

**EFFICACITE RELATIVE DES GRILLES DE CORRECTION  
D'UN TEST DE JUGEMENT SITUATIONNEL POUR  
LA SÉLECTION DU PERSONNEL DE CENTRES D'APPELS**

**Thèse**

SABRINA POIRIER

**Doctorat en relations industrielles de l'Université Laval  
offert en extension à l'Université du Québec en Outaouais**

Doctorat (Ph.D.)

DEPARTEMENT DES RELATIONS INDUSTRIELLES  
UNIVERSITE DU QUEBEC EN OUTAOUAIS  
Gatineau, Canada

DEPARTEMENT DES RELATIONS INDUSTRIELLES  
UNIVERSITE LAVAL  
Québec, Canada



## SOMMAIRE

---

Dans le domaine des centres d'appels, l'affectation d'agents au service à la clientèle efficaces et satisfaits représente une activité cruciale pour toute direction dédiée aux ressources humaines de ces unités d'affaires. Or, pour parvenir à prendre de bonnes décisions d'embauche et de rejet, des instruments de mesure et d'évaluation des compétences sont souvent mis à contribution. À cet effet, les tests de jugement situationnel (TJS) sont de plus en plus utilisés pour la sélection du personnel. L'objet de cette thèse est donc de vérifier l'efficacité relative des différentes méthodes d'élaboration des grilles de correction des TJS. En effet, jusqu'à présent, trois méthodes ont été privilégiées par les concepteurs de tests (Weekley, Ployhart, & Holtz, 2006) : (1) l'approche rationnelle ou théorique (2) l'approche empirique et (3) le recours à des experts. La possibilité qu'une combinaison de ces différentes approches puisse conduire à de meilleurs résultats est également explorée.

Pour y parvenir, le résultat total obtenu à un TJS a été mis en lien avec une évaluation du rendement global effectuée par un supérieur immédiat en y appliquant huit grilles de correction différentes. Au total, un échantillon de 312 employés œuvrant dans des emplois de téléopérateurs au sein d'une grande institution financière québécoise ont participé à cette recherche. Dans l'ensemble, les résultats indiquent qu'une approche empirique permet généralement d'obtenir de meilleures statistiques descriptives en termes de distribution et de dispersion des scores pour un processus de sélection. Cependant, des analyses corrélationnelles indiquent qu'une approche multiple ou méthode hybride, basée sur une intégration des informations fournies par différentes sources d'informations (empirique, théorique et basé sur les experts) pour l'élaboration d'une grille de correction d'un TJS, permet de mieux prédire le rendement des employés ( $r=0,247^{**}$ ) comparativement à l'utilisation de la théorie pour définir les attentes ( $r=0,162^{**}$ ), le recours à des experts de contenu œuvrant dans le domaine de pratique ( $r=0,164^{**}$ ) ou l'utilisation d'une approche empirique ( $r=0,154^*$ ).

Mots clés : Test de jugement situationnel, grilles de correction, validité critériée, centre d'appels.

## TABLE DES MATIERES

---

Sommaire .....	iii
Table des matières .....	iv
Liste des tableaux .....	vi
Liste des figures .....	vii
Remerciements .....	ix
Introduction.....	1
PARTIE 1 : Contexte théorique .....	5
1. La sélection de personnel.....	8
1.1 Ses principales composantes.....	9
1.2 La dotation dans les centres d'appels .....	14
1.3 L'évaluation des différences individuelles dans les centres d'appels .....	19
2. Le test de jugement situationnel.....	21
2.1 Sa description et ses caractéristiques .....	23
2.2 Le développement d'un test de jugement situationnel.....	26
2.3 La grille de correction.....	30
2.3.1 Approche théorique .....	31
2.3.2 Approche des experts.....	32
2.3.3 Approche empirique.....	34
2.3.4 Approche hybride.....	36
2.3.5 L'attribution des pointages .....	37
2.4 Le concept de validité.....	41
2.4.1 Stratégie de validation basée sur le contenu .....	43
2.4.2 La validité de contenu du test de jugement situationnel.....	48
2.4.3 Stratégie de validation basée sur la relation avec d'autres variables .....	53
2.4.4 La validité critériée du test de jugement situationnel .....	55
2.4.5 Impact de la grille de correction sur la validité critériée.....	57
2.5 Le concept de fidélité .....	58
2.5.1 La fidélité du test de jugement situationnel .....	61
2.6 Ses principales limites .....	62
3. Le rendement en emploi .....	67

3.1	La définition du rendement.....	68
3.2	La mesure du critère.....	72
PARTIE 2 : Cadre d'analyse .....		76
4.	Conceptualisation de la recherche.....	77
4.1	Questions de recherche .....	77
4.2	Hypothèses de recherche.....	79
PARTIE 3 : Méthodologie.....		82
5.	Opérationnalisation de la recherche.....	82
5.1	Instruments de mesure pour le prédicteur .....	82
5.2	Élaboration des grilles de correction.....	87
5.2.1	Grille de correction théorique .....	88
5.2.2	Grille de correction par les experts de contenu .....	89
5.2.3	Grille de correction empirique .....	90
5.2.4	Grille de correction hybride .....	94
5.3	Instrument de mesure pour le critère.....	95
5.4	Déroulement de l'expérimentation.....	98
5.5	Échantillon .....	99
5.6	Méthodes d'analyse .....	100
PARTIE 4 : Résultats.....		103
6.1	Analyses préliminaires.....	103
6.1.1	Convergence entre les grilles de correction.....	103
6.1.2	Variabilités entre les grilles de correction.....	105
6.1.3	Analyses descriptives liées au critère.....	106
6.2	Vérifications des hypothèses.....	107
6.2.1	Analyses descriptives liées au prédicteur.....	107
6.2.2	Analyses corrélationnelles entre prédicteur et critère .....	110
6.3	Analyses statistiques supplémentaires .....	112
6.3.1	Comparaison inter-groupe .....	112
PARTIE 5 : Discussion.....		116
6.	Analyse critique .....	117
7.	Forces et limites de la recherche.....	126
8.	Pistes de recherches.....	130
Conclusion .....		134

Références bibliographiques.....	136
Annexe A : Dimensions évaluées par le T3C.....	153
Annexe B : Statistiques descriptives pour le T3C .....	154
Annexe C : Structure factorielle du T3C .....	155
Annexe D : Contre-validation de l'approche empirique.....	156
Annexe E :Critères d'évaluation du rendement supérieur.....	157
Annexe F : Histogrammes de distribution des résultats .....	158
Annexe G : Diagrammes quantile-quantile .....	166
Annexe H : Diagrammes de dispersion .....	167

## LISTE DES TABLEAUX

---

Tableau 1 : Illustration des caractéristiques « Quantité vs. Qualité » d'un centre d'appels.....	17
Tableau 2 : Exemples de questions illustrant les deux types de directives associées au TJS.....	24
Tableau 3 : Typologie du format de questions des TJS en fonction du type de directive.....	25
Tableau 4: Comparaisons entre le test de jugement situationnel et le centre d'évaluation.....	26
Tableau 5 : Corrélations obtenues entre différents facteurs et TJS en fonction du type de directive.....	49
Tableau 6 : Dimensions génériques du rendement.....	72
Tableau 7 : Exemple de résultats à un des items du test.....	92
Tableau 8 : Échelle d'évaluation du rendement.....	97
Tableau 9 : Matrice de corrélations entre les grilles de correction.....	104
Tableau 10 : Pourcentages de différences entre les « bonnes réponses ».....	106
Tableau 11 : Statistiques descriptives pour le critère.....	107
Tableau 12 : Répartition de l'échantillon en fonction de la mesure du critère.....	107
Tableau 13 : Statistiques descriptives pour chacune des grilles de correction.....	108
Tableau 14 : Corrélation entre les scores obtenus à chacune des grilles et le rendement.....	110
Tableau 15 : Intervalles de confiance des corrélations.....	112
Tableau 16 : Test de signification T en fonction du rendement.....	113
Tableau 17 : Différences de moyennes standardisées .....	114

## LISTE DES FIGURES

---

Figure 1 : Modèle liant l'acquisition des ressources humaines à la gestion stratégique.....	7
Figure 2 : Modèle illustrant l'ensemble des composantes d'un processus de sélection.....	10
Figure 3 : Différentes stratégies de validation en sélection de personnel.....	43
Figure 4 : Processus de définition du contenu d'un instrument de mesure à partir de l'emploi.....	46
Figure 5 : Modèle conceptuel des facteurs influençant la validité de construit des TJS.....	51
Figure 6 : Modèle conceptuel pour examiner l'intelligence pratique, émotionnelle et sociale.....	52
Figure 7 : Caractéristiques du critère.....	73
Figure 8 : Modèle conceptuel de la recherche.....	78
Figure 9 : Exemple d'une question du T3C.....	84
Figure 10 : Critères de référence pour la grille de correction empirique.....	91
Figure 11 : Schème expérimental.....	99



À mes deux filles,  
Amélie et Anne-Sophie.

