

Evaluation du questionnaire et validation des échelles de mesure

1 Les critères de qualité des instruments de mesure

Selon Evrard & al. [2000, p.288], deux critères importants permettent de juger de la qualité d'un instrument de mesure : il s'agit de sa fiabilité (ou fidélité) et de sa validité.

1.1 La fiabilité

La fiabilité de l'outil de mesure est satisfaisante lorsque celui-ci est suffisamment stable et cohérent : le même phénomène mesuré plusieurs fois avec le même instrument doit permettre d'obtenir le même résultat, et les énoncés constituant l'outil de mesure doivent permettre au répondant d'apporter des réponses cohérentes. La fiabilité peut donc être évaluée de plusieurs manières :

- La stabilité de l'outil peut être estimée dans le temps en administrant le même questionnaire à la même population à deux moments différents (méthode du test-retest). Elle peut également être évaluée auprès de sujets différents, en administrant en même temps le même questionnaire à deux échantillons différents (méthode des deux moitiés ou « split-half »). En pratique cette méthode des deux moitiés consiste soit à scinder l'échantillon de l'enquête en deux moitiés et vérifier que les résultats obtenus sur les deux moitiés sont suffisamment corrélés, soit à scinder en deux les items d'une échelle de mesure et vérifier que les deux demi-échelles donnent des résultats corrélés sur l'ensemble de l'échantillon [Igalens & Roussel – 1998].

- La méthode recommandée pour mesurer la cohérence interne (appelée aussi homogénéité ou consistance) des échelles de mesure est généralement le calcul du coefficient Alpha de Cronbach [Cronbach - 1951]. Ce coefficient permet de vérifier si tous les items se réfèrent à des notions communes, autrement dit si chaque item présente une cohérence avec l'ensemble des autres items de l'échelle [Igalens & Roussel – 1998, p141].

Il peut se calculer en utilisant la formule simplifiée suivante [extraite de Peterson – 1994] :

$$\alpha = \frac{N \cdot \bar{r}}{1 + (N - 1) \cdot \bar{r}}$$

avec : N = nombre d'énoncés de l'échelle et \bar{r} = moyenne de tous les coefficients de corrélation entre les énoncés.

Le coefficient varie entre $-\infty$ et $+1$; une échelle est considérée comme d'autant plus fiable que Alpha est proche de $+1$.

Les niveaux généralement admis en sciences sociales sont les suivantes (d'après Peterson – 1994 ; Evrard & al. – 2000) :

Alpha > 0,9	Fiabilité très forte mais qui risque d'être artificielle en sciences sociales si elle dépasse 0,95 (probabilité de redondance sémantique entre les items liée à la multiplication d'énoncés trop proches) ¹³⁴
Alpha compris entre 0,8 et 0,9	Bonne fiabilité (niveau recommandé dans les analyses confirmatoires)
Alpha compris entre 0,7 et 0,8	Fiabilité acceptable dans une phase exploratoire
Alpha compris entre 0,6 et 0,7	Fiabilité éventuellement acceptable pour une recherche à caractère exploratoire (sinon, l'échelle doit être épurée)
Alpha compris entre 0,5 et 0,6	Fiabilité limitée (l'échelle doit être épurée)
Alpha < 0,5	Fiabilité insuffisante (l'échelle doit être révisée ou abandonnée)

Une méta- analyse menée par Peterson [1994] sur 882 articles lui a permis d'isoler les niveaux moyens constatés en fonction des échelles étudiées : dans les articles intégrant des échelles de mesure d'attitude, l'alpha moyen est de 0,76 (0,79 pour les études intégrant une mesure de l'implication). Igalens & Roussel proposent de d'utiliser ces résultats comme norme de fiabilité [1998, p.145]. Dans le cadre de notre étude, nous testerons donc l'homogénéité des échelles de mesure grâce au coefficient Alpha de Cronbach, en nous référant à ces résultats comme guide d'interprétation.

1.2. La validité

La validité d'un outil de mesure correspond à sa capacité à mesurer effectivement et efficacement le construit visé. Cette notion est propre au champ des sciences sociales, dans lesquelles la recherche porte souvent sur des concepts abstraits, non directement observables [Drucker & al. 1999]. Cette validité doit être envisagée à plusieurs niveaux : on distingue habituellement la validité de contenu, la validité de trait et la validité prédictive.

1.2.2 La validité de contenu ou validité faciale

Il s'agit d'une évaluation subjective mais systématique de la manière dont le contenu d'une échelle représente bien le construit à mesurer [Malhotra – 2004 p 209]. Ce type d'évaluation est en général réalisé par le chercheur, qui peut s'appuyer également sur le jugement de ses pairs : le critère de validité repose alors sur la notion d'accord intersubjectif. Une échelle de mesure sera donc jugée valide si elle fait l'objet d'un consensus auprès d'un certain nombre d'experts. La vérification de la validité de contenu apparaît à la fois comme la première étape et le minimum

¹³⁴ Le coefficient alpha est sensible au nombre d'énoncés dans une échelle : la multiplication d'énoncés sémantiquement proches fait augmenter mathématiquement le niveau de l'alpha [Cortina – 1993]

requis lorsque l'on souhaite mesurer l'adéquation de l'outil de mesure par rapport aux concepts à mesurer [Schrieschen & al. – 1995].

En pratique, tester la validité de contenu d'un outil de mesure consiste à soumettre celle-ci au jugement de deux types d'évaluateurs : il s'agit en premier lieu des experts du domaine étudié (chercheurs ou professionnels connaissant bien le sujet) et en second lieu des représentants de la population à qui est destinée l'étude. A l'issue de cette opération, une première épuration des différentes échelles peut être réalisée en supprimant les items peu compréhensibles ou mal adaptés à la population visée, ainsi que les items peu représentatifs du concept, ambigus ou redondants [Roussel – 2005].

En ce qui concerne la présente étude, la validité de contenu sera évaluée lors d'un pré-test auprès d'un échantillon composé de d'experts du domaine (responsables d'agence) et d'intérimaires.

1.2.2. La validité de construit ou validité de trait

Tester la validité de construit dans une logique quantitative revient généralement à évaluer la représentativité statistique des items d'une échelle de mesure [Drucker & al.. 1999] : il s'agit pour le chercheur de s'assurer de la validité convergente et discriminante du construit. La validité convergente est assurée si l'on peut vérifier que les énoncés issus d'une échelle et destinés à mesurer un construit particulier soient suffisamment corrélés entre eux, ou que différents outils de mesure du même construit donnent des résultats convergents. La validité discriminante est assurée lorsque les items de l'échelle se distinguent suffisamment des items censés mesurer d'autres phénomènes voisins.

Plusieurs méthodes sont utilisables pour apprécier la validité convergente et discriminante : l'une des plus souvent recommandée est la matrice multitrait- multiméthodes [MTMM, Campbell & Friske – 1959], qui consiste à mesurer un construit multidimensionnel en utilisant plusieurs méthodes : l'application de cette méthode dans le champ de l'implication organisationnelle conduit par exemple le chercheur à mesurer les trois composantes de l'implication (affective, calculée et normative) en utilisant plusieurs échelles de mesure (ex : échelles tri-dimensionnelle de Meyer & Allen et d'O'Reilly & Caldwell). Cette méthode est rigoureuse, mais difficile à employer dans une étude telle que la notre, comportant plusieurs échelles de mesure, car elle alourdit considérablement le questionnaire et le rend difficilement testable.

La difficulté de mise en place de la méthode MTMM conduit souvent les chercheurs à lui préférer l'analyse factorielle, dont l'utilisation dans le domaine de la GRH progresse constamment depuis les années 60 [Igalens & Roussel – 1998, p.153]. Ce type d'analyse permet de juger à la fois de la validité convergente et discriminante des échelles de mesure. On peut schématiquement affirmer

que la validité convergente (capacité à mesurer précisément le concept) est assurée lorsque les résultats de l'analyse factorielle montrent que les items d'une échelle sont suffisamment corrélés à un facteur, et la validité discriminante (capacité à mesurer uniquement le concept) est assurée lorsque ces items sont corrélés de manière claire à un seul axe si le construit théorique sous-jacent est considéré comme unidimensionnel, ou à plusieurs axes (sans chevauchement) lorsque le construit est considéré comme multidimensionnel.

Dans le cadre de la présente étude, nous utiliserons donc l'analyse factorielle pour évaluer la validité des échelles de mesure multiples destinées à mesurer certaines variables. Bien que plusieurs échelles de notre questionnaire aient déjà fait l'objet de procédures de validation, nous entreprendrons de nouveaux tests, en suivant les propositions d'Igalens & Roussel [2005 p.106] qui recommandent de procéder à une vérification de fiabilité et de validité pour les questionnaires incorporant des outils préexistants, mais destinés à être utilisés dans un contexte différent de celui pour lequel ils ont été conçus.

1.2.3. La validité prédictive ou validité nomologique

La validité nomologique ou prédictive d'un construit peut se définir comme « le degré auquel les prévisions fondées sur un concept, qu'un instrument est censé mesurer, sont confirmées » [Zaltman & al- 1973, cités par Drucker Godard & al. – 1999 p 259]. Comme le remarquent Igalens & Roussel [1998], cette validité prédictive ne peut être évaluée qu'au moment de l'étude des relations empiriques entre les concepts. Il s'agit alors de vérifier que les prédictions théoriques concernant les relations entre les concepts mesurés sont vérifiées par des tests portant sur les données empiriques. Dans le cas de notre recherche, la validité nomologique pourra donc être évaluée lors de la phase de discussion des résultats obtenus suite aux tests réalisés sur les hypothèses de notre modèle (*voir chapitre suivant*).

2. Le pré-test du questionnaire

Le pré-test de notre questionnaire d'enquête poursuit deux objectifs : évaluer la validité de contenu et la faisabilité. Ce pré-test a été auprès d'un échantillon de 11 personnes : 2 chercheurs, 5 intérimaires et 4 responsables d'agences.

Nous souhaitons principalement évaluer lors de ce pré-test la transférabilité des échelles de mesure de certaines variables auprès d'une population de salariés intérimaires. Il faut en effet remarquer que la validité de contenu au sens strict ne constituait pas un problème majeur, étant donnée que les échelles de mesure que nous avons choisi d'utiliser ont toutes été préalablement testées auprès d'échantillons de salariés ; certaines d'entre elles (ex : échelle d'implication organisationnelle, échelle de soutien organisationnel perçu) ont même été utilisées dans de très

nombreuses études : on peut donc considérer que leur validité de contenu a été déjà éprouvée au sein de la communauté scientifique.

Il s'agissait donc surtout pour nous de vérifier que le questionnaire d'enquête était adapté à la population visée, et que les différents énoncés pouvaient s'appliquer à la situation des salariés intérimaires.

Sur le plan pratique, nous avons demandé aux salariés intérimaires de remplir le questionnaire en notre présence, en faisant part de leurs remarques ou de leurs difficultés au fur et à mesure de l'exercice. En ce qui concerne les responsables d'agence et les enseignants, nous leur avons demandé de lire et commenter le document.

