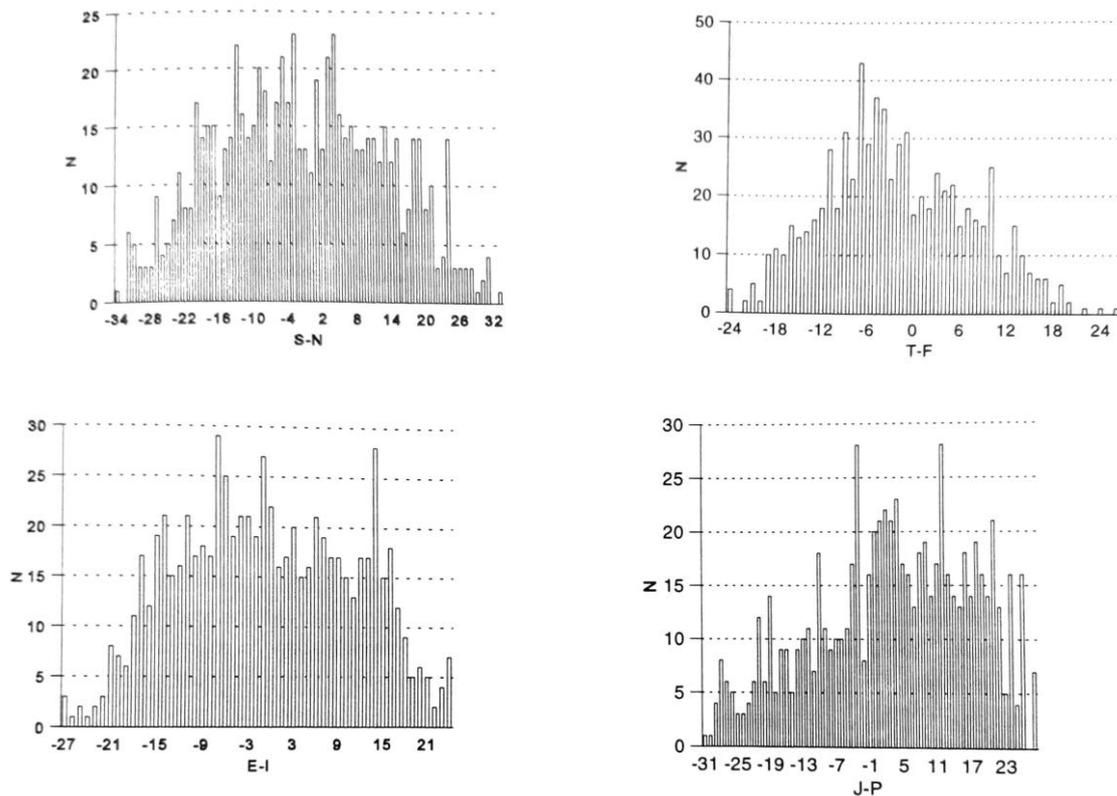


Figure 1.1 : distribution des scores sur chaque dimension

Y a-t-il des types dans le MBTI ?

La démarche des auteurs du MBTI a été d'établir des types en fonction de considérations théoriques, sur la base d'une dichotomisation des réponses. Ainsi, un sujet qui manifeste une préférence pour les pôles I, N, F et P sera classé dans le type INFP. La combinaison des préférences sur les 4 dimensions conduit à la création des 16 types :

ESTJ	ESTP	ISTJ	ISTP
ESFJ	ESFP	ISFJ	ISFP
ENTJ	ENTP	INTJ	INTP
ENFJ	ENFP	INFJ	INFP

Les suggestions du manuel ne s'arrêtent pas là : on pourrait encore, a priori, considérer les 8 types de Jung (tableau 4.3 du manuel), forgés sur la combinaison de 3 lettres ou bien les 4 tempéraments proposés par Keirsey et Bates (présentés p. 50) ; on pourrait aussi, a priori, considérer les combinaisons des fonctions avec la dimension E - I (tableau 4.5 du manuel), ce qui ajouterait 8 possibilités nouvelles aux 8 types de Jung ; et on pourrait encore s'intéresser à toutes les autres combinaisons ne faisant intervenir que deux lettres, présentées dans le manuel

(p 46-50), comme les "introvertis méthodiques" (IJ), les "factuels logiques" (ST), les "intuitifs adaptables" (NP), etc. Dans chaque cas, ces combinaisons déterminent 4 groupes.

Jung	combinaisons fonctions et tempéraments E-I		autres combinaisons			
ES-P	ES	SJ	IJ	ST	SJ	TJ
IS-J	IS	SP	IP	SF	SP	TP
EN-P	EN	NT	EP	NF	NP	FP
IN-J	IN		EJ	NT	NJ	FJ
E-TJ	ET					
I-TP	IT					
E-FJ	EF					
I-FP	IF					

La question qui se pose est de savoir si l'une ou l'autre de ces propositions a des fondements empiriques. Comme l'ensemble des propositions faites dans le manuel introduit des classifications en 16, 8 ou 4 types, il nous suffit d'examiner ces trois cas de figure. Cette question sera étudiée par analyses en clusters.

La différence entre analyse en clusters et construction de types a priori est illustrée par un exemple présenté dans la figure 1.2 : elle présente quatre individus fictifs pour lesquels on dispose d'une information chiffrée (ici de 0 à 20) sur quatre dimensions.