## REMERCIEMENTS

---

Je me suis visualisée des dizaines de fois en train d'écrire les prochaines lignes. Je leur attribue presque autant d'importance qu'à l'ensemble de cette thèse ; un accomplissement qui n'aurait jamais dépassé le stade du projet sans le soutien et les encouragements d'une multitude de personnes.

Il va sans dire que mes premières pensées de reconnaissance vont d'abord à mon co-directeur, Monsieur André Durivage. J'essaie de lui dire régulièrement « merci » pour sa confiance et sa façon de me transmettre son savoir, mais j'ai enfin l'opportunité de lui témoigner toute ma gratitude. C'était déjà il y a 15 ans, la première fois où il m'a accordé la chance de contribuer à ses projets académiques et professionnels. L'événement marquant de cette belle aventure est certainement le fait qu'il m'ait encouragé à suivre un cours à l'UQO pour découvrir l'univers des RHs. Une expérience qui a changé l'ensemble de mon parcours professionnel et personnel et un match parfait qui m'anime encore. Tout au long de la réalisation de ce projet de doctorat, André a été mon exemple de détermination et d'optimisme, en plus de mettre en place toutes les conditions gagnantes qui m'ont permis d'atteindre cet objectif ambitieux. J'aimerais remercier également Monsieur Éric Gosselin, aussi co-directeur de cette thèse, qui a aussi eu un impact important dans ce processus. Par ses conseils judicieux, il a su guider mes efforts vers l'obtention de ce grade de Ph.D. J'ai toujours beaucoup apprécié sa disponibilité, son écoute et la qualité de nos échanges. J'en profite enfin pour remercier sincèrement les membres de mon comité de thèse ainsi que l'ensemble du personnel, des professeurs et des étudiants du département des relations industrielles de l'UQO que j'ai côtoyés au cours de ce parcours. Vous avez tous contribué au développement de mes compétences et je vous en suis reconnaissante.

Dans un autre registre, je ne peux pas passer sous silence l'ensemble de mes collègues de chez EPSI. Autant ceux qui ont cru en moi dès le début en m'offrant leur appui inconditionnel, Estelle, Sylvie, Franco, que ceux qui m'ont accompagné dans cette aventure à différents moments au cours des huit dernières années, Mario, Philippe, Mathieu, Nicolas, Stéphanie, Anjali, Jacynthe ainsi que tous les autres epsiens qui m'accompagnent au quotidien. Chacun d'entre vous peut s'attribuer le mérite de plusieurs pages de cette thèse. Fidèles à vous-mêmes, vous m'avez toujours fait la démonstration que lorsqu'on travaille en équipe, il est possible de réaliser de grandes choses et je vous en suis très reconnaissante.

Finalement, outre mes parents, les membres de ma famille et mes amis proches qui m'ont toujours encouragé dans ce projet, je veux surtout remercier mon mari Jacques qui m'a toujours donné le goût de me dépasser et qui n'a jamais douté de mes capacités à atteindre cet objectif. Comme dit le dicton : « Qui prend mari, prends pays ». À l'aube de nos dix-huit ans ensemble, je ne regretterai jamais une minute de m'être exilée en Outaouais pour construire cette belle vie avec lui et la signification de cette thèse ne serait rien si je ne pouvais pas partager ma fierté avec lui et nos deux charmantes filles, Amélie et Anne-Sophie.

## INTRODUCTION

---

Récemment, la réalité de notre quotidien nous a conduits à nous attendre à ce que l'ensemble des organisations, tant publiques que privées, et les fournisseurs de services avec lesquels nous transigeons soient en mesure de répondre à nos besoins et problèmes à une vitesse impensable il y a de cela à peine quelques années. Pour faire face à cette réalité, les organisations ont donc adopté les technologies de l'information et de la communication pour améliorer leur capacité de prestation de service en mettant en place au cœur de leurs opérations un centre d'appels (ou centre de contact client) dont les activités sont assurées par des agents de service à la clientèle dédiés au traitement des demandes des consommateurs, bénéficiaires ou prospects.

Dans le domaine de la gestion des ressources humaines de ces centres d'appels, le recrutement et la sélection de leurs représentants revêtent un caractère fondamental pour la pérennité et la compétitivité d'un établissement. En effet, comme il s'agit du premier et souvent du seul intervenant de l'organisation face aux clients, l'identification d'employés talentueux et capables de faire face à leurs problèmes quotidiens et futurs constitue une préoccupation majeure pour tout responsable de l'embauche d'un centre d'appels. Or, comme plusieurs de ces unités souffrent encore aujourd'hui d'un taux de roulement<sup>1</sup> important et étant donné que ces postes de « première ligne » représentent la plupart du temps une porte d'entrée afin de progresser dans l'organisation, il demeure crucial que ces agents soient choisis avec attention. À cet égard, certains auteurs, comme Cappelli (2003), vont même jusqu'à suggérer que l'adoption de programmes et de techniques permettant d'atteindre une meilleure adéquation entre le poste et la personne peut contribuer à diminuer ces mouvements de personnel à l'extérieur des frontières de l'organisation. Par conséquent, on observe aujourd'hui un accroissement de la demande pour l'élaboration de processus de sélection qui permettront l'affectation des personnes à des activités de service à la clientèle, de vente ou de recouvrement dans lesquelles elles seront efficaces et satisfaites.

---

<sup>1</sup>Taux de roulement : Pourcentage représentant le nombre d'employés qui quittent l'organisation pour une période donnée par rapport au nombre moyen d'employés de l'organisation au cours de la même période (Saba, T. et Dolan, S. (2013). *La gestion des ressources humaines : tendances, enjeux et pratiques actuelles*. Montréal: Pearson.).

Les procédures de sélection sont mises en œuvre de façon instrumentale afin d'attribuer la bonne ressource au bon endroit. Par contre, ces dernières aspirent également à être stratégiques en fonction des liens qu'elles visent à maintenir entre performance individuelle et performance organisationnelle. De cette manière, des pratiques de sélection adaptées et alignées peuvent contribuer au succès des organisations (Huselid, 1995).

Or, la dotation d'un poste comporte des risques tant pour l'organisation que pour les candidats admis dans la course. Ainsi, toutes méthodes de sélection de personnel permettant de limiter les risques tout en maximisant les bienfaits de leur utilisation devraient être privilégiées. Que ce soit pour respecter les contraintes administratives ou législatives, pour une représentation valide des responsabilités de l'individu au sein de l'organisation ou pour l'évaluation de la valeur scientifique d'un outil d'évaluation, la recherche dans le domaine vient en appui au responsable d'un processus de sélection afin qu'il puisse agir avec rigueur et éthique. À ce sujet, les écrits permettent notamment de recenser les meilleures pratiques et les caractéristiques d'une panoplie de méthodes d'évaluation utilisées lors de l'embauche. Ces instruments conduisent habituellement à choisir, parmi un ensemble de candidats, ceux qui ont le meilleur potentiel pour effectuer correctement le travail qu'on leur assigne. En effet, depuis plus de cent ans, chercheurs et praticiens se préoccupent de la capacité de ces méthodes d'évaluation à prédire le rendement futur en emploi. À travers les décennies, la recherche a d'ailleurs permis d'établir que des critères adéquatement développés avaient des liens évidents avec le rendement attestant ainsi de la validité prédictive de plusieurs mesures d'évaluation et de sélection de personnel (Schmidt et Hunter, 1998).