Deux remarques importantes sont apparues de manière récurrente dans le discours des professionnels du secteur (intérimaires et responsables d'agence) : le questionnaire original, qui comportait 98 questions a été jugé trop long, et les échelles de mesure de type Likert ont été jugées répétitives, avec certains énoncés considérés comme trop complexes ou peu réalistes.

En premier lieu, nous avons constaté que c'est la longueur du questionnaire qui a fait l'objet des critiques les plus sévères : les intérimaires ont jugé le questionnaire trop long, et les responsables d'agence nous ont affirmé que l'administration d'une telle enquête était quasi-impossible, compte tenu de la population concernée. Plusieurs responsables nous ont affirmé procéder eux-mêmes à des enquêtes de satisfaction auprès de leurs intérimaires, et ont mis l'accent sur le très faible taux de retour obtenu. L'une des responsables d'agence ayant participé au pré-test nous a par exemple confié les difficultés et les réticences des intérimaires peu qualifiés vis-à-vis de la « paperasse », même lorsqu'il s'agit de documents importants comme les contrats ou les certificats de travail.

En second lieu, la présence de plusieurs groupes de questions avec des formats de réponse sous forme d'échelles (de type Likert ou sémantique-différentielles) a dérouté plusieurs répondants. Le côté répétitif de l'exercice a été souvent souligné, et plusieurs items ont posé des difficultés de compréhension ou ont été jugés inadaptés.

On peut illustrer ces difficultés en prenant l'exemple de l'énoncé n°10 de l'échelle d'implication dans la valeur-travail (« Je continuerai probablement à travailler même si je n'avais pas besoin d'argent »). Plusieurs répondants intérimaires ont jugé cette question fantaisiste ou mal posée : un intérimaire nous a par exemple fait remarquer que tout dépendait du type travail et du montant de la somme d'argent¹³⁵. Certains responsables d'agence ont pour leur part considéré que cette question était peu pertinente vis à vis de répondants exerçant des professions pénibles et peu valorisées.

Face à l'ensemble de ces critiques, nous avons donc choisi de réduire la longueur de plusieurs

¹³⁵ Il est intéressant de noter que cette limitation a été également mise à jour par Harpaz, [1988] lors d'enquêtes sur la valeur-travail incorporant ce type de « questions de loterie ».

échelles, en éliminant les items jugés peu pertinents ou difficiles à comprendre par les répondants et les experts. Nous avons également modifié la rédaction de certaines questions pour les rendre plus accessibles. La liste des items écartés à la suite du pré-test est présentée en annexe 6. La réduction de la longueur des échelles peut apparaître critiquable, en raison des risques d'affaiblissement de la validité et de la fiabilité des outils, mais ces risques doivent être comparés aux risques de lassitude des répondants, d'effet de halo et de multiplication de non-réponses causés par un questionnaire trop complexe.

La réduction du nombre d'items dans les échelles de mesure d'attitudes ne réduit pas nécessairement leur fiabilité, à condition d'en conserver un nombre suffisant : il est généralement admis qu'un minimum de trois items peut être suffisant pour assurer la fiabilité d'une échelle d'attitude multiple [Cook & al. – 1981], et que les échelles courtes sont considérées comme un moyen efficace de minimiser les biais de réponse [Schmitt & Stultz – 1985, Schrieschen & al. 1990]¹³⁶. Igalens & Roussel [2005, p 106] recommandent ainsi de se limiter à 4 à 6 items par construit ou variable, afin de limiter la longueur du questionnaire. Des versions réduites du questionnaire d'implication organisationnelle de Meyer & Allen ont par exemple été employées dans les études sur l'implication des intérimaires : on peut citer en exemple Connelly, Gallagher & Gilley [2006], qui utilisent 4 items par dimension d'implication.

A la suite du retrait d'un certain nombre d'énoncés, et de la reformulation de certains autres, notre questionnaire d'enquête a été réduit à 86 items. Cette version (présentée en annexe 7) a été soumise à une procédure de validation destinée à évaluer la fiabilité et la validité de construit des différentes échelles multiples utilisées dans notre modèle, que nous allons détailler dans les paragraphes suivants.

3 Constitution de l'échantillon et administration de l'enquête

3.1 Constitution de l'échantillon

3.1.1 La cible de l'enquête

Notre étude demeure exploratoire : les salariés intérimaires ont fait l'objet de peu d'études empiriques dans le domaine de la GRH en France : leurs attitudes (en particulier leur implication au travail) sont encore peu explorées.

Nous avons choisi de réunir un échantillon de convenance, fondé sur la variété des profils d'intérimaires : nous avons tenté de réunir des salariés des deux sexes, d'âge et de profession et de niveaux de qualification variés. Cette méthode empirique est la plus souvent employée dans les

¹³⁶ Une méta analyse menée par T. Hinkin [1995] portant sur 277 échelles de mesure utilisés dans des études menées en sciences sociales montre que les échelles les plus fréquemment utilisées comportent 4 items (suivies des échelles comportant 3, puis 5 items)

recherches quantitatives dans le domaine de la GRH [voir Igalens & Roussel – 1998 p121], et dans les sciences sociales en général : elle consiste pour le chercheur à contacter toutes les personnes possibles correspondant à la définition précise de la population ciblée.

Nous avons vu dans le deuxième chapitre que les salariés intérimaires ont des profils socioprofessionnels spécifiques : ils sont moins qualifiés que la moyenne des salariés, et ils travaillent principalement dans l'industrie et le BTP, et, dans une moindre mesure, dans le secteur tertiaire. Afin de varier notre échantillon tout en nous adaptant aux particularismes de la population enquêtée, nous avons pris contact avec plusieurs agences d'intérim, spécialisées dans des domaines variés. Quatre agences ont accepté de participer à notre enquête (deux agences généralistes, une agence spécialisée dans le BTP et une agence spécialisée dans l'industrie).

3.1.2. La détermination de la taille de l'échantillon

La détermination de la taille de l'échantillon est dépendante des techniques d'analyse de données choisies. Dans notre cas, nous avons prévu d'utiliser les méthodes d'analyse factorielle lors de l'analyse exploratoire des échelles de mesure, puis des méthodes d'équation structurelles lors de l'analyse confirmatoire et d'une partie des tests d'hypothèse (en association avec des analyses de variance et de régression).

La taille minimale conseillée pour un échantillon destiné à faire l'objet d'une analyse factorielle exploratoire varie selon les auteurs entre 4 et 10 répondants par variables [Hinkin – 1995]. Dans les études comportant plusieurs échelles, ce nombre se calcule en fonction du nombre d'énoncés de l'échelle la plus longue (dans notre étude, il s'agit de l'échelle d'implication organisationnelle, qui comporte 14 énoncés). Dans ce cadre, la taille minimale de l'échantillon serait donc de 70 salariés et la taille correspondant au niveau d'exigence le plus élevé devrait être de 140 répondants.

En ce qui concerne l'usage des méthodes d'équation structurelles, Roussel & al [2002, p. 49] recensent différentes recommandations, et proposent des seuils minimum concernant le nombre de répondants (au moins 5 individus par paramètre à estimer dans un modèle d'équations structurelles, avec en tout état de cause un échantillon de 100 à 150 individus minimum). Dans un ouvrage plus récent Roussel [2005] recommande finalement de réunir au moins 200 individus lorsque des procédures d'analyse exploratoire et confirmatoire sont envisagées. Nous nous sommes finalement fixés un objectif de 200 répondants, permettant de satisfaire les seuils minimums exigés. La taille de cet échantillon nous apparaît également satisfaisante, car elle est supérieure à la moyenne des échantillons d'intérimaires réunis dans les études que nous avons collectées dans notre revue de littérature (*voir chapitre 2*).

3.2. Mode d'administration

Lors de nos entretiens exploratoires et des pré-tests du questionnaire avec les directeurs d'agence, nous avons envisagé les différents modes d'administration d'enquête.

Nous nous sommes rapidement rendus compte que l'administration de l'enquête par voie postale risquait d'être vouée à l'échec, car plusieurs responsables d'agence ont insisté sur le fait que le contact par courrier avec les intérimaires était souvent difficile (particulièrement semble-t-il pour les salariés des secteurs du BTP). La longueur de notre enquête et la complexité de certaines questions était également jugées excessives (malgré la reformulation et la suppression de certaines d'entre elles) : un responsable d'une grande enseigne nous a par exemple communiqué des questionnaires internes de satisfaction ne dépassant pas 10 énoncés pour lesquels il ne parvenait pas à obtenir un taux de retour suffisant.

Nous avons alors opté pour une auto administration encadrée du questionnaire, en contact direct [Roussel- 2005 p 254]. Cette méthode présente l'avantage de pouvoir expliquer aux salariés en face à face le « mode d'emploi » du questionnaire, mais elle réclame beaucoup de temps car les salariés intérimaires doivent être contactés individuellement : les ETT n'organisent en effet pratiquement jamais des réunions avec leurs salariés, ou des séances de formation collectives durant lesquelles les enquêtes auto administrées par contact direct sont réalisables auprès d'un ensemble de salariés.

Les intérimaires ont donc été sollicités individuellement pour remplir le questionnaire lors de leur passage dans l'agence, en restant sur place : ceci a permis de communiquer des consignes claires sur la manière de remplir le document, et de fournir des explications éventuelles. Nous avons contrôlé l'administration d'une partie des enquêtes, et les autres ont été confiées à des personnes travaillant dans les agences de travail temporaires.

Cette méthode s'est avérée efficace et rentable : la qualité globale des questionnaires recueillis est bonne (peu d'entre eux ont dû être éliminés ; le nombre de valeurs manquantes dans les questionnaires conservés est faible). L'enquête a été administrée dans quatre agences de travail temporaire (trois d'entre elles étaient situées dans la région PACA et une en Bretagne).

Nous avons recueilli un total de 217 questionnaires : 9 d'entre eux ont été éliminés pour différentes raisons (plusieurs pages non remplies, incohérences dans le renseignement des échelles de type Likert ou lassitude manifeste du répondant l'ayant conduit à interrompre le questionnaire).

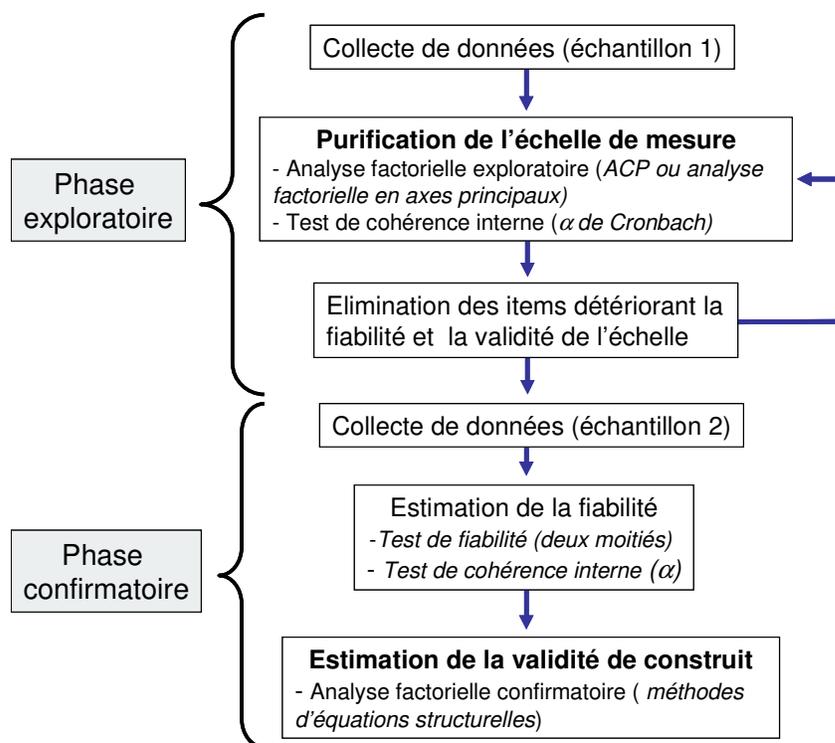
Au final, 208 questionnaires exploitables ont été conservés pour constituer notre échantillon d'enquête.