Dans l'approche *a priori*, les individus P1 et P2 sont considérés comme appartenant au même type psychologique (ils sont tous deux INFP car ils ont une préférence pour ces pôles). Il en est de même pour les individus P3 et P4 (tous deux ESTF).

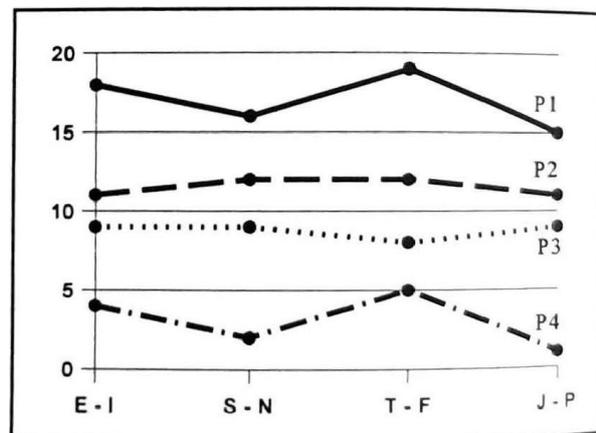


Figure 1.2 : Profils de quatre personnes fictives

Pourtant, au vu de ces profils, il apparaît clairement que P2 et P3 se ressemblent plus que P1 et P2 d'une part et que P3 et P4 d'autre part. Dans l'analyse en clusters (approche empirique *a posteriori*), il est fort probable que P2 et P3 soient classés dans le même groupe typologique.

Autrement dit, le système de classification a priori conduit, du fait de la dichotomisation forcée des dimensions, à créer des types psychologiques relativement peu homogènes (cas des individus P1 et P2). Il est préférable, selon nous, de proposer une classification plus homogène qui tienne compte du caractère continu des dimensions. Cette critique n'est pas seulement d'ordre technique : elle a des implications pratiques évidentes sur la description des types psychologiques du MBTI. Cette même figure permet d'illustrer un autre aspect. L'analyse en clusters, dont l'objectif est de classer les sujets dans des groupes en fonction de leur

ressemblance, travaille sur les relations entre sujets. Il est donc nécessaire de calculer des relations entre sujets et plusieurs indicateurs peuvent être utilisés à cette fin. Les plus classiques sont la distance euclidienne (ou son carré) et la corrélation.

Lorsque la corrélation est utilisée, seul le profil des ressemblances est pris en compte. Dans la figure 1.2, la corrélation nous indiquerait que P1 et P4 se ressemblent beaucoup, car la différence d'élévation des scores n'interviendrait pas (leur profil est presque parallèle, bien que l'un soit beaucoup plus élevé que l'autre). En optant pour un tel coefficient, nous n'aurions aucune chance de retrouver les typologies du MBTI, puisque l'option prise dans l'établissement de la typologie maximise les différences d'élévation des profils, en dichotomisant les scores.

Lorsque la distance est utilisée, les différences d'élévation du profil sont prises en compte. Selon cet indicateur de relation entre les sujets, les sujets P1 et P4 sont très différents.

Dans les analyses en clusters que nous effectuerons, nous utiliserons la distance euclidienne, ce qui nous situera dans la perspective la plus favorable à une validation de la typologie. Ces distances sont calculées sur les scores factoriels⁹ des individus, scores issus d'une analyse selon les axes principaux suivie d'une rotation varimax (afin de disposer de variables orthogonales). Tous les items ont été codés 1 ou 0 selon la procédure déjà décrite plus haut¹⁰ et les non-réponses recodées de manière aléatoire, comme précédemment. La méthode d'analyse en clusters utilisée est celle de Ward.

Comme nous nous situons, dans le cas présent, dans une démarche confirmatoire, nous retiendrons les solutions en 16, 8, et 4 groupes, afin d'examiner si l'une ou l'autre de ces solutions est interprétable dans l'esprit du MBTI.

Pour chacune de ces solutions, nous produisons l'effectif de chaque groupe et la distribution des sigles de 4 lettres du « type »¹¹ des personnes qui le composent (tableau 1.14). Les autres combinaisons possibles de lettres correspondant à d'autres typologies que celle qui propose 16 groupes peuvent être examinées à partir de cette même présentation.

Aucune de ces solutions ne permet de retenir l'une quelconque des typologies proposées. Examinons les brièvement tour à tour.

⁹ Nous avons vu au préalable que les analyses factorielles fournissaient quatre dimensions bien interprétables.

¹⁰ Nous avons vu précédemment que le fait d'utiliser une codification simple ou pondérée n'avait pratiquement pas d'incidence sur les résultats.

¹¹ Nous utilisons les "types" issus d'une codification simple, afin de garder une cohérence avec les analyses en clusters qui utilisent des scores issus de cette même codification simple. Nous avons vu précédemment (tableau 11) que la plupart des personnes conservaient le même "type" selon le système de codification adopté.

Solution en 16 groupes

Si on recherche 16 types, les personnes ENFJ paraissent se concentrer dans le groupe D mais ce groupe D contient aussi de nombreux ENFP et ESFJ. On retrouve par ailleurs de nombreuses personnes de ces "types" (ENFP, ESFJ) dans d'autres groupes (dans les groupes A et F pour les premières, dans les groupes E et H pour les secondes). Chaque ligne ou colonne du tableau pourrait être prise ainsi tour à tour sans pouvoir constater de correspondance manifeste entre un groupe et un des 16 "types". Le groupe P paraît isoler les sujets INFJ, mais nombre de sujets de ce "type" se trouvent aussi dans les groupes I et N. Le meilleur "type" que l'on puisse trouver est probablement le ESFP, mais il ne rassemble guère plus de la moitié de ses représentants théoriques dans un groupe (G) qui contient lui-même plus de la moitié d'autres "types".