À cet effet, les tests de jugement situationnel (TJS) sont de plus en plus utilisés en sélection du personnel, notamment dans le cadre de la dotation de postes reliés au service à la clientèle (McDaniel, Morgeson, Finnegan et al., 2001). Dans sa définition la plus large, le TJS inclut tout test élaboré en vue de mesurer le jugement dans des contextes reliés au travail (McDaniel et al., 2001). Parmi les avantages qui leur ont valu une popularité croissante, on note le fait qu'ils sont étroitement reliés au contenu de l'emploi donnant lieu ainsi à une appréciation plus favorable de la part des candidats qui les perçoivent comme un moyen d'évaluation juste et équitable (Bauer, Maertz Jr, Dolen et al., 1998). Ils sont également moins sujets aux impacts négatifs ou « *adverse impacts* » (Clevenger, Pereira, Wiechmann et al., 2001; Motowidlo et Tippins, 1993; Nhung T. Nguyen, Biderman et McDaniel, 2005).

Cependant, comme l'objectif de la sélection du personnel est de prédire le plus justement possible le rendement et le comportement au travail d'un candidat une fois qu'il sera embauché, il est décevant de constater que les études tendent à démontrer que la validité critériée moyenne des TJS s'élève à 0,26 (McDaniel, Hartman, Whetzel et al., 2007). Ces résultats sont intéressants, mais nettement inférieurs aux coefficients de validité obtenus au moyen d'autres exercices d'évaluation, tels que l'entrevue structurée ou les échantillons de travail pour lesquels les coefficients s'élèvent respectivement à 0,51 et 0,50 (Schmidt et Hunter, 1998). À première vue, cette situation pourrait être attribuable en partie à l'approche adoptée par les concepteurs des TJS lors de l'élaboration de la clé de correction. En effet, la validité d'un TJS est peut-être en grande partie reliée à l'efficacité relative de sa grille de correction. Jusqu'à présent, trois méthodes ont été privilégiées par les concepteurs de tests (Weekley, Ployhart et Holtz, 2006) : (1) l'approche rationnelle ou théorique (2) l'approche empirique et (3) le recours à des experts. Dans le premier cas, les bonnes réponses sont déterminées à partir de la théorie ou du modèle qui a été utilisé au moment du développement des mises en situation et des choix de réponses. La seconde approche consiste à établir les bonnes réponses en fonction d'analyses statistiques, telles que la corrélation entre les choix de réponses et un critère externe (par exemple, l'expérience, le rendement, etc.). Finalement, dans la dernière approche, les bonnes réponses sont déterminées par un groupe d'experts en fonction de leur expérience et de leur connaissance de l'emploi.

L'objectif de ce projet de recherche est donc d'évaluer si l'une ou l'autre de ces trois approches permettant de déterminer la grille de correction d'un TJS conduirait à une meilleure efficacité relative en sélection de personnel, notamment par rapport à la distribution et la dispersion des scores des participants ainsi que sur le plan de la prédiction du rendement en emploi. La possibilité qu'une combinaison de ces trois approches permette d'obtenir de meilleurs résultats en comparaison avec l'utilisation isolée de chacune de ces approches sera également examinée.

D'abord, la revue de la documentation permettra de positionner le contexte théorique de la problématique par rapport au monde du travail et de la gestion des ressources humaines. Le thème de la sélection de personnel et ses principales composantes, notamment pour la dotation dans un contexte organisationnel de centres d'appels, seront également abordés. Par la suite, la discussion portera sur les tests de jugement situationnel et leurs principales caractéristiques psychométriques. Finalement, il sera nécessaire de présenter le concept que l'on tente de mesurer par cet outil, c'est-à-dire la notion du rendement en emploi.

Ce faisant, l'ensemble de la recension des écrits permettra d'exposer le cadre d'analyse, c'est-à-dire les questions de recherche ainsi que les hypothèses. La méthodologie utilisée afin d'obtenir les données nécessaires à la validation des hypothèses de recherche sera ensuite présentée incluant la nature des instruments administrés, la population cible ainsi que les méthodes d'analyse utilisées pour y parvenir. Enfin, la dernière partie portera sur les résultats obtenus ainsi que les pistes de discussion qui s'en dégagent.

## PARTIE 1 : CONTEXTE THEORIQUE

---

Qu'elle soit coopérative, communautaire, publique ou commerciale, toute organisation a besoin de gérer ses ressources humaines (RH) d'une façon efficace et doit s'intéresser à l'ensemble des individus qui la compose, qu'ils soient cadres ou exécutants (Boivin, 1987). De ce fait, lorsque l'on s'intéresse à la gestion des ressources humaines (GRH), on y découvre un ensemble de moyens, d'activités ou de programmes mis de l'avant pour acquérir, maintenir, développer et déployer de façon efficace les personnes exerçant ou susceptibles d'exercer un travail productif au sein d'une organisation donnée (Boivin, 1987). Or, l'ouverture des marchés étrangers, la rapidité des évolutions technologiques, les changements démographiques et la transformation des valeurs sociales portent la réflexion sur l'importance du potentiel humain qui compose une organisation (Burke et Ng, 2006; Pettersen, 2000). En fait, l'importance critique des RH pour le succès organisationnel est apparue de plus en plus manifeste dans le monde de la gestion (Guérin, Le Louarn et Wils, 1988). La GRH devient alors un ensemble d'activités qui visent la gestion des talents et des énergies des individus dans le but de contribuer à la réalisation de la mission, de la stratégie et des objectifs de l'organisation (Saba et Dolan, 2013).

Dans cette perspective, les milieux académiques et les professionnels de la GRH ont généralement adopté une vision portant sur la stratégie pour tenter d'améliorer les performances organisationnelles grâce au capital humain (M. L. Lengnick-Hall, Lengnick-Hall, Andrade et al., 2009; Wright, Dunford et Snell, 2001). Ainsi, par la gestion stratégique des ressources humaines qui compose le capital humain d'une organisation, on considère que les milieux qui adoptent une certaine stratégie d'affaires exigent des pratiques RH différentes que celles adoptées par des organisations privilégiant une stratégie d'affaires alternative (Delery et Doty, 1996). Par ailleurs, la littérature dans ce domaine a démontré que les pratiques de gestion des ressources humaines peuvent contribuer au rendement financier de l'organisation, et ce, particulièrement lorsqu'elles sont alignées avec la stratégie d'affaires de l'organisation (Huselid, 1995).

La gestion stratégique des ressources humaines devient alors l'atout essentiel du management par les ressources de l'organisation ou de l'approche « *Resource-Based View* » (Becker et Huselid, 2006; M. L. Lengnick-Hall et al., 2009; Wright et al., 2001). La théorie du management par les ressources décrit la façon dont les aptitudes, les compétences et les caractéristiques des employés créent un avantage concurrentiel durable et défendable dans le cadre d'une démarche proactive (Arrègle, 2006; Phillips,