4. La validation des échelles de mesure

Un certain nombre de variables de notre modèle sont évaluées par des échelles de mesure multiples de type échelles de Likert. L'étude de la validité des construits mesurés par ce type d'échelle, extrêmement répandues en sciences de gestion, suit le plus souvent un certain nombre d'étapes extraites de la démarche générale de construction d'échelles de mesure proposée initialement par G. Churchill [1979] dans le domaine du marketing, et connue sous le nom de « paradigme de Churchill ». La méthode que nous allons utiliser a été exposée par Roussel [2005] : elle s'appuie sur une actualisation et un aménagement de la démarche originale de Churchill, qui prend en compte les principales critiques et améliorations proposées par différents auteurs depuis le milieu des années 80 [ex : Gerbing & Anderson – 1988 ; Fabrigar & al. – 1999 ; Mc Callum & al. – 1999]. La démarche générale de validation comporte deux grandes phases : une phase exploratoire destinée à purifier les instruments de mesure en utilisant une méthode itérative, et une phase confirmatoire visant à vérifier la fiabilité et la validité des instruments épurés.

Ces deux étapes peuvent être schématisées comme suit :

Figure 2 : La démarche de validation de construit dans le cas de variables mesurées par des échelles multiples (d'après Churchill- 1979 ; Roussel – 2005)



4.1 Analyse exploratoire

Afin de se rapprocher de la démarche recommandée par Churchill [1979], nous avons choisi effectuer nos opérations d'analyse exploratoire et confirmatoire sur deux échantillons différents, en divisant notre échantillon d'enquête initial ($n = 208$) en deux sous-échantillons comportant respectivement 68 et 140 répondants, en nous inspirant de la procédure choisie par plusieurs auteurs [ex : Belghiti & Briole- 2004, Manville – 2005]

Nous avons donc réuni un échantillon de 68 salariés intérimaires, qui permet de respecter le nombre minimum requis de répondants sur toutes les échelles de notre étude :

Echelle	Nombre d'items	Nombre de répondants	Rapport nombre d'items / nombre de répondants.
Soutien organisationnel perçu	6	68	11 répondants par item
Locus of control	6	68	11
Equité perçue	4	68	17
Implication organisationnelle	14	88 ¹³⁷	5,7
Implication professionnelle	8	68	8,5
Implication dans les valeurs du travail	7	68	9,7

4.1.1. Procédure et choix méthodologiques

La purification des instruments de mesure sera réalisée en utilisant conjointement l'analyse de cohérence interne (coefficient alpha de Cronbach) et l'analyse factorielle. En suivant les recommandations de Gerbing & Anderson [1988], nous évaluerons en premier lieu la dimensionnalité des échelles de mesure avant de mesurer leur cohérence interne.

Il existe deux méthodes principales pour réaliser une analyse factorielle de type exploratoire : l'analyse en composantes principales (ACP) et l'analyse factorielle « classique » en facteurs communs et spécifiques (AFCS, dite aussi analyse en axes principaux ou en facteurs principaux). L'ACP est longtemps restée la méthode la plus utilisée : elle est par exemple recommandée dans plusieurs ouvrages méthodologiques en langue française [ex : Evrard & al. – 2000 ; Igalens & Roussel – 1998], et est également recommandée par Churchill [1979] lorsqu'il s'agit d'épurer les échelles de mesure.

Il apparaît cependant que l'ACP est remise en cause lorsqu'elle est utilisée dans le cadre de procédures de validation d'échelles destinées à mesurer des construits latents. Roussel [2005 p.258] remarque que cette technique constitue une approximation de l'analyse factorielle classique : elle conduit à surestimer les contributions factorielles et les communautés (part de variance expliquée par les facteur communs), et à sous-évaluer les corrélations entre les facteurs. Ce phénomène s'explique principalement par une prise en considération différente de la variance selon les deux techniques [Baillargeon – 2003 p.5, Roussel- 2005 p.258] :

¹³⁷ 20 salariés supplémentaires ont été sélectionnés dans le reste de l'échantillon afin de parvenir au ratio requis

- Dans les modèles d'analyse factorielle classique, la variance de chaque variable est décomposée en trois parties : la variance commune (causée par les facteurs latents), la variance spécifique (assujettie à aucun facteur) et la variance d'erreur. L'analyse factorielle ne porte que sur la part de variance commune, et tente d'en extraire les facteurs sous-jacents¹³⁸.
- Dans l'ACP au contraire, aucune différence n'est faite entre les trois types de variance, et l'analyse porte sur la variance totale. Etant donné que les saturations et les communautés sont alors surestimées, et les corrélations entre facteur sous-estimées, la solution factorielle obtenue peut donc se révéler plus facile à interpréter, mais elle incorpore une part d'erreur.

Roussel [2005 p. 258] s'appuie alors sur plusieurs contributions récentes dans le domaine de l'analyse des données [ex : Nunnally & Bernstein – 1994 ; Falissard – 2001 ; Preacher & mc Callum – 2003] pour recommander parmi les différentes techniques l'analyse factorielle en axes principaux (AFAP¹³⁹) lorsque l'objectif est de construire un modèle linéaire comportant des variables latentes. Un autre intérêt de la technique d'analyse en axes principaux relevé par Fabrigar et ses collègues [Fabrigar & al. – 1999] est qu'elle peut s'appliquer dans le cas où la multinormalité des variables n'est pas assurée (au contraire de l'ACP, qui nécessite en principe une distribution multinormale des variables). Or, les variables non distribuées normalement sont relativement fréquentes dans le domaine des sciences sociales, tout particulièrement lorsque l'on travaille sur des variables d'intervalle de type Likert. Donada & Mbengue [1999, p 387], qui remarquent que les techniques de l'ACP et de l'analyse factorielle classique fournissent des résultats souvent comparables, résument clairement le critère de choix en fonction des objectifs du chercheur : « si l'objectif de recherche est simplement de résumer les données, alors l'ACP est le meilleur choix. Par contre, si l'objectif de recherche est de mettre en évidence une structure sous-jacente aux données (c'est-à-dire identifier des variables latentes ou des construits) alors l'analyse factorielle en facteurs communs et spécifiques est le choix qui s'impose »¹⁴⁰.

Etant donnée l'orientation de notre recherche, nous adopterons donc l'analyse factorielle « classique », en facteurs communs et spécifiques. Pour être plus précis nous utiliserons en

¹³⁸ Au niveau pratique, la variance commune est estimée par le coefficient de corrélation multiple (R^2). Avant de débiter l'analyse factorielle, les valeurs de R^2 sont placées dans la diagonales de la matrice de corrélation, en remplacement des valeurs 1 (qui restent utilisées dans les procédures d'ACP).

¹³⁹ Il existe plusieurs techniques pour l'analyse factorielle classique disponibles dans les logiciels d'analyse des données : méthode des moindres carrés non pondérés, méthode du maximum de vraisemblance, alpha-factorisation... Nous utiliserons en priorité la méthode d'analyse en axes principaux, recommandée par Roussel [2005 p. 258] pour l'épuration des échelles de mesure de construits.

¹⁴⁰ Une approche originale de la différenciation entre ACP et AFCS est proposée par Lauri Tarkkonen, professeur de mathématiques à l'université d'Helsinki (cité par C. Durand – 2003). Un extrait est disponible en annexe 8.

priorité l'analyse factorielle en axes principaux (AFAP) qui s'accommode d'une distribution s'écartant de la normale pour les variables analysées. En cas d'échec de l'analyse factorielle en axes principaux, nous utiliserons l'ACP afin de parvenir à dégager une structure factorielle pour toutes nos échelles de mesure¹⁴¹.

Une fois la méthode d'extraction des facteurs décidée, il faut procéder à l'interprétation, voire à l'épuration des solutions factorielles obtenues.

Les procédures recommandées pour l'épuration des échelles de mesure font appel à diverses techniques d'interprétation et règles de décisions propres à l'analyse factorielle.

Il est important de noter que la plupart de ces règles et recommandations ont un caractère empirique, arbitraire et subjectif [Evrard & al. – 2000 ; Preacher & mc Callum – 2003 ; Roussel – 2005] : elles tirent leur validité d'une forme de consensus au sein de la communauté des chercheurs.

Notre démarche d'épuration comprendra 4 étapes :

Etape 1 : Analyse des possibilités de factorisation de l'échelle.

La question préalable à toute analyse factorielle menée sur une série de données consiste à se demander si celles-ci sont factorisables, autrement dit « si elles forment un ensemble suffisamment cohérent pour qu'il soit raisonnable d'y chercher des dimensions communes » [Evrard & al. – 2000, p.377]. Plusieurs techniques permettent d'apprécier cette caractéristique :

- **L'examen de la matrice de corrélations** entre les énoncés de l'échelle doit faire apparaître plusieurs corrélations suffisamment élevées entre les variables.

- **L'utilisation du test MSA** (*measure of sampling adequacy*) de Kaiser, Meyer & Olkin (également appelé test KMO) qui permet de mesurer la capacité de factorisation au niveau de l'échelle ou de chacun des énoncés¹⁴². Un niveau de MSA élevé (proche de 1) pour l'échelle indique qu'il existe une solution factorielle statistiquement acceptable qui représente les relations entre les variables.

Les seuils d'acceptabilité préconisés par Kaiser [cité par Norusis/SPSS – 1993] sont les suivants :

MSA < 0,5	Inacceptable (échelle non factorisable)
0,5 < MSA < 0,6	Médiocre
0,6 < MSA < 0,7	Moyen
0,7 < MSA < 0,8	Bon
0,8 < MSA < 0,9	Très bon
MSA > 0,9	Excellent

- **L'utilisation du test de sphéricité de Bartlett**, qui vérifie l'hypothèse nulle selon laquelle

¹⁴¹ La méthode d'analyse factorielle en axes principaux ne permet pas toujours en effet d'obtenir une solution (dans certains cas, l'analyse ne converge pas, ce qui ne permet pas d'extraire les facteurs attendus). L'ACP présente l'avantage de toujours fournir une solution, qu'il convient bien évidemment d'interpréter.