Si on recherche seulement 8 types dans ces 16 groupes, l'entreprise n'est pas plus couronnée de succès : le groupe A paraît contenir une majorité de "FP" ; mais de nombreux "SFP", se trouvent aussi dans les groupes C, D, H ou O. Tout au plus peut-on dire que le type de Jung "E-TJ" se concentre dans le groupe B.

Solution en 8 groupes

Si on examine la correspondance avec la typologie de Jung, on retrouve la concentration des "E-TJ" dans le groupe B. Ils représentent 73 % du groupe. Les 7 autres types théoriques ne peuvent y être retrouvés.

Solution en 4 groupes

L'examen attentif des combinaisons de préférences deux à deux (c'est-à-dire de 24 possibilités) montre que le groupe A concentre les personnes "FP" (121, soit 91 % du groupe A) ou "NF" (120, soit 90 % du groupe A) ou "NP" (117, soit 88 % du groupe A). Mais ces regroupements de personnes ne représentent respectivement que 55 %, 37 %, 51 % de l'ensemble des "FP", "NF" ou "NP".

Compte tenu de ces résultats, ni la typologie en 16 groupes du MBTI, ni celle de Jung, ni les tempéraments de Keirsey et Bates ne se trouvent validés. Ces résultats, en fait, ne nous étonnent pas : les différentes propositions typologiques examinées se fondent toutes sur une bipartition des quatre dimensions fondamentales mesurées par le MBTI et ces bipartitions ne paraissent pas, elles-mêmes, fondées empiriquement.

Mais cette invalidation des différentes typologies examinées ne signifie pas qu'il soit impossible de regrouper les personnes selon qu'elles se ressemblent plus ou moins sur les quatre dimensions considérées. Il s'agit dès lors de faire des analyses en clusters en suivant une démarche exploratoire.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
n	38	73	67	60	35	33	44	46	28	30	59	62	41	67	29	16
enfj	2	8	37	27	1		1	1	2				1	1		
enfp	28		12	14		26	2				5	3	2			
entj		31	3	1							7					
entp		7	2			4	1				15					
esfj				13	21		8	18					1			
esfp	5			2	6		20						2			
estj		24		2			5			8	1		2			
estp							5									
infj	1		6		6				10		1	4	1	32		16
infp	2		6	1	1	3					1	51	3	14		
intj			1					1	16	22	10		3	9		
intp											18		2	5		
isfj		1					2	22				3	7	5	15	
isfp							1						4		5	
istj		2						4				1	8	1	9	
istp												1	5			

	A	B	C	D	E	F	G	H
n	133	75	127	79	74	101	59	83
enfj	2	8	63	1	3	1		1
enfp	57		26	2		1	5	
entj		31	3				7	
entp	4	7	3	1		2	15	
esfj			13	29	18	1		
esfp	5		2	26		2		
estj		24	2	5		10	1	
estp				5				
infj	5		6		10	1	1	48
infp	56		7			3	1	14
intj			1		1	3	10	9
intp						2	18	5
isfj				8	38	22		5
isfp	3			2		9		
istj					4	39		1
istp	1					5	1	

	A	B	C	D
n	133	261	180	157
enfj	2	72	2	4
enfp	57	31	3	
entj		41		
entp	4	25	3	
esfj		13	30	18
esfp	5	2	28	
estj		27	15	
estp			5	
infj	5	7	1	58
infp	56	8	3	14
intj		13	3	10
intp		18	2	5
isfj		1	30	43
isfp	3		11	
istj		2	39	5
istp	1	1	5	

Tableau 1.14 : distributions des 16 types a priori dans 16 groupes, dans 8 groupes et dans 4 groupes issus d'analyse en clusters

Analyses exploratoires

Le choix du nombre de groupes typologiques est, en analyse en clusters, une démarche essentiellement empirique. Sur la base de la chute des coefficients d'erreur et de l'évolution des effectifs intra-groupes, nous avons été amenés à retenir deux solutions. La première, à cinq groupes, reste à un niveau assez général ; la seconde, à dix groupes, est plus affinée.

Les résultats présentés ci-après n'ont pas de caractère définitif. Les groupes dégagés doivent être considérés comme provisoires et illustrent surtout une autre manière d'envisager les résultats du MBTI. *In fine*, l'intérêt de tels groupes, plus fondés sur le plan empirique, n'apparaîtra qu'en fonction d'études de validation complémentaires. Notre objectif est surtout, pour l'instant, de montrer l'applicabilité d'une telle démarche.

Une classification à cinq types

Pour simplifier la présentation des cinq groupes, la description portera uniquement sur l'un des pôles de chaque dimension (sachant qu'il existe une corrélation inverse presque parfaite entre les scores des deux pôles).