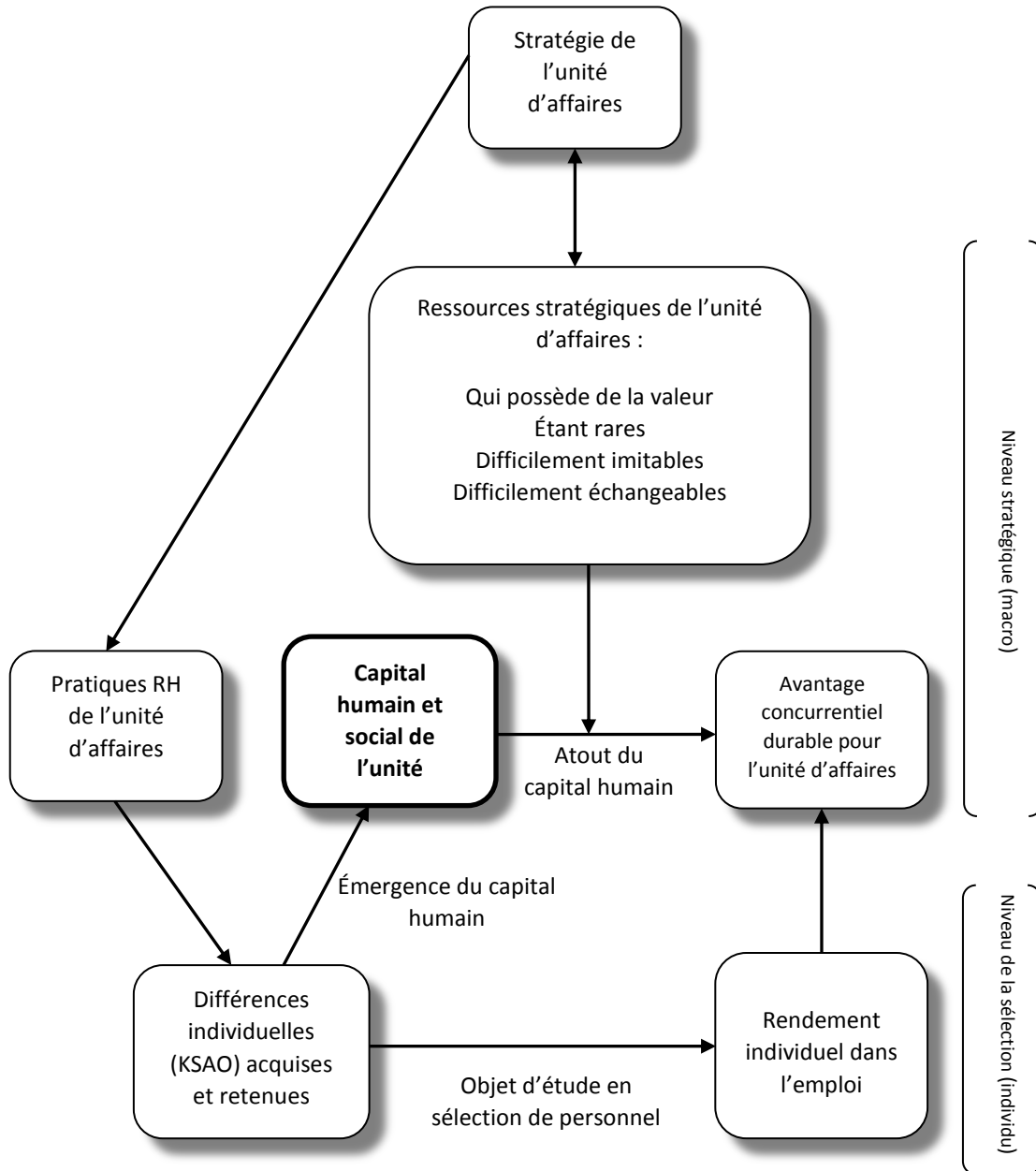
2012). Selon cette approche, des ressources seront des actifs stratégiques si elles consolident la position concurrentielle de l'entreprise en étant difficilement imitables, substituables et échangeables (Arrègle, 2006). En bref, la présence ou l'absence de ces actifs stratégiques expliquent les écarts de performance entre les organisations qui entrent en compétition à un moment précis sur un marché (Arrègle, 2006).

De cette manière, les auteurs vont mettre de l'avant le rôle des ressources humaines, non seulement comme source d'avantages concurrentiels, mais également comme atout compétitif sur lequel reposent tous les autres avantages (Guérin et al., 1988). L'avantage concurrentiel réfère aux capacités, aux ressources, aux relations et aux décisions qui permettent à une organisation de capitaliser sur les opportunités de l'environnement et de maîtriser les menaces à l'intérieur de son secteur d'activités (C. A. Lengnick-hall et Lengnick-hall, 1988).

Par ailleurs, la théorie du capital humain stipule que l'acquisition et la rétention du savoir spécifique à une organisation représentent un facteur déterminant pour son efficacité (Ployhart et Weekley, 2010). Ainsi, par capital humain, on fait alors référence à l'ensemble des connaissances, des capacités, des habiletés et autres caractéristiques (ou « KSAOs » pour *Knowledge, Skills, Abilities and Other Attributes*), tacites ou explicites, des personnes qui composent l'organisation (Crook, Todd, Combs et al., 2011; Ployhart et Weekley, 2010; Wright et al., 2001). En somme, les processus de la GRH sont appelés à créer de la valeur en mettant en place, notamment, des pratiques visant l'acquisition du capital humain chez les employés et leur affectation à un endroit où ils seront efficaces.

Cependant, comme le font remarquer Ployhart et Weekley (2010) dans leur modèle conceptuel liant l'acquisition des ressources humaines à la gestion stratégique d'une organisation, il ne suffit pas d'opérationnaliser une mesure de sélection et d'identifier les bonnes personnes au bon endroit pour participer au rendement organisationnel. En effet, l'ensemble des pratiques RH et les ressources humaines composant une organisation doivent créer un avantage durable au-delà du niveau individuel pour contribuer efficacement à la stratégie. L'illustration que ces auteurs en font permet de positionner les pratiques de sélection de personnel à l'intérieur du cadre plus large de la gestion stratégique des ressources humaines (voir figure 1 à la page suivante).

**Figure 1**  
**Modèle liant l'acquisition des ressources humaines à la gestion stratégique**  
 (Ployhart et Weekley, 2010, p. 200, traduction libre)



## 1. LA SELECTION DE PERSONNEL

---

Dans cette perspective, il est généralement admis que la sélection de personnes compétentes pour la réalisation de leurs tâches a un effet crucial sur la productivité et la prestation de services des entreprises privées et des organismes gouvernementaux (Hunter et Hunter, 1984; Schmidt et Hunter, 1998). En effet, si les employés ne possèdent pas les « KSAOs » leur permettant d'effectuer correctement leur emploi, tout programme axé sur les ressources humaines visant à améliorer la performance organisationnelle risque d'être voué à l'échec (Gatewood, Feild et Barrick, 2008).

Dans sa définition la plus large, la sélection de personnel représente une combinaison d'activités menant à une prise de décision et au choix de candidat(s) plutôt que d'autres pour un emploi ou un rôle au sein d'une organisation (Edenborough, 2007; Farr et Tippins, 2010). De cette manière, trouver la personne qui répond le mieux aux exigences d'un emploi représente l'objectif premier de chaque décision de sélection. Ainsi, tout en prenant en considération les enjeux légaux et sociaux ainsi que les intérêts à court et à long terme du candidat et de l'entreprise, le but d'un processus de sélection est de maximiser l'utilisation de la main-d'œuvre en affectant les bonnes personnes aux emplois au bon moment. Ces décisions peuvent inclure toute action ou tout mouvement de personnel relié au fait d'embaucher une nouvelle personne au sein de l'organisation, mais également celles de déplacer un employé à l'intérieur ou à l'extérieur de la structure organisationnelle de façon planifiée : mutation, promotion, redéploiement, accès à un programme de formation, utilisation de personnel temporaire ou « *decruting* » (Bechet, 2008; Catano, 2010; Farr et Tippins, 2010).

Pour parvenir à prendre une décision, les activités de sélection impliquent donc de recueillir de l'information sur ces personnes, de façon à évaluer leurs différences individuelles (Gatewood et al., 2008). À la base de la sélection de personnel se retrouve donc l'idée qu'une variation sur le plan des caractéristiques individuelles est reliée à une variation sur le plan du rendement dans l'emploi (Guion, 1965). Par conséquent, un processus de sélection impliquera : « *a process of measuring a person's knowledge, skills, abilities, and personal style to evaluate the characteristics and behavior that are relevant to (predictive of) successful job performance* » (Silzer et Jeanneret, 2011, p. 271). Comme Hogan (2012) le résume, cette cueillette d'informations est basée sur ces hypothèses fondamentales : (1) les êtres humains se distinguent par des traits ou des caractères importants ; (2) ces traits peuvent être quantifiés ; (3) ces traits sont assez stables ; et (4) les mesures de ces traits sont liées au comportement.