¹⁴² L'adéquation (MSA) des items pris individuellement se trouve sur la diagonale de la matrice de corrélation anti-image (disponible en sortie sur le logiciel SPSS) : une valeur >0,5 est considérée comme acceptable.

toutes les corrélations seraient égales à zéro. Ce test est très sensible au nombre de cas; il est presque toujours significatif ce nombre de cas est suffisamment élevé (c'est-à-dire s'il y a plus de 5 observations par variable).

Etape 2 : Choix du nombre d'axes factoriels à retenir

Dans les études visant à isoler des construits latents, on peut considérer que chaque axe factoriel représente une dimension du construit. Les règles empiriques suivantes permettent d'aider à fixer le nombre de facteurs retenus :

- **La fixation d'un seuil minimum de variance expliquée.** On considère généralement qu'il faut retenir suffisamment d'axes pour restituer au minimum 50% de la variance de l'échelle [Igalens & Roussel – 1998] .
- **Le critère des valeurs propres (règle de Kaiser)** recommande de ne conserver que les facteurs dont les valeurs propres (*eigenvalues*, représentant schématiquement la quantité d'information extraite par les facteurs) sont supérieures à 1.
- **L'examen de la courbe des valeurs propres (*scree test*)** : La courbe des valeurs propres est construite en mettant en rapport le nombre de facteurs (en abscisse) avec leur valeur propre (en ordonnée). Cette courbe fait en principe apparaître un point d'inflexion (coude) à partir duquel la courbe s'aplatit : seuls les facteurs situés à gauche du point d'inflexion devraient être retenus. Ce test fournit une aide à la décision surtout lorsque le nombre de facteurs est important. Il existe également des test plus précis, notamment **le test MAP proposé par Velicer** [Zwick & Velicer – 1986], qui permet de déterminer le nombre d'axes s'ajustant le mieux à la distribution étudiée.

Etape 3 : Elimination des items nuisant à la validité des construits (épuration)

L'objectif de cette étape est de parvenir à une structure factorielle claire, c'est-à-dire une situation dans laquelle chaque énoncé de l'échelle est significativement corrélé à un facteur et un seul. Plusieurs techniques et règles de décision peuvent et doivent être conjointement utilisées lors de cette étape :

- **Le critère des communalités** : La communalité ou communauté d'un énoncé mesure la part de variance de cet énoncé restituée par les facteurs retenus. En deçà de certains seuils, la contribution des énoncés à l'explication du facteur est considérée comme médiocre, et peut conduire à leur élimination de l'échelle¹⁴³. Les valeurs de seuils proposées par Roussel [2005, p. 264] sont les suivantes :

¹⁴³ L'usage de ce seul critère pour éliminer une variable doit être évité : l'examen des contributions factorielles doit être également réalisé. Un énoncé qui a une faible communauté mais qui contribue clairement à un facteur et un seul peut être conservé.

Communauté > 0,8	Très bonne représentation (forte contribution)
Communauté comprise entre 0,65 et 0,8	Bonne représentation
Communauté comprise entre 0,4 et 0,65	Représentation moyenne
Communauté < 0,4	Représentation médiocre (élimination possible)

- **Le critère des contributions factorielles (*loadings*)** : la contribution (ou saturation) d'un énoncé sur un axe factoriel correspond à la corrélation entre cet énoncé et l'axe considéré. Ce critère est important, car il est à la base des jugements de validité convergente et discriminante. Il est en général admis qu'une contribution est considérée comme significative si elle est au moins égale à 0,3. Au-delà de 0,5, on peut la considérer comme acceptable.

Deux règles de décision sont en général associées à ces valeurs :

- Il est conseillé de supprimer les énoncés n'ayant aucune contribution supérieure ou égale à 0,5 sur au moins l'un des facteurs, afin d'améliorer la validité convergente de l'échelle
- Il est conseillé de supprimer les énoncés ayant des contributions supérieures à 0,3 sur plusieurs facteurs, afin d'améliorer la validité discriminante de l'échelle.

- **Le jugement du chercheur** : Si l'échelle testée a déjà fait l'objet de recherches antérieures, l'interprétation des axes factoriels peut être guidée par le niveau des connaissances accumulées sur le construit. Si l'on prend en exemple l'implication organisationnelle, de nombreux travaux ont été menés sur l'échelle tridimensionnelle de Meyer & Allen et ont pu valider une structure en trois dimensions (affective calculée et normative) : la connaissance préalable de ces trois dimensions peut guider l'interprétation des itérations factorielles.

La technique des rotations factorielles fournit une aide précieuse à la décision, en permettant souvent de rendre plus lisible le tableau des contributions factorielles. Elles doivent être utilisées lorsque la première itération fournit des résultats difficiles à interpréter. Deux types de rotations sont généralement utilisées : rotations orthogonales (Varimax, Quartimax et Equamax sur SPSS) et rotations obliques (Direct Oblimin ou Promax sur SPSS). Les rotations orthogonales sont fondées sur l'hypothèse que les axes factoriels sont indépendants, et les rotations obliques admettent la possibilité que les axes soient corrélés (ce qui est souvent le cas dans les sciences sociales, ou des construits peuvent posséder plusieurs dimensions corrélées¹⁴⁴). Certains auteurs [ex : Igalens & Roussel – 1998 ; Pedahzur & al – 1991] recommandent de réaliser systématiquement les 2 types de rotation, et de retenir la solution la plus facilement interprétable.

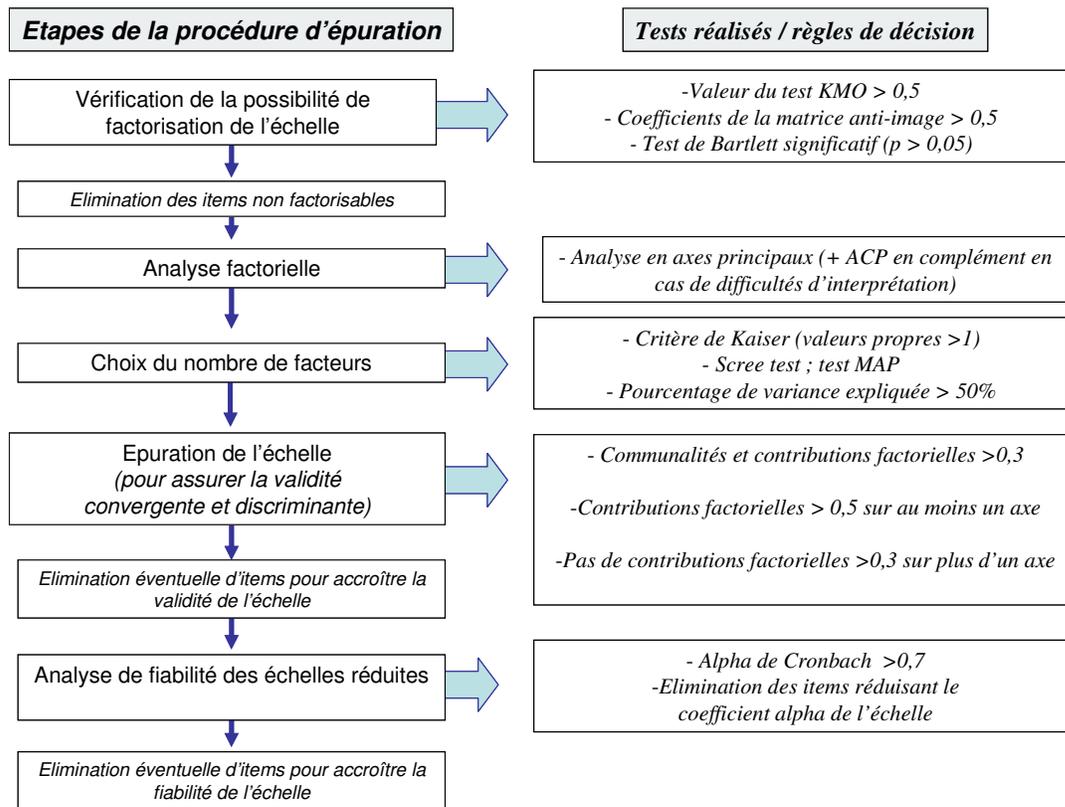
¹⁴⁴ L'échelle d'implication organisationnelle de Meyer et Allen fournit un bon exemple de construit multidimensionnel avec sous-dimensions corrélées : l'implication normative apparaît généralement corrélée à l'implication affective dans les études empiriques (voir la méta-analyse de Meyer & al. – 2002)

Etape 4 : Test de la fiabilité des échelles obtenues

Le test de la fiabilité des échelles obtenues après épuration éventuelle est réalisé en utilisant le coefficient **Alpha de Cronbach**. Il est encore possible d'épurer les échelles à ce stade en supprimant les items dont le retrait permet d'augmenter significativement le coefficient alpha.

Les règles guidant notre procédure d'épuration peuvent finalement être résumées dans le schéma suivant :

Figure 3 : La procédure d'épuration des échelles de mesure.



Une remarque importante doit être faite en conclusion de ce paragraphe méthodologique. Lorsque l'analyse factorielle amène à conclure que des items doivent être éliminés, il est souvent nécessaire d'arbitrer entre la fiabilité de l'échelle et la validité de construit : la suppression d'énoncés peut réduire le niveau du coefficient alpha, mais permet en contrepartie d'obtenir des structures factorielles plus claires, garantes d'une meilleure validité convergente et discriminante. Nous privilégierons les résultats issus de l'analyse factorielle, sous réserve de conserver des coefficients alpha voisins de 0,7, qui demeure la norme la plus commune d'acceptabilité. Nous tenterons également de conserver au minimum trois énoncés par construit ou dimension, afin de pouvoir assurer une représentation suffisamment riche des variables de notre modèle.

4.1.2. Analyse factorielle exploratoire et épuration des échelles de mesure

Les analyses factorielles et les tests d'homogénéité ont été réalisés à l'aide du logiciel SPSS (version 11). Nous avons inclus dans cette analyse toutes les échelles multidimensionnelles de type Likert utilisées dans notre modèle : Locus of control (LOC), volition, valeur intérimaire, équité relative perçue, soutien organisationnel perçu (SOP), autonomie perçue, implication organisationnelle, implication dans la profession et implication dans les valeurs du travail. Les outils de mesure de l'insécurité perçue et de la précarité objective ne sont pas étudiés dans cette partie, car nous verrons qu'il ne s'agit pas de variables latentes classiques mesurées grâce à des échelles. Nous développerons le processus de validation de ces deux construits dans le paragraphe 5 de cette section.

Nous allons à présent présenter les principaux résultats de l'analyse exploratoire des échelles de mesure. Les détails des analyses et des procédures d'épuration sont disponibles en annexe 8.

4.1.2.1 Echelle de locus of control

L'échelle de Locus of control professionnel a fait apparaître une capacité de factorisation médiocre : plusieurs corrélations faibles et/ou négatives apparaissent sur la matrice de corrélations, le niveau du test KMO est médiocre (= 0,57) et deux énoncés sont mal représentés sur la diagonale de la matrice anti-image. La procédure d'épuration nous conduit à éliminer ces items (LOC 1 & LOC2).