	E	S	T	J	sexe (% femmes)
échantillon total	10.6	10.9	9.6	12.5	40.9 %
1 (n = 157)	7.2	11.3	8.4	15.8	45.9 %
2 (n = 180)	9.7	16.6	10.6	13.4	30.0 %
3 (n = 133)	11.1	6.7	6.2	6.0	57.9 %
4 (n = 127)	14.0	8.3	8.0	13.2	53.5 %
5 (n = 134)	12.2	9.7	14.5	13.3	20.9 %

Tableau 1.15 : Description d'une classification des sujets en cinq groupes (moyennes aux échelles et pourcentage de femmes)

Les résultats concernant les moyennes aux échelles peuvent être visualisés sur la figure 1.3.

Nous ne rentrerons pas dans une description détaillée de ces groupes. En quelques mots, ils se caractérisent de la manière suivante:

- Le groupe 1 se caractérise surtout par un niveau moyen élevé d'introversion et par un mode d'approche du monde extérieur centré sur le jugement. Il n'y a pas de préférence marquée en ce qui concerne le type de perception habituellement choisi pour recueillir des informations (échelle S-N) ou le critère de jugement habituellement utilisé (échelle T-F). Il y a une légère surreprésentation des femmes dans ce groupe.
- Le groupe 2 a des niveaux moyens sur les échelles E-I, T-F et J-P. Il se particularise surtout par un score moyen très élevé sur l'échelle S-N qui indique que les sujets de ce groupe privilégient massivement la sensation (par opposition à l'intuition) dans la prise d'informations. Ce groupe se compose majoritairement de personnes de sexe masculin.
- Le groupe 3 a un profil très spécifique sur trois des quatre échelles. Les personnes de ce groupe privilégient l'intuition dans le recueil d'informations, le sentiment comme critère de jugement et la perception comme mode d'approche du monde extérieur. Ce groupe comporte proportionnellement beaucoup plus de femmes que d'hommes.
- Le groupe 4 est assez largement extraverti, privilégie plutôt l'intuition et le sentiment mais ces préférences sont moins marquées que dans le groupe 3. Par ailleurs, une autre

différence majeure entre ces deux groupes est que ce dernier utilise plutôt le jugement comme mode d'approche du monde extérieur. C'est aussi un groupe composé majoritairement de femmes.

- Le groupe 5 a des scores moyens sur les échelles E-I (légère préférence pour l'extraversion) et S-N. Il ne se démarque pas des groupes 1, 2 et 4 pour l'échelle J-P. Sa particularité majeure est son score élevé sur l'échelle T-F qui indique une préférence assez marquée pour la pensée (par opposition au sentiment) comme critère de jugement dans la prise d'une décision. Ce groupe comporte proportionnellement beaucoup plus de femmes que d'hommes.

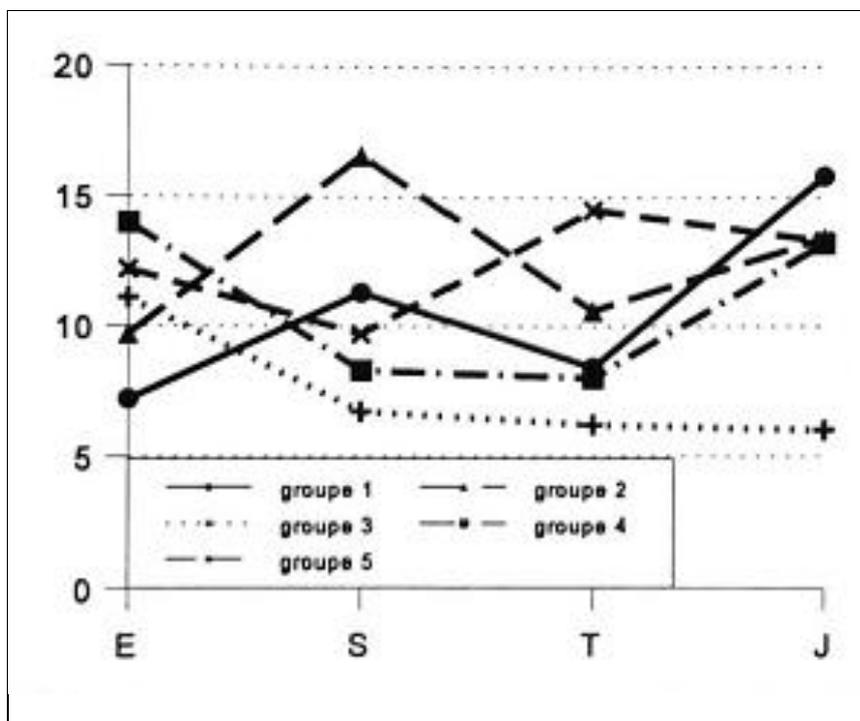


Figure 1.3 : Scores moyens des cinq groupes typologiques aux échelles du MBTI (scores bruts)

Comme la classification des sujets ne comporte que 5 groupes, on ne s'étonnera pas de ne pas trouver de correspondance directe avec les 16 types distingués dans le manuel. Il y a toutefois un certain recouvrement, ainsi que l'atteste le tableau 1.16 qui montre, si besoin est, que la typologie proposée n'est pas totalement arbitraire (on peut constater par exemple que le groupe 4 compte plus de 40 % de personnes de type théorique ENFJ alors qu'elles ne sont que 8,5 % dans l'échantillon total ; dans le groupe 3, on trouve principalement des personnes de type théorique ENFP et INFP, etc.).