Selon une perspective théorique, la discipline de la sélection du personnel se présente donc comme : « une méthodologie scientifique basée sur la mesure de différences individuelles visant à choisir la bonne personne pour un poste en particulier, tout en mesurant empiriquement l'efficacité et la justesse des décisions » (Vinchur, 2007, p. 194, traduction libre). D'un point de vue historique, cette problématique « adéquation personne-poste » a acquis une grande importance au cours de la Première Guerre mondiale où les responsables de l'armée américaine ont sollicité l'implication des psychologues en vue de créer des tests permettant de déterminer rapidement le profil psychologique des millions de recrues de façon à les assigner un poste adéquat. Ainsi, depuis plus de cent ans, les chercheurs travaillent à améliorer les méthodes de développement et d'évaluation des instruments de sélection de personnel (Farr et Tippins, 2010). En effet, la mesure et l'évaluation des caractéristiques individuelles doivent être réalisées avec des procédures valides, fidèles et défendables, c'est-à-dire supportées par des évidences empiriques recueillies à partir de méthodes scientifiques reconnues (Catano, 2010; Pettersen, 2000).

---

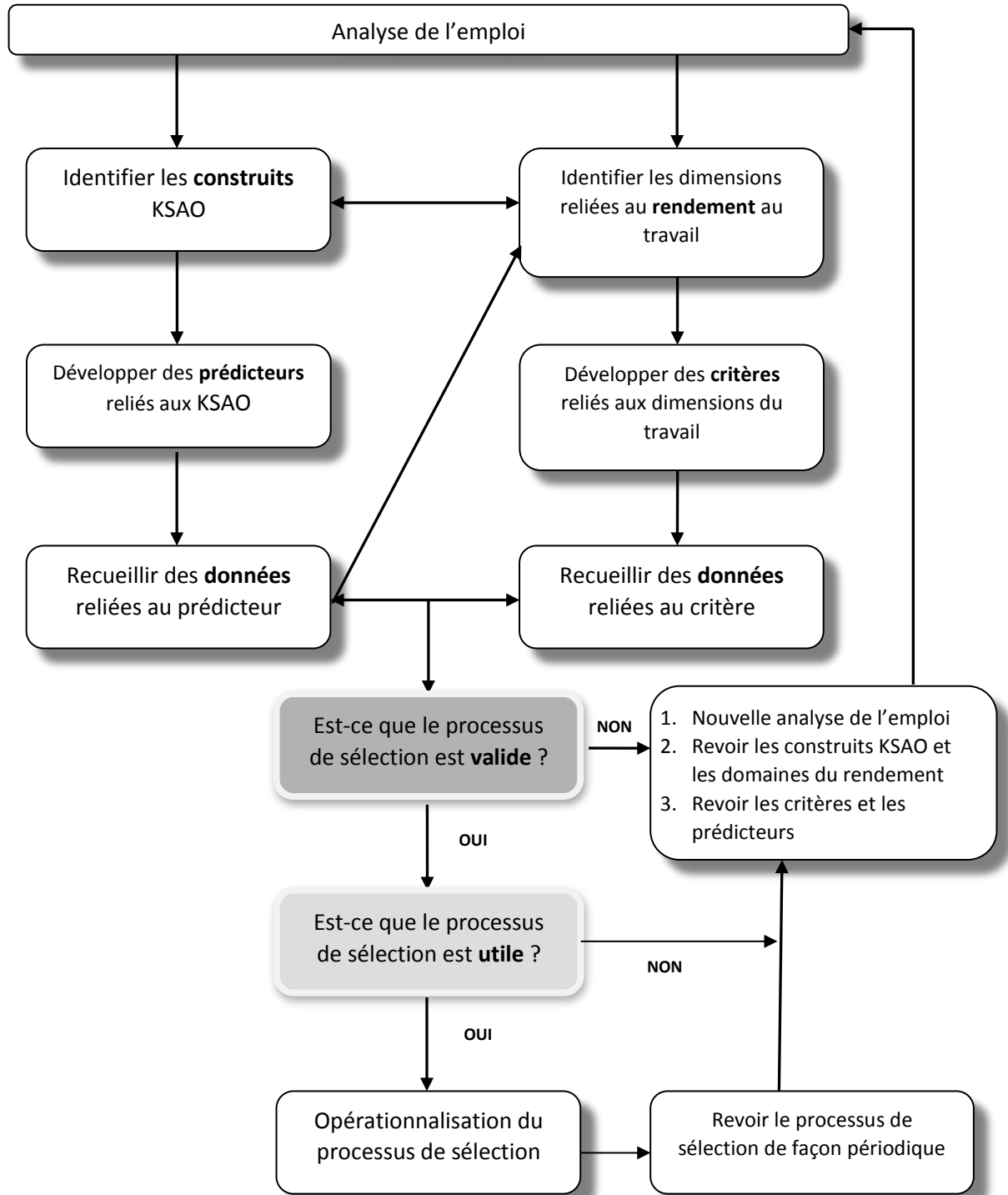
## 1.1 SES PRINCIPALES COMPOSANTES

---

Dans le monde organisationnel, la gestion des activités de sélection consiste donc à mettre en place des stratégies et des mécanismes qui les aideront à prédire le succès au travail, et ce, avec le plus de précision possible. Comme les décisions de sélection peuvent s'avérer coûteuses et entraîner des conséquences émotionnelles néfastes pour les candidats, il est primordial de prendre les moyens qui diminueront les risques qui y sont reliés. Pour y parvenir, le responsable du processus de sélection doit donc recueillir de l'information exacte et pertinente sur les personnes ayant posé leur candidature ainsi que sur les exigences de l'emploi et les normes de rendement qui y sont associés, et ce, tout en respectant les contraintes administratives, législatives et les règles d'éthique liées à la profession (Pettersen, 2000).

Lorsque l'on s'intéresse à l'ensemble des composantes d'un processus de sélection qui relie l'aspect fonctionnaliste avec le caractère scientifique de la discipline, il est possible de se référer au modèle illustré par Catano (2005). Ce modèle est intéressant puisqu'il met l'accent sur deux questions d'importance primordiale : « est-ce que le processus de sélection est valide ? » et « est-ce que le processus de sélection est utile ? ».

**Figure 2**  
**Modèle illustrant l'ensemble des composantes d'un processus de sélection**  
 (Catano, 2010, p. 25, traduction libre)



D'abord, comme l'objectif de tout processus de sélection est de maximiser l'adéquation entre les individus disponibles et un poste à combler, un portrait clair et détaillé de cet emploi représente la première étape logique dans le développement d'un processus de sélection (Gatewood et al., 2008). Notons d'ailleurs que le terme « analyse de l'emploi » sera généralement privilégié à celui d' « analyse de poste ». En effet, les pratiques en sélection de personnel abordent la plupart du temps l'évaluation des caractéristiques propres à un regroupement de postes plutôt qu'à un seul (Catano, 2010).

Comme il le sera exposé plus loin, cette première étape d'analyse est primordiale autant pour l'identification des domaines de rendement devant faire l'objet d'une mesure que pour le choix des instruments d'évaluation. D'ailleurs, elle se doit d'être conduite avec rigueur et éthique étant donné que la discrimination débute souvent avec l'analyse de l'emploi et les biais qui y sont parfois associés (Cook, 2009). Finalement, cette collecte d'informations sur la nature de l'emploi servira également à établir les bases de comparaison entre le rendement attendu et le rendement démontré lors de la validation de la procédure de sélection.

À la suite de l'analyse de l'emploi, il s'agit donc pour le responsable des ressources humaines de choisir les méthodes et les instruments de mesure qui seront appliqués et à partir desquels les mesures du rendement futur seront prises. Or, le concept d'utilité permet de mesurer l'importance de la décision prise lors de la sélection et il permet généralement d'orienter le processus d'embauche et le choix à l'égard des instruments d'évaluation. Comme le rapporte Sekiou (2001), l'utilité est une quantité théorique qui traduit en une seule valeur la différence entre les coûts et la qualité des décisions de sélection. Bien qu'il ne soit pas aisé de calculer le retour sur l'investissement d'une méthode de sélection, grâce à différentes approches comme celle sur le calcul du gain économique de Brogden-Cronbach-Gleser<sup>2</sup> et ses méthodes pour estimer la valeur monétaire des différences de rendement entre les employés, il est possible d'évaluer l'amélioration produite par l'application d'un instrument plutôt qu'un autre en termes monétaires. Par ailleurs, la configuration d'un processus prendra en compte les coûts associés à une erreur de sélection (attribuer le poste à une personne non suffisamment qualifiée pour le faire ou rejeter une candidature qui aurait eu un rendement supérieur dans l'emploi) afin de déterminer l'utilité d'une ou plusieurs méthodes d'évaluation. Par exemple, lorsque l'on prend en considération les coûts associés à la conduite d'un processus de recrutement (affichage, analyse des

---

<sup>2</sup> Cette approche est connue sous le nom de ses trois auteurs : Brogden (1949), Cronbach et Gleser (1965). Source : Pettersen (2000).

candidatures reçues, entrevues de pré-sélection, etc.), les heures investies dans l'accueil et la formation d'un nouvel employé ainsi que le salaire versé à un individu qui ne rencontre pas les exigences de productivité de l'organisation, il est plausible de croire que plus les niveaux hiérarchiques et les salaires des postes à doter seront élevés, plus les processus de sélection seront configurés d'une façon à en augmenter la validité et à en minimiser le risque de mauvaise décision. En somme, des considérations pratiques et financières viendront souvent influencer les activités liées à l'évaluation des candidatures. À cet égard, l'efficacité d'un processus visant l'embauche massive d'employés de première ligne sera souvent prise en compte de manière prioritaire avant son implantation, c'est-à-dire son caractère simple, rapide et peu coûteux.