Les résultats obtenus avec une échelle réduite à 4 items sont les suivants :

<i>Enoncés de l'échelle</i>	<i>Communautés</i>	<i>Contributions factorielles</i>
LOC 3	,290	,538
LOC 4	,200	,447
LOC 5	,632	,795
LOC 6	,364	,603
Alpha = 0,68	Variance totale expliquée = 37 %	

La solution factorielle extraite est insatisfaisante, la variance totale expliquée atteint à peine 37%, la plupart des communautés sont à un niveau insuffisant, et les contributions factorielles sont peu élevées en moyenne. La fiabilité apparaît légèrement en deçà des normes d'acceptabilité (< 0,7).

La qualité générale de l'échelle de locus of control est donc médiocre : étant donné qu'il ne s'agit pas d'une variable centrale dans notre modèle de recherche, nous décidons de ne pas la prendre en compte cette attitude pour la suite de notre recherche.

4.1.2.2. Echelle de préférence pour l'intérim (volition)

L'échelle se révèle moyennement factorisable (KMO = 0,61). La première itération nous conduit à une solution restituant pratiquement 50% de la variance de l'échelle, avec les résultats suivants :

<i>Enoncés de l'échelle</i>	<i>Communautés</i>	<i>Contributions factorielles</i>
Volition 1	,603	,777
Volition 2	,183	,428
Volition3	,553	,744
Alpha = 0,67	Variance totale expliquée = 48,6%	

La solution factorielle proposée pour l'échelle réduite est unidimensionnelle ; elle restitue moins de 50% de la variance, et les trois énoncés contribuent significativement sur un seul facteur. Le deuxième énoncé apparaît toutefois très mal représenté : son retrait limite le nombre d'énoncés de l'échelle, mais permet d'obtenir une meilleure fiabilité : le coefficient Alpha passe à 0,73, et les deux énoncés expliquent 80% de la variance de l'échelle. Nous décidons donc de réduire l'échelle à deux items (volition 1 et volition3).

4.1.2.3 : Echelle d'employabilité perçue (valeur intérimaire)

L'échelle se révèle peu factorisable (KMO = 0,54). L'usage de l'analyse factorielle en axes principaux ne permet pas de converger vers une solution. Nous utilisons donc l'analyse en composantes principales (ACP). La première itération nous conduit à une solution restituant plus de 50% de la variance de l'échelle, avec les résultats suivants :

<i>Enoncés de l'échelle</i>	<i>Communautés</i>	<i>Contributions factorielles</i>
Employabilité 1	,824	,908
Employabilité 2	,329	,573
Employabilité 3	,575	,759
Alpha = 0,63	Variance totale expliquée = 57%	

La fiabilité de l'échelle est inférieure à la norme habituelle, mais demeure acceptable dans une étude à caractère exploratoire. Nous la conservons donc en l'état en attendant les résultats de l'analyse factorielle confirmatoire, qui nous permettront de décider définitivement de son maintien ou de son retrait.

4.2.1.4 : Echelle d'autonomie perçue

L'échelle d'autonomie révèle une capacité de factorisation trop faible (test KMO = 0,47) : le premier item (correspondant à la possibilité de refuser des missions) est très mal représenté sur la diagonale de la matrice anti-image. Nous décidons donc de limiter l'échelle à deux items, qui ont trait à la faculté de l'intérimaire à négocier sur son salaire ou le contenu de son travail.

Le coefficient Alpha de fiabilité pour ces deux items est de 0,86. Nous décidons donc de conserver l'échelle réduite à deux énoncés.

4.1.2.5. Echelle de soutien organisationnel perçu

La capacité de factorisation de l'échelle de soutien perçu est apparue bonne (KMO = 0,79).

Suite à la première itération de l'analyse factorielle, un énoncé mal représenté dans le tableau des communautés et présentant une contribution factorielle trop faible ($< 0,3$) a été éliminé (POS 1) : la variance expliquée est passée à 49 %. On peut noter que cet item est le plus généraliste (« mon agence s'intéresse vraiment à mon bien-être »). Les items décrivant des actions précises ont tous été sélectionnés, ce qui indique que les sujets interrogés ont peut-être été plus à l'aise pour former des jugements cohérents sur le soutien perçu de la part de l'agence à partir de propositions concrètes.

Les résultats obtenus avec une échelle réduite à 5 items sont les suivants :

<i>Enoncés de l'échelle</i>	<i>Communautés</i>	<i>Contributions factorielles</i>
POS2	,483	,695
POS3	,325	,570
POS4	,423	,651
POS5	,443	,665
POS6	,786	,886
Alpha = 0,81	Variance totale expliquée = 49 %	

La solution factorielle proposée correspond à un construit unidimensionnel, en adéquation avec les résultats d'Eisenberger & al. [1986], auteurs de l'échelle originale. Les contributions factorielles sont significatives, et regroupées sur un seul axe. Le niveau du coefficient alpha est très satisfaisant. Nous conservons donc l'échelle de soutien organisationnel perçu sous sa forme réduite à cinq énoncés.

4.1.2.6. Echelle d'équité relative perçue

La capacité de factorisation de l'échelle est apparue suffisante (test KMO = 0,78)

La première itération nous a conduit à une solution restituant pratiquement 50% de la variance de l'échelle, avec les résultats suivants :

<i>Énoncés de l'échelle</i>	<i>Communautés</i>	<i>Contributions factorielles</i>
Équité 1	,456	,675
Équité 2	,654	,809
Équité 3	,470	,685
Équité 4	,502	,709
Alpha = 0,79	Variance totale expliquée = 49,3 %	

La solution factorielle proposée pour l'échelle réduite est unidimensionnelle ; elle restitue environ 50% de la variance, et les quatre énoncés contribuent significativement sur un seul facteur. Nous pouvons considérer que la structure factorielle proposée est claire. Le niveau de fiabilité de l'échelle étant acceptable, nous la conservons sous sa forme originale.

4.1.2.7. Echelle d'implication organisationnelle

L'échelle d'implication organisationnelle présente une bonne capacité de factorisation (test KMO = 0,85). La première itération conduit à une solution en deux axes (variance totale expliquée = 57%), confirmée par l'examen du graphique des valeurs propres et le test MAP de Velicer, qui s'avère assez difficilement interprétable.

La structure la plus claire apparaît à la suite d'une rotation Varimax :

<i>Énoncés de l'échelle</i>	<i>Communautés</i>	<i>Contributions factorielles</i>	
		<i>Axe1</i>	<i>Axe2</i>
IOA1	,587	,766	
IOA2	,796	,896	
IOA3	,401	,636	
IOA4	,682	,806	
IOA5	,616	,778	
IOC1	,412	-,365	,570
IOC2	,598		,751
IOC3	,467		,673
IOC4	,391		,542
IOC5	,416		,647
ION1	,489	,669	
ION2	,384	,605	
ION3	,421	,418	0,452
ION4	,619	,753	
Variance totale expliquée = 49 %			

La matrice des contributions montre clairement que la dimension affective de l'implication est très bien représentée dans sa quasi-totalité, sur le même axe. On note également que la dimension normative apparaît clairement, mais de manière problématique : elle se confond totalement avec la dimension affective. Cette difficulté est bien connue des chercheurs ayant travaillé sur le modèle tri-dimensionnel d'Allen et Meyer [ex : Dunham & al.- 1994 ; Ko & al. – 1997] : la

distinction entre les trois composantes de l'implication est généralement prouvée par analyse factorielle confirmatoire dans les études anglo-saxonnes [Meyer & al. – 1990 ; Hackett & al. – 1994] et françaises [Paillé – 2002 ; Belghiti & Briole – 2004], mais les dimensions affective et normative sont significativement corrélées. La dimension calculée est également bien représentée sur le deuxième axe. Nous pouvons remarquer que deux items (IOC 1 et ION 3) apparaissent complexes, car ils présentent des contributions significatives ($> 0,3$) sur deux axes. Afin d'obtenir la solution la plus claire possible, nous choisissons d'éliminer successivement ces deux derniers items, et nous procédons à deux nouvelles itérations. La solution retenue à l'issue de la dernière itération comporte toujours deux axes (la variance totale expliquée passe à 53%). Les contributions factorielles de l'échelle réduite à 12 items, après rotation Varimax sont présentées dans le tableau suivant :

Enoncés de l'échelle	Communautés	Contributions factorielles	
		Axe1	Axe2
IOA1	,614	,780	
IOA2	,774	,878	
IOA3	,379	,615	
IOA4	,694	,809	
IOA5	,627	,780	
ION1	,458	,648	
ION2	,407	,613	
ION4	,651	,772	
IOC2	,622		,780
IOC3	,487		,694
IOC4	,390		,581
IOC5	,360		,591
Variance totale expliquée = 53 %			
Alpha sous échelle implication affective normative= 0,79			
Alpha sous-échelle calculée = 0,75			
Alpha échelle globale = 0,84			

La solution obtenue après rotation Varimax apparaît claire et interprétable. La situation présentée dans la matrice ci-dessus est cependant originale : les dimensions affective et normative de l'implication sont bien représentées sur le premier axe, mais elles restent parfaitement imbriquées : nous choisissons donc de fusionner les sous-échelles affective et normative (8 items en tout) en un construit commun : l'implication-affective/normative. La sous-dimension calculée, qui regroupe les 4 items contribuant au facteur 2 est nettement apparente, avec des contributions factorielles moyennes. Nous décidons donc de conserver cette solution.

Les analyses de fiabilité doivent désormais être également envisagée au niveau de chaque sous-échelle, étant donné que nous obtenons deux dimensions dans l'échelle d'implication organisationnelle,

Nous pouvons remarquer que l'homogénéité de la dimension affective/normative est très bonne, ce qui nous conforte dans la décision de conserver cette sous-échelle d'implication organisationnelle modifiée.

Le niveau du coefficient alpha de la dimension calculée est plus faible, mais demeure dans les normes acceptables dans le cadre d'une étude exploratoire. Nous avons remarqué que la notion de « *side-bets* » (investissements passés) et de manque d'alternatives étaient peut-être inadaptées au monde de l'intérim. Il est en effet aisé pour un intérimaire de passer d'une agence à l'autre, et les méthodes de fidélisation des agences (consistant avant tout à assurer une continuité des missions) ne sont pas vraiment motivantes pour un intérimaire qualifié, en tout cas pas suffisantes pour constituer un frein à sa mobilité. L'analyse factorielle confirmatoire nous permettra de justifier ou non le maintien de cette dimension calculée.

4.1.2.8. Echelle d'implication dans la profession

L'échelle d'implication dans la profession présente une très bonne capacité de factorisation (test KMO = 0,82). La première itération fait apparaître une solution à deux facteurs, qui expliquent 57% de la variance de l'échelle.