	groupe 1	groupe 2	groupe 3	groupe 4	groupe 5	TOTAL
ENFJ	1.9	0.6	2.3	40.2	3.0	8.5
ENFP	0.0	1.1	41.4	16.5	1.5	10.9
ENTJ	0.0	0.0	0.0	7.1	26.9	6.2
ENTP	0.0	0.0	3.8	1.6	17.2	4.1
ESFJ	12.7	17.8	0.0	14.2	0.7	9.7
ESFP	0.0	12.2	4.5	3.9	0.0	4.5
ESTJ	1.3	7.8	0.0	3.9	19.4	6.4
ESTP	0.0	5.0	0.0	0.0	0.7	1.4
INFJ	31.8	0.6	3.0	5.5	1.5	8.8
INFP	8.3	1.1	37.6	5.5	0.7	10.0
INTJ	6.4	1.1	0.0	0.0	10.4	3.6
INTP	4.5	1.7	3.0	0.0	12.7	4.2
ISFJ	26.8	12.2	0.0	0.8	0.0	8.9
ISFP	0.6	9.4	3.0	0.0	0.0	3.0
ISTJ	5.7	24.4	0.0	0.8	3.0	7.9
ISTP	0.0	5.0	1.5	0.0	2.2	1.9

1.16 : Pourcentages des différents types théoriques dans les cinq groupes empiriques

Une classification à dix types

Nous reprendrons la même démarche que ci-dessus pour décrire la classification empirique en 10 groupes. Comme la logique est la même que précédemment, nous nous contenterons de présenter les résultats sous forme chiffrée.

	E	S	T	J	sexe (% femmes)
échantillon total	10.6	10.9	9.6	12.5	40.9 %
1	5.9	7.9	9.1	14.1	39.8 %
2	13.2	16.5	8.5	11.7	40.5 %
3	14.5	6.6	6.3	5.5	60.6 %
4	12.9	6.2	8.2	13.5	50.7 %
5	7.0	16.7	12.3	14.7	21.8 %
6	15.2	10.6	7.7	12.8	56.7 %
7	14.3	11.7	14.4	15.8	17.3 %
8	8.6	15.1	7.6	17.8	52.7 %
9	7.2	6.8	6.2	6.5	54.8 %
10	9.6	7.2	14.7	10.0	25.4 %

Tableau 1.17 : Description d'une classification en dix groupes (moyennes aux échelles et pourcentage de femmes)

Ces scores moyens sont illustrés graphiquement par la figure I.4. Afin que celle-ci reste lisible, ce sont les différents groupes typologiques qui figurent en abscisse au lieu des scores aux différentes échelles comme c'est traditionnellement le cas.

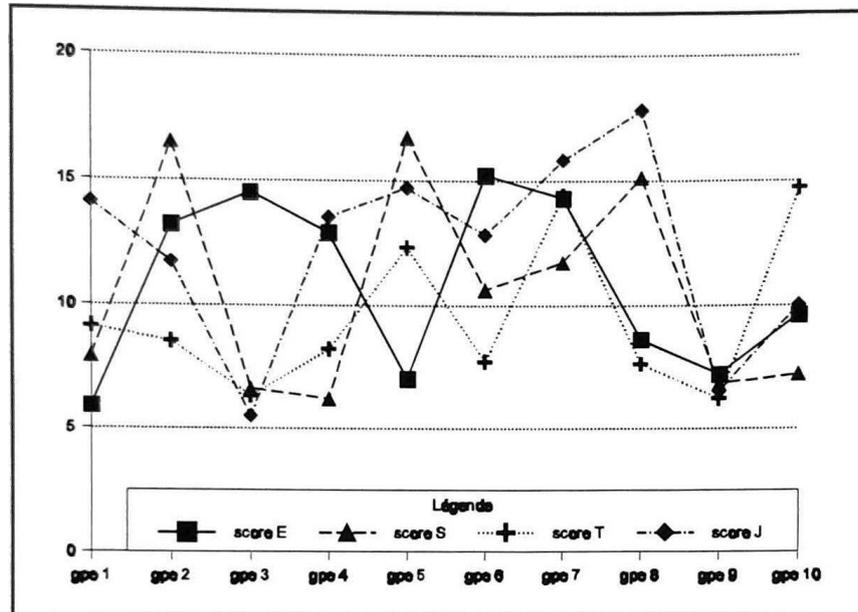


Figure 1.4 : Scores moyens des dix groupes typologiques aux échelles du MBTI (scores bruts)

	1	2	3	4	5	6	7	8	9	10	total
ENFJ	2.4	0.0	4.2	44.8	1.0	35.0	5.3	1.4	0.0	0.0	8.5
ENFP	0.0	1.3	73.2	17.9	1.0	15.0	0.0	0.0	4.8	3.4	10.9
ENTJ	0.0	0.0	0.0	9.0	0.0	5.0	38.7	0.0	0.0	11.9	6.2
ENTP	0.0	0.0	7.0	3.0	0.0	0.0	10.7	0.0	0.0	25.4	4.1
ESFJ	0.0	39.2	0.0	1.5	1.0	28.3	1.3	27.0	0.0	0.0	9.7
ESFP	0.0	27.8	8.5	0.0	0.0	8.3	0.0	0.0	0.0	0.0	4.5
ESTJ	0.0	7.6	0.0	1.5	7.9	6.7	33.3	2.7	0.0	1.7	6.4
ESTP	0.0	8.9	0.0	0.0	2.0	0.0	0.0	0.0	0.0	1.7	1.4
INFJ	49.4	1.3	1.4	10.4	0.0	0.0	2.7	12.2	4.8	0.0	8.8
INFP	15.7	1.3	5.6	9.0	1.0	1.7	0.0	0.0	74.2	1.7	10.0
INTJ	10.8	0.0	0.0	0.0	2.0	0.0	2.7	1.4	0.0	20.3	3.6
INTP	8.4	0.0	0.0	0.0	3.0	0.0	0.0	0.0	6.5	28.8	4.2
ISFJ	6.0	7.6	0.0	1.5	15.8	0.0	0.0	50.0	0.0	0.0	8.9
ISFP	1.2	3.8	0.0	0.0	13.9	0.0	0.0	0.0	6.5	0.0	3.0
ISTJ	6.0	0.0	0.0	1.5	43.6	0.0	5.3	5.4	0.0	0.0	7.9
ISTP	0.0	1.3	0.0	0.0	7.9	0.0	0.0	0.0	3.2	5.1	1.9