Dans le même sens, l'apport de la technologie dans le monde du travail rend l'utilisation du Web de plus en plus répandue non seulement pour la gestion des procédures et des candidatures au sein des organisations, mais également pour l'administration des instruments d'évaluation via des applications de testing en ligne. En effet, suite à un sondage effectué auprès de 125 grandes organisations nord-américaines, Chapman et Webster (2003) ont observé que l'utilisation des nouvelles technologies en évaluation du personnel était motivée par l'amélioration de l'efficacité, l'introduction de nouveaux outils de sélection, la réduction des coûts et l'amélioration de la standardisation des processus. Ajoutons à ces objectifs poursuivis, d'autres avantages de l'administration d'instruments d'évaluation en ligne, telles que la possibilité de modifier rapidement le contenu des moyens d'évaluation, la possibilité d'ajuster rapidement les normes ou les notes de passage des tests utilisés, la possibilité d'administrer les tests 24 heures par jour, 7 jours par semaine, la création de bases de données permettant la conservation et la centralisation des résultats ainsi que la production d'analyses statistiques, etc. (Durivage et Pettersen, 2007; Tippins, Beaty, Drasgow et al., 2006). Ainsi, Internet permet aujourd'hui de standardiser facilement les processus, d'éliminer la lourdeur administrative associée au papier-crayon ainsi que certaines sources d'erreurs humaines, et ce, tout en augmentant la rapidité des décisions.

Or, ces considérations opérationnelles ne doivent pas avoir préséance sur la réalité vécue par le participant au processus de sélection. Certains auteurs vont même jusqu'à parler de la « validité sociale » d'un processus de sélection en insistant sur l'importance d'utiliser des procédures d'évaluation transparentes qui permettent la rétroaction aux candidats (Truxillo et Bauer, 2011). En effet, du point de vue du candidat, une participation à un processus de sélection est une occasion unique d'accéder à un poste dans lequel il sera satisfait et où il pourra se réaliser pleinement (Farr et Tippins, 2010).

C'est pourquoi, on insiste de plus en plus sur l'importance d'offrir aux candidats une description réaliste de l'emploi ainsi que des impacts de cette dernière sur l'efficacité des processus de recrutement (St-Onge, 2013). Le mot « réaliste » fait référence au fait que les meilleures méthodes d'embauche vont mettre l'accent sur ce à quoi le candidat doit exactement s'attendre une fois embauchée afin d'éliminer la distance entre l'image idéale du travail et l'image réelle (Catano, 2010). À cet effet, les processus d'embauche qui fournissent davantage d'informations sur le poste, et cela sans déformation, sont donc en principe plus efficaces puisqu'ils contribuent à améliorer l'adéquation individu-emploi et vont également être associées à différentes variables, telles que la satisfaction en emploi et la rétention dans le poste (Breagh, 1981).

Par ailleurs, un candidat sera également influencé par la perception d'équité dans le traitement de tous les autres candidats. D'abord, l'équité dans la procédure est primordiale : mêmes tests, mêmes règles, mêmes standards, mêmes normes, etc. (Pettersen, 2000). Par ailleurs, en matière d'égalité, l'encadrement législatif concernant les droits de la personne au Canada précise plusieurs caractéristiques protégées par des lois antidiscriminatoires (Demers et Messing, 2001). À cet égard, la *Charte des droits et libertés de la personne*<sup>3</sup> encadre les relations entre un employeur et ses employés, notamment lors d'un processus d'embauche, sur de multiples sujets reliés au respect de la vie privée (article 5), aux facteurs de discrimination (article 10), à la non-discrimination à l'embauche (article 16) ou aux renseignements recueillis en vue d'un emploi (article 18). La Commission des droits de la personne et des droits de la jeunesse se voit d'ailleurs fréquemment interpellée par les salariés vis-à-vis des pratiques de sélection choisies par les organisations. Qui plus est, y est prescrit dans la jurisprudence à cet effet, l'obligation des organisations de justifier non seulement l'usage des méthodes de sélection, mais également la relation entre le contenu de ces méthodes et les emplois visés (Pettersen, 2000). Devant les tribunaux, une organisation ne peut pas simplement invoquer son droit de gérance, elle doit également faire la démonstration du bien-fondé de l'évaluation en regard au poste à combler (Demers et Messing, 2001). Par ailleurs, lorsque les processus d'embauche visent des candidatures provenant de l'intérieur de l'organisation, les conventions collectives viennent également encadrer les processus de sélection ainsi que l'utilisation d'outils d'évaluation.

---

<sup>3</sup> L.R.Q. c. C-12.



En somme, malgré que la décision d'embaucher la personne de son choix appartienne à un employeur, un processus de sélection et les instruments d'évaluation qui y sont associés doit nécessairement respecter le droit à l'égalité ainsi que les dispositions législatives en vigueur. Conséquemment, pour les professionnels de la fonction ressources humaines, leur mission est de trouver les pratiques et les tests les plus appropriés afin d'assurer l'équité et la justice entre les personnes visant à œuvrer dans l'organisation. Pour les représentants des travailleurs, il s'agira de s'assurer que les organisations privilégient l'utilisation d'un mode de prise de décision rigoureux ainsi que des méthodes d'évaluation qui s'éloignent des pratiques arbitraires en les dénonçant, le cas échéant, selon les dispositifs de règlement des litiges prévus (Pettersen, 2000).

Dans tous les cas, un processus de sélection et des instruments d'évaluation qui n'accordent pas d'attention à l'ensemble de ces enjeux n'a pas sa place en gestion des ressources humaines. Non seulement celui-ci ne peut servir à guider les décisions organisationnelles, mais il peut également être nuisible et entraîner des conséquences néfastes, telles que des choix de candidats inadéquats, une évaluation inéquitable, des difficultés à se défendre en cas de grief ou de poursuite devant les tribunaux, sans parler du manque d'éthique à l'égard des personnes injustement traitées (Pettersen, 2000).

---

## 1.2 LA DOTATION DANS LES CENTRES D'APPELS

---

Les centres d'appels sont un phénomène plutôt récent. Même s'ils sont apparus à partir des années 1960 aux États-Unis, ce n'est que depuis le milieu des années 1990 que ce secteur d'activité économique a connu un essor considérable à l'échelle mondiale (Béraud, Colin et Grasser, 2008). Au Canada, par exemple, l'emploi dans ce secteur est passé de 20 000 à 112 000 postes entre 1987 et 2004 (StatistiqueCanada, 2005) cité dans (Marcoux, 2010) et les revenus des organisations y évoluant spécifiquement sont passés de 424 millions à 2,76 milliards de dollars entre 1998 et 2006 (Vincent et McKeown, 2008). Aux États-Unis, Lucas (2001) estimait qu'il y avait plus de 3 millions de personnes qui travaillaient dans une telle organisation au début des années 2000. Qui plus est, ces données sur la croissance de l'emploi et des revenus de ce secteur sont nettement sous-estimées lorsque l'on s'intéresse aux problématiques RH puisque leur compilation exclut l'ensemble des centres d'appels intégrés directement à l'intérieur d'une organisation (Van Jaarsveld, Frost et Walker, 2007) cité dans (Marcoux, 2010). En effet, les travailleurs de centres d'appels sont présents dans tous les secteurs de

l'économie canadienne, notamment dans le secteur des banques, des assurances, des télécommunications, de la vente, des loisirs, de l'assistance et de la sécurité, mais également au sein de l'administration publique. De nos jours, la plupart de grandes organisations utilisent des centres d'appels, qu'ils soient situés localement ou ailleurs dans le monde afin d'offrir du service à leur clientèle et de vendre leurs produits.