La matrice des contributions factorielles se révélant difficile à interpréter, différentes rotations ont été effectuées. Les résultats obtenus après une rotation oblique (direct oblimin) ont été privilégiés, car il s'avère que les deux facteurs sont significativement corrélés ($r = 0,47$), ce qui rend les rotations orthogonales du type Varimax moins pertinentes¹⁴⁵. Les résultats finalement obtenus sont les suivants :

Enoncés de l'échelle	Communautés	Contributions factorielles	
		Axe1	Axe2
IPA1	,568	,649	
IPA2	,716	,709	
IPA3	,554	,720	
IPA4	,613	,846	
IPC1	,623		,717
IPC2	,491		,763
IPC3	,549		,622
IPC4	,476		,506
Variance totale expliquée = 57 %			
Alpha sous échelle implication affective = 0,85			
Alpha sous-échelle calculée = 0,80			
Alpha échelle globale = 0,86			

¹⁴⁵ Les rotations orthogonales présupposent que les axes factoriel sont indépendants (non corrélés). Les rotations obliques sont quant à elles adaptés aux situations dans lesquelles les facteurs sont corrélés (ce qui est fréquemment le cas dans les recherches en sciences sociales).

La solution factorielle obtenue après rotation oblique est claire¹⁴⁶ : nous obtenons deux facteurs nettement différenciés, correspondant chacun à une dimension théorique de l'implication professionnelle : le facteur 1 regroupe les 4 items correspondant à la dimension affective de l'implication professionnelle affective et le facteur 2 regroupe les 4 items correspondant à la dimension calculée. Le niveau des coefficients Alpha pour l'échelle complète, ainsi que pour chaque sous-échelle est très satisfaisant.

Nous décidons de conserver l'échelle bidimensionnelle d'implication dans la profession sous sa forme originale, en distinguant une dimension affective et une dimension calculée, conformément à notre modèle général.

4.1.2.9. Echelle d'implication dans la valeur travail

La capacité de factorisation de l'échelle apparaît bonne (test KMO = 0,74).

La première itération nous conduit à une solution en deux axes factoriels (avec une variance totale expliquée de 41 %). L'examen des communautés et des contributions factorielles après les différentes rotations nous a permis de constater la présence de deux items problématiques (un item complexe : IMPT6, et un item à faible contribution : IMPT4), qui ont été retirés.

La procédure d'épuration nous a permis d'aboutir à une solution comportant un seul axe factoriel, qui a permis d'augmenter légèrement la variance expliquée (44 %), mais celle-ci n'atteint pas 50%.

Les résultats de l'échelle réduite à 5 items sont les suivants :

<i>Enoncés de l'échelle</i>	<i>Communautés</i>	<i>Contributions factorielles</i>
IVT 1	,299	,547
IVT 2	,593	,645
IVT 3	,520	,721
IVT 5	,380	,617
IVT	,593	,770
Alpha = 0,78	Variance totale expliquée = 44 %	

Les contributions factorielles sont significatives sur un seul axe, la variance expliquée est inférieure à la norme, mais le niveau du coefficient alpha de Cronbach est tout à fait acceptable

¹⁴⁶ La matrice présentée, disponible après une rotation oblique (pattern matrix) est destinée à faciliter l'interprétation des résultats : la matrice standard des contributions factorielles appelée matrice structurale fait en effet apparaître naturellement de nombreux chevauchements, résultant de la corrélation entre les axes [Tabachnick & Fidell – 2000]. La « pattern matrix » comporte seulement les corrélation entre les items et les facteurs.

dans le cadre d'une recherche exploratoire. Nous décidons donc de conserver l'échelle d'implication dans les valeurs du travail sous sa forme réduite.

4.1.3. Commentaires

Les échelles modifiées résultant de la procédure d'épuration sont résumées dans le tableau suivant :

Tableau 2 : Echelle de mesures retenue à l'issue de l'analyse exploratoire

<i>Echelle</i>	<i>Fiabilité (coefficient Alpha)</i>	<i>Nombre d'énoncés</i>
Préférence pour l'intérim	<i>0,73</i>	<i>2</i>
Valeur intérimaire perçue	<i>0,67</i>	<i>3</i>
Autonomie perçue	<i>0,86</i>	<i>2</i>
Soutien organisationnel perçu	<i>0,81</i>	<i>5</i>
Equité relative perçue	<i>0,79</i>	<i>4</i>
Implication organisationnelle (échelle complète)	<i>0,84</i>	<i>12</i>
- dimension affective normative	<i>0,79</i>	<i>8</i>
- dimension calculée	<i>0,75</i>	<i>4</i>
Implication dans la profession (échelle complète)	<i>0,86</i>	<i>8</i>
- dimension affective	<i>0,80</i>	<i>4</i>
- dimension calculée	<i>0,85</i>	<i>4</i>
Implication dans la valeur-travail	<i>0,78</i>	<i>5</i>

Cinq échelles ont été réduites (volition, autonomie perçue, soutien perçu, implication organisationnelle, implication dans la valeur-travail), et une échelle n'a pas été retenue en raison de sa qualité insuffisante (locus of control).

La principale conclusion à tirer de cette première analyse porte sur les échelles d'implication. L'implication organisationnelle telle que nous l'envisageons pour les intérimaires est tout à fait spécifique : elle comporte une dimension affective/normative homogène, décrivant un attachement à l'agence de travail temporaire, et une dimension calculée décrivant une implication « contrainte ».

Nous pouvons en conclure à ce stade de la recherche que les liens organisationnels très particuliers que tissent les intérimaires avec leur agence de référence ont un impact sur leur implication : il semble que l'attachement à l'agence soit fondée à la fois sur un attachement affectif et sur un sentiment de devoir : on peut penser que les intérimaires reconnaissent le rôle clé de leur agence dans le nombre et la qualité des missions, et que leur attachement comporte une part de gratitude, qui implique un besoin de réciprocité.

En ce qui concerne l'implication dans les valeurs du travail, une approche multidimensionnelle de ce construit a été envisagée lors de la construction du questionnaire, mais n'a pas été retenue à

l'issue de l'analyse exploratoire. Il est sans doute nécessaire dans l'avenir de mettre au point un outil de mesure multidimensionnel nouveau, plus fiable que celui que nous avons utilisé.

Les échelles de mesure épurées vont à présent faire l'objet d'une analyse factorielle confirmatoire.

4.2. Analyse confirmatoire

4.2.1. Les méthodes d'équations structurelles, outils de l'analyse factorielle confirmatoire

L'analyse factorielle confirmatoire est utile pour valider les structures factorielles issues de l'analyse exploratoire. La différence majeure avec l'analyse factorielle exploratoire est que le chercheur présuppose une série de liens entre les items et les construits qu'ils sont censés mesurer, et va tester la qualité de cette structure établie a priori. Ces tests sont en général effectués en recourant à des méthodes d'équation structurelles, les analyses factorielles confirmatoires étant considérées comme des modèles particuliers d'équation structurelle [Roussel & al. – 2002].

Les méthodes d'équations structurelles avec variables latentes et erreurs de mesure [Jöreskog-1978] ont été mises au point pour permettre de procéder à des tests de relations causales entre plusieurs variables latentes explicatives et une ou plusieurs variables latentes à expliquer.

Lors des phases d'analyse factorielle confirmatoire, ces méthodes d'équations structurelles sont utilisées pour tester des modèles dits de mesure, c'est-à-dire des relations entre plusieurs variables observées (dans notre cas, il s'agit des items appartenant à une échelle de mesure d'attitude) et une variable latente correspondant au construit que les indicateurs sont censés mesurer. L'intérêt de ces modèles pour des analyses factorielles est notamment lié à la prise en compte des erreurs de mesure¹⁴⁷.

Le test d'un modèle de mesure par les méthodes d'équations structurelles consiste à mesurer la qualité d'ajustement d'un schéma de relations théoriques (proposé par le chercheur) aux données empiriques issues d'un échantillon : plusieurs logiciels spécialisés permettent de réaliser ces mesures ; les plus diffusés sont Lisrel [Jöreskog & Sörbom – 1996], Amos [Arbuckle- 1995], et EQS [Bentler- 2001].

Lors du test, il est possible de mesurer la qualité d'ajustement d'un seul modèle, mais également de comparer plusieurs modèles alternatifs concurrents : cette seconde procédure est plus riche et permet de compléter efficacement l'analyse exploratoire. Dans notre étude, il sera par exemple

¹⁴⁷ Les erreurs de mesure peuvent être calculées pour toutes les variables (les estimations des contributions factorielles tiennent compte uniquement de la variance commune, ce qui permet de disposer d'estimations d'erreurs pour toutes les mesures). Ceci permet de réaliser des estimations plus réalistes des contributions factorielles des items.

intéressant de comparer la structure du construit d'implication organisationnelle issue de notre étude exploratoire avec un modèle théorique « standard » comportant trois dimensions clairement discriminées.

La qualité de l'ajustement du modèle aux données (*goodness of fit*) est évaluable par des indices d'ajustement (*fit indices*) : ces indices sont nombreux (pour une présentation détaillée, on peut se référer à Roussel & al. -2002 p. 62 et suiv. et à Sharma & al. - 2005). Ceux-ci sont généralement classés en trois groupes :

- Les indices absolus évaluent dans quelle mesure le modèle proposé reproduit correctement les données collectées. Les principaux indices absolus sont présentés dans le tableau ci-dessous, avec leurs normes habituelles d'interprétation [Roussel & al. – 2005 ; Sharma & al. – 2005].

Tableau 3 : Les indices de mesure d'ajustement absolus

Indice	Description	Valeur-clé
Chi-deux	Indice permettant d'évaluer dans quelle mesure le modèle postulé reproduit les données collectées. Il est très dépendant de la taille de l'échantillon (le modèle est presque toujours rejeté si le nombre d'observation est > 200)	Pas de valeur clé. Probabilité p associée < 0,05
GFI et AGFI (<i>goodness of fit index et adjusted goodness of fit index</i>)	Indices analogues au coefficient R ou R2 dans la régression (ils mesurent la part de variance-covariance des variables expliquée par le modèle). Ils sont peu sensibles à la taille de l'échantillon mais sensibles à la complexité du modèle. Ils ne sont pas très efficaces pour détecter les modèles mal spécifiés	> 0,9
Gamma 1 et 2	Indices représentant les variances-covariances du GFI et de l'AGFI	> 0,9
PNI PNNI	Indices de non centralité : mesurent la non adéquation à partir d'une matrice de variance-covariance induite sur la population totale. Ils sont peu sensibles à la taille de l'échantillon	PNI le plus faible possible PNNI > 0,95
RMR et SRMR (<i>root mean square residual et standardized RMR</i>)	Appréciation des résidus. Le RMR est utilisable si la matrice des données de départ est une matrice de corrélations et le SMRM est compatible avec une matrice des variances-covariances.	< 0,05 Le plus proche possible de 0
RMSEA (<i>root mean square error of approximation</i>)	Indice évaluant la différence moyenne attendue par degré de liberté dans l'échantillon. Il est indépendant de la taille de l'échantillon et de la complexité du modèle.	Si possible < 0,05 Acceptable jusqu'à 0,08

- Les indices incrémentaux sont fondés sur une comparaison entre le modèle testé et un modèle plus restrictif dit modèle de base (il s'agit en général d'un modèle nul dans lequel les variables observées seraient non corrélées). Les indices incrémentaux mesurent l'amélioration de l'ajustement lorsque le modèle testé est comparé au modèle nul. Le tableau ci-dessous présente les principaux indices d'ajustement incrémentaux disponibles, avec leurs normes d'interprétation :

Tableau 4 : Les indices de mesure d'ajustement incrémentaux

<i>Indice</i>	<i>Description</i>	<i>Valeur-clé</i>
NFI (<i>normed fit indice</i>)	Indice représentant la proportion de covariance totale entre les variables expliquées par le modèle testé lorsque le modèle nul est pris comme référence. Sa valeur est sous-estimée dans le cas d'échantillons de faible taille	> 0,9
TLI ou NNFI (<i>non normed fit index</i>)	Indice comparant le manque d'ajustement du modèle à tester à celui du modèle de base. Il efficace pour détecter les modèles mal spécifiés et assez peu sensible à la taille de l'échantillon [Sharma & al. – 2005]	> 0,9
IFI (<i>incremental fit index</i>)	Indice comparable au NNFI mais moins sensible pour les petits échantillons	> 0,9
CFI (<i>comparative fit index</i>)	Indice mesurant la diminution relative du manque d'ajustement	> 0,9

- Les indices de parcimonie permettent de choisir entre plusieurs modèles équivalents celui qui présente la meilleure parcimonie, c'est-à-dire celui qui permet d'estimer les données en recourant au minimum de paramètres à estimer : ces indices permettent de sélectionner les modèles les plus « économes », c'est-à-dire ceux dont la structure de relations est la plus simple. Le tableau ci-dessous présente les principaux indices de parcimonie, avec leurs normes d'interprétation.