Tableau 1.18 • Pourcentages des différents types théoriques dans les dix groupes empiriques

6. Conclusion

Dans ces analyses secondaires du MBTI, nous avons examiné successivement différents aspects permettant de questionner la validité de l'instrument.

Les analyses factorielles ont confirmé, de manière générale, les résultats présentés dans le manuel : les quatre dimensions extraites des analyses sont facilement interprétables.

On peut regretter toutefois que le manuel ne fournisse pas plus de détails sur la méthodologie utilisée. Il reste aussi des items dont le pattern factoriel apparaît problématique. Ce résultat ne remet pas fondamentalement en cause la validité intraconcept du questionnaire, mais ce dernier aurait probablement pu encore être amélioré.

Les analyses de fidélité ont permis d'apporter des informations non présentées dans le manuel. Les coefficients sont satisfaisants dans l'ensemble, sauf pour l'échelle T-F où la consistance interne est tout juste acceptable étant donné le nombre d'items. L'exploitation plus détaillée des analyses factorielles qui vient d'être évoquée aurait probablement permis d'améliorer encore la fidélité des échelles qui reste, rappelons-le, satisfaisante dans l'ensemble.

Après examen du contenu des items forgeant les quatre échelles, il nous a semblé que la signification des dimensions en termes de processus ou de fonctions influençant la perception ou la décision relevait d'une interprétation assez indirecte : l'échelle sensation - intuition pourrait être interprétée comme une dimension d'ouverture à l'expérience, l'échelle pensée - sentiment pourrait être interprétée comme une dimension d'agréabilité, et l'échelle jugement perception renvoie très fortement au caractère consciencieux.

En bref, il existe bien quatre dimensions dans le MBTI mais il est moins sûr qu'elles appréhendent des processus influençant la perception ou la décision : le contenu de ces échelles est assez classique et rappelle fortement quatre des dimensions du modèle de la personnalité en cinq facteurs.

Les analyses concernant les systèmes de cotation des items et de calcul des scores adoptés pour le MBTI montrent que, sur le plan empirique, la complexité de ces systèmes n'est pas justifiée.

En ce qui concerne la cotation des items, le passage d'un système complexe (pondération des réponses selon l'item et selon le sexe) à un système beaucoup plus simple (cotation dichotomique identique pour les hommes et les femmes) ne modifie que marginalement les résultats. Étant donné le peu de différences observées entre cotation simple et cotation pondérée et différentielle (qu'il s'agisse des résultats des analyses factorielles, des corrélations entre les scores aux échelles, ou des préférences pour un pôle des échelles sur l'autre pôle), le système simplifié devrait probablement être privilégié.

En ce qui concerne le calcul éventuel des "scores continus" pour chacune des quatre dimensions, notre investigation a montré qu'il se soldait simplement par l'ajout ou le retrait de 1 point à un score qui pouvait être obtenu très directement. La complexité du calcul préconisé dans le manuel, outre le fait qu'elle ne se justifie pas, ne fait qu'introduire une déformation arbitraire des scores, qui, bien que mineure (elle ne joue

que sur un point) ne paraît pas légitime. Elle a pour effet d'établir une préférence entre les pôles de chaque échelle, là où les réponses de la personne n'en manifestent pas.

Enfin, nous avons questionné la validité des 16 types typologiques proposés dans le manuel, ainsi que celle d'autres typologies reposant sur ces quatre dimensions : la classification en huit types de Jung, la classification en quatre tempéraments de Keirsey et Bates.

Aucune de ces typologies n'a reçu de confirmation empirique par les analyses pratiquées. Cela n'étonne pas puisque nous avons vu au préalable que ces typologies se fondaient sur une bipartition des réponses selon les deux pôles de chaque échelle, sans que celle-ci se laisse légitimer par les données empiriques : les distributions de réponses ne présentent pas de bimodalité, ce qui signifie qu'une majorité de sujets ne présente pas de préférence nette pour un pôle ou l'autre de chaque dimension ; de plus, une préférence est à l'occasion imposée (dans le cas où un sujet se montre très exactement entre les deux pôles de chaque échelle).

Considérant que ces typologies ne reposaient que sur des critères a priori sans justification empirique, nous avons proposé d'autres systèmes de classification en utilisant une démarche exploratoire, résolument empirique. Les analyses nous ont conduits à retenir deux classifications possibles, l'une en cinq groupes, l'autre en dix types. Ces classifications présentent l'avantage de refléter une certaine réalité (les groupes sont déterminés en fonction de la similarité des profils de réponse des individus). Des données externes complémentaires seraient nécessaires pour évaluer l'intérêt de ces nouvelles "typologies".