Selon Dawson (2004), on peut définir un centre d'appels comme « *a physical location where calls are placed, or received, in high volume for the purpose of sales, marketing, customer service, telemarketing, technical support or other specialized business activity* », notamment, le service après-vente, les réclamations, le recouvrement de créances, etc. À cet égard, il est utile de présenter les caractéristiques qui peuvent distinguer les centres d'appels entre eux. Selon le Système de classification des industries de l'Amérique du Nord<sup>4</sup>, il y a des établissements dont l'activité principale consiste à recevoir des appels, alors que d'autres ont pour mission de faire des appels téléphoniques. Plus spécifiquement, les travailleurs d'un centre d'appels « *handle incoming (inbound) customer calls for such things as assistance, order processing, and administrative assistance (e.g., complaints or billing questions) or place calls (outbound) to customers in order to obtain information or sell products and services (e.g., customer satisfaction surveys, marketing, sales, and customer follow-up)* (Lucas, 2001). Par ailleurs, dans ces deux cas, le centre d'appels peut être internalisé à l'organisation ou encore externalisé, c'est-à-dire que les ressources technologiques et humaines de cet établissement travaillent pour un tiers (Zapf, Isic, Bechtoldt et al., 2010).

Comme le fait remarquer Buscatto (2002, p. 100) à propos des centres d'appels : « Fondés sur une alliance du téléphone et de l'informatique, ils incarnent une modernité technologique accessible à tous en tout point du territoire ». En fait, la réalité virtuelle permet à tout consommateur ou bénéficiaire d'un service d'obtenir réponse à ses questions 24 heures par jour et 7 jours par semaine, et ce, sans se déplacer. À l'ère où l'on considère le client comme l'un des biens les plus précieux d'une organisation, les unités opérationnelles et structures professionnelles qui se chargent de faire le lien entre eux ont donc acquis une importance notable. D'ailleurs, cette réalité n'est pas étrangère au fait que plusieurs sociétés utilisent maintenant la dénomination « Centre Contact Client » et « Centre d'appels » de façon

---

<sup>4</sup> Cité dans Vincent, R. et McKeown, L. (2008). Tendances dans l'industrie des centres d'appels. Ottawa : Statistiques Canada.

interchangeable afin de prendre en compte la dimension qualitative de la relation client qui est implicite dans les activités de ces unités d'affaires.

Bien que l'activité de relation clientèle par téléphone nous est souvent présentée comme une activité simple, répétitive et organisée selon un mode proche du modèle taylorien (travail déqualifié, très contrôlé, effectué dans des conditions de travail difficiles), la réalité des centres d'appels est évidemment beaucoup plus nuancée étant donné que la nature de leurs opérations varie considérablement d'un établissement à un autre (durée des appels, types de clients, etc.) (Marcoux, 2010; Van Jaarsveld et al., 2007). En fait, dans plusieurs centres d'appels, il existe non seulement une certaine forme de complexité sur le plan du contenu du travail, mais également par rapport à l'organisation du travail elle-même :

« L'observation d'un certain nombre de facteurs tels que la variété des tâches réalisées, l'absence de scripts, la mobilisation d'applications informatiques importantes, la durée des appels, les objectifs quantitatifs/qualitatifs imposés, les formes de contrôle utilisées et les possibilités d'échanges au niveau du collectif de travail, mais aussi la formation initiale et le temps nécessaire pour un opérateur de devenir pleinement « opérationnel » montre que l'organisation du travail s'inscrit plutôt dans un cadre de « professionnalisation » de l'activité » (Béraud et al., 2008, p. 5).

Dans le même sens, le caractère multiple et évolutif des procédures en place dans un centre d'appels, par exemple concernant les produits offerts, la réglementation en vigueur, les technologies utilisées, etc., fait que, dans la plupart des cas, il s'agit pour l'agent du centre d'appels de s'adapter en situation au traitement de contextes différents plutôt que de mettre en œuvre indistinctement des modes opératoires applicables à un nombre réduit de situations (Cihuelo, 2010; Van Jaarsveld et al., 2007). Dans le même sens, ces derniers doivent constamment mettre leurs connaissances et leurs capacités à jour afin d'offrir aux clients un service de qualité qui respecte les besoins et attentes de ces derniers (Van Jaarsveld et al., 2007).

Finalement, les stratégies de gestion des ressources humaines mises en place dans un centre d'appels doivent prendre en considération que l'organisation du travail peut s'articuler autour de deux principes contradictoires, soit, d'une part, sur une production axée sur le volume et les coûts, et d'autre part, par rapport à une personnalisation du service (Marcoux, 2010). De cette manière, il est utile pour le spécialiste des ressources humaines d'établir l'importance relative de chacun de ces principes et

d'établir la nature distinctive des opérations du centre d'appels afin de s'y aligner. Pour y parvenir, Taylor et ses collègues (2002) suggèrent que la diversité des opérations d'un centre d'appels peut être appréhendée en se référant à la dichotomie « Quantité vs. Qualité » pour laquelle ils ont identifié différentes caractéristiques se situant aux extrémités de chaque pôle (voir tableau 1 ci-dessous).

**Tableau 1**  
**Illustration des caractéristiques « Quantité vs. Qualité » d'un centre d'appels**  
**(Taylor et al., 2002, p. 136)**

Quantité	Qualité
Interactions avec le client de nature simple	Interactions avec le client de nature complexe
Activité routinière	Individualisation du service
Objectifs difficiles à atteindre	Objectifs flexibles
Script de réponse à suivre de façon stricte	Réponse flexible – pas de script
Temps consacré à l'appel très serré	Souplesse concernant le temps consacré à l'appel
« <i>Wrap-up times</i> » très serrés	Priorité accordée à la satisfaction du client
Fort pourcentage du temps de travail au téléphone	Possibilité d'accomplir d'autres tâches
Géré en fonction de statistiques	Statistiques modifiées par des critères de qualité
Cycle de la tâche très court	Long cycle de la tâche
Haut volume d'appels	Faible volume d'appels
Faible valeur de chaque appel	Forte valeur de chaque appel
Peu d'autonomie laissée à l'opérateur	Haut niveau d'autonomie de la part de l'opérateur
La nature des appels est simple	La nature des appels est complexe
Service de livraison de masse	Personnalisation

Qu'ils soient désignés comme téléopérateurs, télévendeurs, préposés, agents ou représentants au service à la clientèle, c'est souvent à partir de leur capacité à satisfaire les clients que sera jugée une organisation. Ainsi, attribuer un poste au service à la clientèle à un nouveau candidat n'est pas une activité banale et dénuée de coûts. Des sommes importantes sont donc investies annuellement par un centre d'appels non seulement, pour attirer des individus talentueux, mais également pour sélectionner les candidats démontrant le meilleur potentiel en fonction des objectifs poursuivis. Selon une étude de Van Jaarsveld et ses collègues réalisée en 2007 concernant l'industrie canadienne des centres d'appels, plus de 45,5% de ces organisations vont utiliser des tests psychométriques ou d'aptitudes lors d'un processus de recrutement. Par ailleurs, les centres d'appels ayant été sondé dans cette étude ont indiqué dépenser en moyenne 3 495,00\$ pour recruter, pré-sélectionner, choisir et former un nouvel employé. Lorsque l'on pense aux mouvements de personnel qui sont habituellement très élevés dans les centres d'appels (moins de 6 ans de stabilité en emploi en moyenne et près de 30% des agents qui ont moins d'un an d'ancienneté tel que rapporté dans l'étude de Van Jaarsveld et ses collègues réalisée

en 2007), on réalise rapidement que l'impact financier du recrutement est très important pour ces organisations. En bref, les frais de dotation représentent plus de la moitié de l'ensemble des coûts de fonctionnement d'un centre d'appels (Duder et Rosenwein, 2001).