Tableau 5 : Les indices de parcimonie

<i>Indice</i>	<i>Description</i>	<i>Valeur-clé</i>
Chi-deux /ddl (chi-deux normé)	Indice pouvant être utilisé pour comparer le niveau de parcimonie de plusieurs modèles	Ne doit pas dépasser 5. Si possible < 2
PNFI	Indices possédant les mêmes caractéristiques que le NFI. Permet de comparer des modèles alternatifs	Le plus faible possible
PGFI	Indice possédant les mêmes caractéristiques que le GFI. Permet de comparer des modèles alternatifs	Le plus faible possible
AIC CAIC	Ces indices visent à pénaliser les modèles complexes. Il permettent de comparer des modèles alternatifs, et sont assez peu sensibles à la taille de l'échantillon	Le plus faible possible

Le choix d'une liste d'indice d'ajustement adaptés à notre cadre d'étude peut être faite en fonction des critères suivants :

- La faible taille de nos échantillons amène à sélectionner en priorité des indices a priori peu sensibles à la taille de l'échantillon, ou réservés à des échantillons de faible taille : PNFI, NNFI, CN, CFI, TLI, RMSEA,
- L'usage des logiciels Lisrel et Amos nous amène à privilégier les indices d'ajustement disponibles en standard sur cet outil (ex : Gamma 1 et 2 non disponibles)

- La comparaison de modèles de taille différente encourage à utiliser les indices ajustés (AGFI, chi deux normé)
- Il paraît utile de faire figurer les différents types d'indices (absolus, incrémentaux et de parcimonie) afin de pouvoir effectuer un choix multicritères permettant de maximiser la qualité de la comparaison [Roussel – 2005]

En combinant tous ces critères, nous retiendrons finalement les indices et les seuils d'interprétation suivants:

<i>Indices absolus</i>			<i>Indices incrémentaux</i>		<i>Indices de parcimonie</i>		
GFI	AGFI	RMSEA	IFI	CFI	χ^2 /ddl	AIC	PNFI
> 0,9	> 0,9	≤ 0,08	> 0,9	> 0,9	< 5	le plus faible (comparaison)	le plus faible (comparaison)

4.2.2. La démarche d'analyse retenue

La méthode d'analyse confirmatoire que nous allons suivre est inspirée de la démarche présentée dans plusieurs travaux consacrés à la mesure de l'implication [Roussel & al. – 2002 ; Perrot – 2003 ; Belghiti & Briole – 2004 ; Manville – 2005] ; il s'agit d'une procédure générale comportant quatre étapes, qui seront réalisées avec l'aide des logiciels Lisrel version 8.8 [Jöreskog & Sörbom – 2006] et Amos version 7 [Arbuckle – 2006]. Nous avons préalablement procédé à l'examen des valeurs manquantes dans l'échantillon. Cette procédure est importante dans le cadre d'une analyse par méthodes d'équations structurelles, car l'estimation par les logiciels Amos ou Lisrel nécessite une distribution sans valeurs manquantes. Nous avons examiné la distribution des indicateurs des variables latentes afin de vérifier si les valeurs manquantes étaient distribuées de manière aléatoire. Cet examen nous a conduit à retirer deux observations dont les valeurs manquantes n'étaient pas distribuées normalement (il s'agissait de répondant n'ayant pas renseigné tous les items d'une échelle). L'échantillon a donc été réduit à 138 observations. Les valeurs manquantes restantes ont été remplacées en utilisant l'algorithme EM (*expected maximization*) disponible sur le logiciel Lisrel 8.

- La première étape consiste à examiner la multi-normalité des variables des échelles :

Nous examinerons dans un premier temps les indicateurs de déviation par rapport à la normalité (coefficients de dispersion et de concentration) pour chaque variable de l'échelle, puis le coefficient de concentration multivarié (coefficient de Mardia) de l'échelle complète¹⁴⁸.

Cet examen est indispensable afin de déterminer la méthode d'estimation des modèles de mesure qui sera retenue. La méthode la plus utilisée est celle du maximum de vraisemblance (*maximum*

¹⁴⁸ Selon Roussel & al [2002 p 47] le coefficient de symétrie (Skewness) ne doit pas dépasser |3| et le coefficient de concentration (Kurtosis) peut être admis jusqu'à |8|. Le coefficient de Mardia doit être inférieur à |3|.

likelihood), qui nécessite une distribution approximativement normale des variables. Si la multinormalité des variable ne peut pas être assurée, même de manière approximative, nous utiliserons une méthode d'estimation dite robuste disponible dans le logiciel Lisrel (robust maximum likelihood), spécialement adaptée aux distributions non normales [Vautier & al. – 2005 ; Mels – 2006].

- La seconde étape consiste dans l'élaboration de plusieurs modèles concurrents à tester.

Dans le cas de l'analyse factorielle confirmatoire, la structure des modèles est relativement simple : il s'agit de mettre en rapport une ou plusieurs variables latentes avec des indicateurs de mesure.

Lorsque le construit apparaît comme unidimensionnel (tous les énoncés contribuent sur un seul facteur latent), nous testerons la version initiale de l'échelle contre la version réduite issue de l'analyse factorielle exploratoire (ex : en ce qui concerne l'échelle de soutien organisationnel perçu, nous comparerons l'échelle complète à 6 items et l'échelle réduite à 5 items obtenue après analyse factorielle exploratoire). Lorsque le construit est multidimensionnel (ex : implication dans la profession) nous comparerons le modèle issu de l'étude exploratoire à sa version initiale, et à une version unidimensionnelle (dans laquelle tous les indicateurs sont reliés à un seul facteur latent). Cette procédure nous permettra de juger de l'intérêt de conserver des construits multidimensionnels.

- La troisième étape consiste à sélectionner parmi les modèles testés celui présentant les meilleurs indices d'ajustement, en fonction des seuils retenus. L'interprétation des différents indicateurs de qualité d'ajustement peut être délicate, surtout lorsque les modèles sont proches. Il est généralement recommandé dans ce cas de choisir le modèle le plus parcimonieux (en se fondant sur les indicateurs adéquats).

- La quatrième et dernière étape vise à s'assurer de la fiabilité, de la validité convergente et de sa validité discriminante du modèle retenu.

● **La fiabilité de l'échelle** retenue sera évaluée par un coefficient ρ^2 de Jöreskog (ou ρ^2 de Ksi). Ce coefficient mesure la fiabilité de cohérence interne d'un construit latent mesuré par une série d'indicateurs (selon la même logique que le coefficient Alpha de Cronbach). Il est recommandé lors des phases d'analyse factorielle confirmatoire [Gerbing & Anderson – 1988]

Le coefficient ρ^2 de Jöreskog (non standardisé) se calcule selon la formule suivante :

$$\rho_{\xi} = \frac{\left(\sum_{i=1}^p \lambda_i \right)^2 \text{var}(\xi)}{\left(\sum_{i=1}^p \lambda_i \right)^2 \text{var}(\xi) + \sum_{i=1}^p \text{var}(\delta_i)}$$

Lambda = contribution factorielle de l'indicateur *i*

K_{xi} = variance de l'échelle

Epsilon = erreur de mesure associée à l'indicateur *i* (variance d'erreur de l'indicateur)

Ses règles d'interprétation sont les mêmes que celles de l'alpha de Cronbach [Roussel & al. 2002] : une échelle peut être considérée comme suffisamment cohérente dans une phase confirmatoire à partir du moment où le coefficient rho est au moins supérieur à 0,70 (un seuil de 0,8 est préférable).

• En ce qui concerne la validité convergente de l'échelle retenue, celle-ci sera assurée si deux conditions sont remplies [Roussel & al. – 2002 p. 79] :

- Chaque item doit contribuer significativement sur un construit latent : on considère que c'est le cas à partir du moment où le test t associé à la contribution factorielle de l'indicateur est supérieur à 1,96 (pour un risque d'erreur de 5%).

- Chaque indicateur standardisé doit partager plus de variance avec son construit latent qu'avec son erreur de mesure¹⁴⁹. On considère que c'est le cas à partir du moment où la moyenne des variances entre le construit et ses mesures est supérieure à 0,5. La formule de calcul associée proposée par Fornell & Larcker [1981] correspond à la variance moyenne extraite (VME) ou « rho de validité convergente ». La formule de calcul est la suivante :

$$\rho_{vc(A)} = \frac{\sum_{i=1}^n (\lambda_i^2)}{\sum_{i=1}^n (\lambda_i^2) + \sum_{i=1}^n \text{var}(\epsilon_i)}$$

lambda = contribution factorielle standardisée de l'indicateur *i*

epsilon = erreur de mesure associée à l'indicateur *i*

• Lorsque le construit est multidimensionnel (c'est-à-dire qu'il fait apparaître plusieurs facteurs latents corrélés), il faut enfin procéder à un examen de la validité discriminante de chacun des facteurs latents. Cet examen peut être fait en comparant les variances partagées par les différents construits latents avec les variances partagées par les construits avec leurs indicateurs : il s'agit de

¹⁴⁹ Le pourcentage de variance d'un item restitué par le construit latent correspond au carré de sa contribution factorielle

vérifier qu'un construit latent partage plus de variance avec ses indicateurs qu'avec le ou les construits voisins [Roussel & al. – 2002 p 189]. On peut également utiliser un test de différence de chi-deux : le modèle multidimensionnel est comparé avec un modèle unidimensionnel (dans lequel tous les indicateurs sont reliés à un facteur unique) : si la différence d'ajustement entre les deux modèles (mesurée par un test de différence de chi-deux) est significative, et que le modèle multidimensionnel présente un meilleur ajustement, une preuve supplémentaire de validité discriminante est apportée.