En bref, l'ensemble des analyses pratiquées sur l'échantillon de 731 sujets qui a servi à mettre au point de la version française du MBTI permet de retenir quatre échelles dans ce questionnaire mais ne permet pas de retenir des types de personnes.

Si l'intérêt du MBTI, comme nous le suggérons au départ, réside dans son originalité par rapport aux autres inventaires de personnalité, ces résultats lui font perdre beaucoup de son intérêt : les "types" ne sont pas confirmés empiriquement ; les "préférences naturelles" pour les pôles des dimensions paraissent largement surfaites ; les quatre dimensions qu'il permet de mesurer paraissent largement similaires à quatre des cinq dimensions du modèle de la personnalité en cinq facteurs.

De ce dernier point de vue, une dimension fondamentale est passée sous silence dans cette représentation de la personnalité : le névrosisme ou la stabilité émotionnelle. Alors pourquoi utiliser le MBTI ?

Bibliographie

Abdi, H. (1987). Introduction au traitement statistique des données expérimentales. Grenoble . P.U.G.

Aldenderfer, M.S. & Blashfield, R.K. (1984). Cluster analysis. Beverly Hills : Sage publications.

APA (1985). Standards for education and psychological testing. Washington : American Psychological Association.

Barrick, M.R. & Mount, M.K. (1991). The Big Five personality dimensions and job performance. *Personnel Psychology*, 44, 1-26.

Beaufils, B. (1994). L'enseignant, le chercheur et les tests : de la transparence du calque à l'opacité du logiciel. In M. Huteau (Ed.) (1994). Les techniques psychologiques d'évaluation des personnes. Paris : EAP (pp. 199-206).

Bollen, K.A. (1989). Structural equations with latent variables. New York : Wiley.

Bray, J.H. & Maxwell, S.E. (1986). Multivariate analysis of variance. Beverly Hills : Sage Publications.

Briggs-Myers, I. & McCaulley, M.H. (traduction et adaptation de Cailloux, G. , & Cauvin, P.) (1996). Indicateur Typologique de Myers-Briggs (forme G révisée). Manuel. Paris : Les Editions du Centre de Psychologie Appliquée.

British Psychological Society (1989). Psychological testing: guidance for the user. Leicester : British Psychological Society.

Bruchon-Schweitzer, M. & Femeux, D. (1991) LES méthodes d'évaluation du personnel utilisées pour le recrutement en France. *L'orientation Scolaire et Professionnelle*, 20, 71-88.

Bunderson, C.V., Inouye, D.K. & Olsen, J.B. (1989). The four generations of computerized educational measurement. In R.L. Linn (Ed). *Educational Measurement*. New York : Macmillan (pp. 367-407).

Carroll, J.B. (1993). Human cognitive abilities. A survey of factor analytic studies. Cambridge: The Cambridge University Press.

Cattell, H.E.P. (1996). The original Big Five: A historical perspective. *European Review of Applied Psychology*, 46, 5-14.

Chan, W. & Bentler, P.M. (1993). The covariance structure analysis of ipsative data. *Sociological Methods & Research*, 22, 214-247.

Christensen, P.R., & (1955). Ship Destination Test. Berverly Hill, Calif.: Sheridan Psychological Services.

Conn. S.R. & Rieke, M.L. (1994). The 16 PF Fifth edition technical manual. Champaign, IL Institute for Personality and Ability Testing, Inc.:

Cornwell, J.M. & Dunlap, W.P. (1994). On the questionable soundness of factoring ipsative data: a response to Saville & Willson (1991). *Journal of Occupational and Organizational Psychology*, 67, 89-100.

Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354.

Dickes, P., Tournois, J., Flieller, A., & Kop, J.L. (1994). *Ln psychométrie*. Paris : PUF.

Digman, J.M. (1990). Personality structure : the emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.

Dunlap, W.P. & Cornwell, J.M. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research*, 29, 115-126.

ECPA (1995). *SOSIE : Manuel d'application*. Paris : Editions du Centre de Psychologie Appliquée.

ECPA (1996). *TEAI : Tests d'évaluation des aptitudes intellectuelles : Manuel*. Paris : Les Editions du Centre de Psychologie Appliquée.

Gendre, F. (1976). *L'analyse statistique multivariée*. Genève : Droz.

Gough, H.G. (1952). On making a good impression. *Journal of Educational Research*, 46, 33-42.

Green, R.F., Guilford, J.P., Christensen, P.R., & Comrey, A.L. (1953). A factor-analytic study of reasoning abilities. *Psychometrika*, 18, 135-160.

Guilford, J.P. (1967). *The nature of human intelligence*. New York: McGraw.

Guilford, J.P., Hoepfner, R., & Petersen, H. (1965). Predicting achievement in ninth-grade mathematics from measures of intellectual-aptitude factors. *Educational & Psychological Measurement*, 25, 659-682.

Gustafsson, J-E. & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, 28, 407-434.

Guttman, L., & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence*, 15, 79-103.

Hertzka, A.F., & (1991). *Test des syllogismes*. Manuel. Paris : Les Editions du Centre de Psychologie Appliquée.

Hicks, L.E. (1970). Some properties of ipsative, normative and forced-choice normative measures. *Psychological Bulletin*, 74, 167-184.

Huteau, M. (Ed.) (1994). *Les techniques psychologiques d'évaluation des personnes*. Paris : EAP.

NFPC (1996). Les bilans personnels et professionnels de compétences. Document interne. Luxembourg : NFPC.

Joras, M. (1995). Le bilan de compétences. Paris : PUF (QJ no 2979).

John, O.P. (1990). The "Big Five" factor taxonomy: dimensions of personality in the natural language and in questionnaires. In L.A. Pervin (Ed). Handbook of personality. Theory and Research. Hove : Guilford Press (pp. 66-100).

Jöreskog, K.G. & Sörbom, D. (1993). Lisrel 8 : User's reference guide. Chicago : Scientific Software International, Inc.

Kettner, N.W., Guilford, J.P., & Christensen, P.R. (1956); A factor analytic investigation of the factor called general reasoning. Educational & Psychological Measurement, 16, 438-453.

Kline, P. (1993). Personality : The psychometric view. New York: Routledge.

Koehn, D. (1996). Handwriting analysis in pre-employment screening. The Online Journal of Ethics, 1, 1 (<http://condor.depaul.edu/ethics/hand.html>).

Landy, F.J. (1986). Stamp collecting versus science: validation as hypothesis testing. American Psychologist, 41, 1183-1192.

Levy-Leboyer, C. (1996). Evaluation du personnel : quelles méthodes choisir ? Paris : Les Editions d'Organisation.

Lyon-Caen, G. (1991). Les libertés publiques et l'emploi. Paris : La documentation française.

Matarazzo, J. D. (1992). Psychological testing and assessment in the 21st century. American Psychologist, 47, 1007-1018.

McCrae, R.R. & Costa, P.T. (1989) Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. Journal of Personality, 57, 17-40.

Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 35, 1012-1027.

Messick, S. (1989). Validity. In R.L. Linn (Ed). Educational Measurement. New York Macmillan (pp. 13-103).

Nihira, K., Hoepfner, R., & Memfield, P.R. (1964). A factor analysis of the semantic-evaluation abilities. Rep. Psychol. lab. Univer. Southern Calif, No. 32.

Paulhus, D. (1984). Two-component models of social desirable responding. Journal of Personality and Social Psychology, 46, 598-609.

Prieto, J.M., Gouveia, V.V. & Fernandez, M.A. (1996). Evidence on the Primary Source-Trait Structure in the Spanish 16 PF, 5th Edition. European Review of Applied Psychology, 46, 33-43.

Robertson, I. & Smith, M. (1989). Personnel selection methods. In M. Smith & I. Robertson (Eds). *Advances in selection and assessment*. Chichester : Wiley (pp. 89-112).

Rolland, J.-P. (1993). Validité de construct de "marqueurs" des dimensions de la personnalité du modèle en cinq facteurs. *Revue Européenne de Psychologie Appliquée*, 43, 317-337.

Russel, M.T. & Karol, D.L. (1993). *The 16PF Fifth edition administrator's manual*. Champaign, IL : Institute for Personality and Ability Testing, Inc.

Russell, M.T., Karol, D.L., Gillet, I., Mogenet, J.-L. & Rolland, J.-P. (1995). *16 PF5 de R.B. Cattell : manuel*. Paris : Editions du Centre de Psychologie Appliquée.

Saggino, A. & Kline, P. (1996). The location of the Myers-Briggs type indicator in personality factor space. *Personality and Individual Differences*, 21, 591-597.

Saville, P. & Willson, E. (1991). The reliability and validity of nonnative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology*, 64, 219-238.

Snow, R.E., Kyllonen, P.C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates, 47-103.

Tenopyr, M.L. (1988). Artfactual reliability of forced-choice scales. *Journal of Applied Psychology*, 73, 749-751.

Tett, R.P., Jackson, D.N. & Rothstein, M. (1991). Personality measures as predictors of job performance: a meta-analytic review. *Personnel Psychology*, 44, 703-742.

The Hatfield Polytechnic Hertfordshire, Psychometric Research Unit (1985). *Graduate and Managerial Assessment : Manuel and User Guide*. Berkshire : NFER-Nelson.

Thurstone, L.L. (1938). Primary mental abilities. *Psychometric Monographs*, I.

Tournois, J. (1996). L'évolution de la mesure des valeurs : de Rokeach Schwartz. *Les Cahiers de Recherche de l'ADEPS n° 18*. Nancy : ADEPS (Université Nancy 2 et CNRS).

Tournois, J., Mesnil, F., & Kop, J.-L. (1998). Autoduperie et hétéroduperie : un instrument de mesure de la désirabilité sociale (sous presse). *Revue Européenne de Psychologie Appliquée*.

Very, P.S. (1974). Differential factor structures in mathematical ability. *Genetic Psychology Monographs*, 75, 169-207.

Warzecha, G. (1991). The challenge to psychological assessment from computer technology. *Revue Européenne de Psychologie Appliquée*, 41, 213-220.

Watson, G.B. (1925). The measurement of fair-mindedness. *Contributions to Education*, NO 176. New York: Bureau of Publications, Teachers College, Columbia University.

Watson, G.B., & Glaser, E.M. (1964). Watson-Glaser critical thinking appraisal manual for Forms Ym and Zm. Ney York: Harcourt Braque Jovanovic.

Watson, G., & Glaser, E.M. (1992). Evaluation de la pensée critique (CTA) : Manuel. Paris : Les Editions du Centre de Psychologie Appliquée.