Nous allons maintenant présenter les principaux résultats de l'analyse confirmatoire pour chacune des échelles retenues ; la procédure complète et détaillée de validation est disponible en annexe 9. Les échelles réduite à deux items (volition et autonomie perçue) n'ont pas fait l'objet d'analyse confirmatoire par les méthodes d'équation structurelles, car ces analyses confirmatoires ne sont en principe pas possibles sur les modèles comportant un seul facteur latent et moins de trois indicateurs, car ces modèles posent des problèmes d'identification¹⁵⁰. Nous avons simplement pu vérifier que la fiabilité de ces deux construits mesurés par le coefficient alpha de Cronbach demeurait bonne sur le deuxième échantillon. Nous obtenons des niveaux proches de ceux calculés sur le premier échantillon (0,68 pour la volition et 0,82 pour l'autonomie)

4.2.3. Les résultats obtenus

Pour chacune des échelles testées, nous allons présenter, le cas échéant, les indices d'ajustement de plusieurs modèles concurrents, ainsi que les coefficients rhô de fiabilité et de validité convergente (correspondant à la variance moyenne extraite : VME).

- Echelle de valeur intérimaire perçue

MODELE	GFI	AGFI	RMSEA	IFI	CFI
<i>Seuils d'acceptabilité</i>	$> 0,9$	$> 0,9$	$\leq 0,08$	$> 0,9$	$> 0,9$
Modèle théorique	0,99	0,97	0,01	1	1

Fiabilité de l'échelle : Coefficient rhô de Jöreskog = 0,73

Variance extraite (rhô de validité convergente de Fornell & Larker) = 0,49

Le modèle présente de très bons indices d'ajustement ; sa fiabilité et sa validité convergente sont acceptables. Nous pouvons donc retenir cette échelle pour la suite de notre étude.

¹⁵⁰ Le nombre de degrés de liberté de ce type de modèles est au mieux égal à zéro, ce qui ne permet pas l'estimation (pour une analyse détaillée, on peut se reporter à Kline (2005), p.174).

- Echelle de soutien organisationnel perçu :

Nous avons comparé deux modèles : le modèle initial à 6 énoncés avec le modèle réduit à 5 énoncés issu de notre étude exploratoire

Modèles	GFI	AGFI	RMSEA	IFI	CFI	Chi deux normé	AIC	PNFI
<i>Seuils d'acceptabilité</i>	> 0,9	> 0,9	≤ 0,08	> 0,9	> 0,9	< 5	<i>le plus faible</i>	<i>le plus faible</i>
SOP 6 items (mod. théorique)	0,98	0,95	0,01	1	1	1,01	34,15	0,52
SOP 5 items (modèle issu de l'étude exploratoire)	0,98	0,93	0,07	0,99	0,99	1,67	28,7	0,39

Les deux modèles étant très proches, nous retenons le modèle à 5 items, plus parcimonieux.

Fiabilité de l'échelle : Coefficient rhô de Jöreskog de l'échelle à 5 items = 0,80

Variance extraite de l'échelle réduite (rhô de validité convergente) = 0,46

La fiabilité est bonne et la validité convergente légèrement en deçà du seuil d'acceptabilité.

Nous décidons donc de conserver l'échelle dans sa version réduite.

- Echelle d'équité relative perçue

L'échelle initiale d'équité perçue ayant été conservée à l'issue de l'analyse factorielle exploratoire, nous testons simplement sa qualité d'ajustement en utilisant des indices absolus et relatifs :

MODELE	GFI	AGFI	RMSEA	IFI	CFI
<i>Seuils d'acceptabilité</i>	> 0,9	> 0,9	≤ 0,08	> 0,9	> 0,9
Modèle théorique	0,99	0,97	0	1	1

Fiabilité de l'échelle : Coefficient rhô de Jöreskog = 0,758

Variance extraite (rhô de validité convergente) = 0,44

Dans le cas de l'échelle d'équité relative perçue, la validité convergente est médiocre (mais demeure proche du seuil de 0,5). Etant donné que l'échelle d'équité perçue présente un bon ajustement aux données empiriques et une fiabilité suffisante, nous décidons donc de la retenir pour la suite de notre étude, malgré une validité convergente en deçà du seuil d'acceptabilité.

- Echelle d'implication organisationnelle

Nous avons confronté trois modèles différents : un modèle unidimensionnel, dans lequel les facteurs affectif, normatif et calculé ne sont pas distingués ; un modèle théorique comportant trois facteurs (affectif, normatif et calculé) représentant les trois dimensions de l'implication dans

la conceptualisation de Meyer & Allen, et enfin le modèle bidimensionnel issu de notre étude exploratoire :

Modèles	GFI	AGFI	RMSEA	IFI	CFI	Chi deux normé	AIC	PNFI
<i>Seuils d'acceptabilité</i>	> 0,9	> 0,9	≤ 0,08	> 0,9	> 0,9	< 5	<i>le plus faible</i>	<i>le plus faible</i>
Modèle théorique	0,86	0,80	0,093	0,95	0,95	1,7	224,9	0,72
Modèle unidimensionnel	0,72	0,61	0,17	0,87	0,87	4,95	431,74	0,69
Modèle issu de l'étude exploratoire	0,91	0,86	0,071	0,97	0,97	1,69	140,8	0,72

Le modèle issu de l'étude exploratoire apparaît clairement supérieur aux deux autres. Nous le conservons donc.

<p>Fiabilité de la dimension affective/normative Coefficient rhô de Jöreskog = 0,89</p> <p>Variance extraite pour la dimension affective : rhô de validité convergente de Fornell & Larker) = 0,50</p>	<p>Fiabilité de la dimension calculée Coefficient rhô de Jöreskog = 0,72</p> <p>Variance extraite pour la dimension calculée : rhô de validité convergente de Fornell & Larker = 0,40</p>
<p>Fiabilité de l'échelle complète = 0,91</p> <p>Variance moyenne extraite de l'échelle totale = 0,47</p>	

En ce qui concerne la **validité discriminante** du modèle testé, la corrélation entre les deux facteurs latents (0,34) est nettement inférieure à toutes les contributions factorielles des items des deux sous-échelles, et que le test de différence de chi-deux entre modèle unidimensionnel et modèle à deux facteurs est significatif (voir annexe 9). On peut donc affirmer que les facteurs sont suffisamment indépendants et que la validité discriminante de l'échelle d'implication organisationnelle est suffisante.

La validité convergente de la sous-échelle d'implication organisationnelle calculée est médiocre : ceci confirme les réserves précédemment faites sur la fiabilité globale de la dimension calculée de l'implication organisationnelle appliquée aux salariés intérimaires. Nous conservons cependant cette échelle pour la suite de notre travail étant donné que sa fiabilité reste à un niveau acceptable dans le cadre d'une étude exploratoire (la validité discriminante est établie et la fiabilité interne de la sous-échelle est supérieure à 0,7 en phase exploratoire et confirmatoire, ce qui la situe dans les normes d'acceptation en sciences sociales).

- Echelle d'implication dans la profession

Nous avons confronté le modèle issu de notre étude exploratoire à un modèle unidimensionnel, dans lequel tous les indicateurs sont corrélés à un facteur latent unique.

Modèles	GFI	AGFI	RMSEA	IFI	CFI	Chi deux normé	AIC	PNFI
Seuils d'acceptabilité	> 0,9	> 0,9	≤ 0,08	> 0,9	> 0,9	< 5	le plus faible	le plus faible
Modèle unidimensionnel	0,83	0,67	0,19	0,85	0,85	6,15	151,9	0,56
Modèle issu de l'étude exploratoire	0,96	0,92	0,051	0,99	0,99	1,36	60,5	0,62

Le modèle issu de l'étude exploratoire présente un bien meilleur ajustement. Ceci confirme le caractère multidimensionnel de l'implication dans la profession.

<p>Fiabilité de la dimension affective/normative Coefficient rho de Jöreskog = 0,85</p> <p>Variance extraite pour la dimension affective : rho de validité convergente de Fornell & Larker) = 0,59</p>	<p>Fiabilité de la dimension calculée Coefficient rho de Jöreskog = 0,77</p> <p>Variance extraite pour la dimension calculée : rho de validité convergente de Fornell & Larker = 0,46</p>
<p>Fiabilité de l'échelle complète = 0,90</p> <p>Variance moyenne extraite de l'échelle totale = 0,53</p>	

La fiabilité de la sous échelle d'implication affective est très satisfaisante (>0,8) et la fiabilité de la sous dimension calculée est acceptable (> 0,75).

La corrélation entre les deux sous-dimensions de l'implication dans la profession est notable (r = 0,43), ce qui montre que les deux sous-échelles mesurent des construits partageant des caractéristiques communes.

En ce qui concerne **la validité discriminante** du modèle testé, on peut constater que la corrélation entre les deux facteurs latents (0,43) est nettement inférieure à toutes les contributions factorielles des items des deux sous-échelles, et que le test de différence de chi-deux entre modèle unidimensionnel et modèle à deux facteurs est significatif (*annexe 9*). On peut donc affirmer que les facteurs sont suffisamment indépendants.

- Echelle d'implication dans les valeurs du travail

Nous avons confronté un modèle unidimensionnel avec le modèle issu de notre étude exploratoire.

Modèles	GFI	AGFI	RMSEA	IFI	CFI	Chi deux normé	AIC	PNFI
Seuils d'acceptabilité	> 0,9	> 0,9	≤ 0,08	> 0,9	> 0,9	< 5	le plus faible	le plus faible
Modèle unidimensionnel	0,95	0,91	0.069	0,96	0,96	1,67	51,37	0,61
Modèle issu de l'étude exploratoire	0,97	0,92	0.082	0,97	0,97	1,93	29,67	0,47

Bien que les deux modèles soient très proches le modèle issu de l'étude exploratoire présente une meilleure qualité d'ajustement, notamment au niveau de la parcimonie. Nous le conservons donc.

Fiabilité de l'échelle : Coefficient rhô de Jöreskog = 0,77

Variance extraite (rhô de validité convergente de Fornell & Larker) = 0,45

La fiabilité est bonne et la validité convergente légèrement en deçà du seuil d'acceptabilité.

Nous décidons donc de conserver l'échelle dans sa version réduite.

Le tableau suivant résume les résultats obtenus sur les six échelles testées :

Echelle	Nombre d'énoncés retenus	Fiabilité (rhô de Jöreskog)	Validité convergente (VME)
Valeur intérimaire	3	0,73	0,49
Soutien perçu	5	0,80	0,46
Equité relative perçue	4	0,76	0,44
Implication organisationnelle			
- échelle complète	12	0,91	0,47
- dimension affective	8	0,89	0,50
- dimension calculée	4	0,72	0,40
Implication dans la profession			
- échelle complète	8		
- dimension affective	4	0,85	0,59
- dimension calculée	4	0,77	0,46
Implication dans la valeur travail	5	0,77	0,45

Nous pouvons remarquer en conclusion que l'analyse confirmatoire nous a permis de valider toutes les structures identifiées sur le premier échantillon, lors de la phase exploratoire. Toutes les échelles retenues présentent de bons indicateurs de fiabilité, mais on constate que leur validité convergente est généralement légèrement inférieure au seuil de 0,5.