Dans le même sens, en raison des exigences d'embauche qui sont relativement peu élevées en termes d'éducation et d'expérience de travail, les centres d'appels demeurent accessibles à un bassin important de chercheurs d'emploi. On s'attendrait alors à ce que le taux de sélection (c'est-à-dire le pourcentage de personnes embauchées lorsque comparé avec le bassin total de personnes ayant posé leur candidature) soit très élevé. Or, l'efficacité des titulaires de postes à adopter chacun de ces rôles au bon moment et de la bonne manière est très variable d'une personne à l'autre. Selon l'étude de Van Jaarsveld et ses collègues réalisée en 2007, le taux de sélection est d'à peine 30% au Canada. À cet égard :

« Au-delà du niveau du diplôme, les compétences requises sont la capacité d'apprentissage, de communication et d'élocution, la capacité à prendre en compte et à comprendre les données des clients, la maîtrise de l'écrit (orthographe, raisonnement et expression) et de plus en plus la capacité commerciale (en fort développement dans la plupart de ces centres). Représentant le plus souvent la première forme de contact avec l'entreprise («la vitrine de l'entreprise»), les responsables des centres de relations avec les clients (CRC) sont sensibles au «savoir-être» des conseillers clientèle et veillent à ce que ces derniers n'aient pas de particularités détectables à l'écoute » (Béraud et al., 2008, p. 5).

Comprendre les problèmes des clients, prendre les bonnes décisions qui s'imposent et mettre en œuvre les activités administratives qui y sont reliées sont en fait tributaires d'une variété de caractéristiques présentes ou non chez les individus d'où l'importance d'identifier et d'évaluer correctement les caractéristiques qui feront la différence dans le rendement futur en emploi. En fait, si la littérature en matière de service à la clientèle a clairement établi que la capacité d'un individu à s'adapter dans le cadre d'une activité client est reliée positivement au rendement individuel et organisationnel, ce n'est que récemment que l'on s'est intéressé à ses principaux déterminants (Gwinner, Bitner, Brown et al., 2005).

Dans cette perspective, Gwinner et ses collègues (2005) se sont demandés : 1) Comment peut-on définir les comportements d'adaptation des employés ? et 2) Quelles sont les différences individuelles qui nous permettent de distinguer les employés qui s'adaptent plus facilement de ceux

pour qui cela est plus difficile ? (Gwinner et al., 2005). Pour ce faire, ils ont défini la capacité d'adaptation dans le cadre d'une activité clientèle comme étant la combinaison de deux composantes, soit la composante interpersonnelle et la composante reliée à l'offre du service. Plus spécifiquement, les résultats de leur étude ont démontré des liens significatifs entre certaines caractéristiques personnelles comme la sensibilité face aux autres, l'orientation vers le service, la motivation intrinsèque et la tolérance face à l'ambiguïté en lien avec la capacité d'une personne à s'adapter dans le cadre d'une relation clientèle. Il s'agit donc de ces principales caractéristiques qui devront être ciblées dans le cadre d'un processus de sélection visant à contribuer stratégiquement aux opérations d'un centre d'appels.

---

### 1.3 L'ÉVALUATION DES DIFFÉRENCES INDIVIDUELLES DANS LES CENTRES D'APPELS

---

Ainsi, malgré que l'environnement organisationnel et divers éléments du processus entre en jeu, les différences individuelles demeurent une variable incontournable dans l'évaluation de l'efficacité des interactions d'une personne avec les situations rencontrées avec les clients. En fait, plus cette dernière se situe en maîtrise de la composante interpersonnelle et la composante reliée à l'offre du service qui lui est assignée, plus les résultats obtenus et le degré de réussite de son intervention dans le cadre d'une relation client sera élevé. D'un point de vue théorique, la notion de différences individuelles est habituellement associée à la description des préférences, des capacités et des dispositions de l'individu qui influence en retour la force et la fréquence des comportements d'une personne (Motowild, Borman et Schmit, 1997). De cette manière, lorsque l'on s'intéresse à la sélection du personnel dans un centre d'appels, le fondement essentiel de la discipline demeure, c'est-à-dire la relation démontrée de façon valide et utile entre les différences individuelles et les exigences de l'emploi. Ainsi, les activités de sélection des agents au service à la clientèle doivent prévoir des méthodes permettant de recueillir le maximum d'informations sur le plan de ces différences individuelles dans le but de prédire leurs comportements futurs au travail et d'améliorer l'efficacité organisationnelle.

Pour y parvenir, plusieurs outils d'évaluation sont à la disposition des organisations. Au-delà de l'entrevue de sélection qui est mise à contribution dans la majorité des processus de sélection des centres d'appels, les tests qui consistent à mettre le candidat dans une situation très similaires à celle de l'emploi demeurent une pratique très répandue dans les organisations depuis plusieurs décennies (Weekley et Jones, 1999). Ces tests, que l'on appelle mise en situation, jeu de rôle, exercice de

simulation ou échantillon de travail, sont des outils recherchés puisque leur objectif est précisément de placer le candidat dans une situation de travail et d'observer ses comportements (Bourhis, 2007). Ils se caractérisent habituellement par leur degré de représentativité ou « *fidelity* » (Catano, 2010). De façon générale, ce concept de « *fidelity* » fait référence au réalisme et à l'authenticité de l'exercice de simulation vis-à-vis de la nature du poste et de la situation en tant que telle. Les exercices qui composent les centres d'évaluation sont un bon exemple de mises en situation détenant un fort degré de réalisme étant même parfois qualifiés de « *authentic assessment* » (Lievens et Patterson, 2011). Dans le domaine des centres d'appels, certaines organisations vont donc privilégier l'utilisation d'un échantillon de travail au cours duquel le candidat doit assumer le rôle d'un agent de service à la clientèle au sein d'une organisation fictive et effectuer des tâches semblables à celles qui se présentent dans l'exercice de ces fonctions. Cet instrument est donc une simulation hautement représentative de l'emploi (*high-fidelity simulation*) et permet aux candidats d'agir de façon réaliste comme s'ils étaient en poste. Ainsi, en simulant le rôle et les responsabilités associés à un poste de cette nature, ce type d'instrument d'évaluation est donc basé sur la prémisse qu'il est possible d'effectuer des prédictions sur la performance future d'une personne en fonction de la performance démontrée lors du test. De façon plus spécifique, les exercices de simulation font directement référence à l'approche behavioriste qui postule la « cohérence comportementale » ou, en d'autres mots : « *the best indicator of future performance is past performance* » (Wernimont et Campbell, 1968). En fait, selon cette approche, la validité d'un échantillon de travail est fondée sur l'idée que le comportement observé avec l'instrument, qui est en fait une réplique du travail réel, se répétera dans les situations de travail ultérieures (Lévy-Leboyer, 2002).

Cependant, le caractère hautement représentatif des emplois et la description réelle du comportement adopté par le candidat qui étaient auparavant la marque de commerce des tests situationnels ont progressivement cédé du terrain à des échantillons de travail pouvant être administrés et corrigés plus rapidement à des coûts moins élevés (Lievens et Patterson, 2011). Ainsi, plusieurs organisations ont récemment privilégié l'introduction de tests à choix multiples dans leur processus de sélection. Par le biais de cette nouvelle méthode, on vise plutôt la mesure du comportement projeté ou de ce que la personne croit qu'elle ferait dans une telle situation (son intention) plutôt que sa description réelle (Weekley et Jones, 1999). Dans le même sens, on s'oriente maintenant vers des exercices de simulation qui présentent un faible niveau de réalisme (*low-fidelity simulation*), cette approche étant plus conviviale et moins onéreuse (Lievens et De Soete, 2012). De manière générale,

dans ces nouvelles formes d'échantillons de travail, le candidat doit se positionner en fonction d'une série de de réponses suite à la lecture d'une courte mise en situation. Ces exercices à choix multiples qui s'apparentent aux tests dits de « jugement situationnel » possèdent un degré d'authenticité plus faible (*low-fidelity simulation*) lorsqu'ils sont comparés avec des exercices de jeux de rôle qui mettent réellement en scène le candidat devant une problématique du travail. Incidemment, on observe, depuis les deux dernières décennies, un intérêt accru envers les tests de jugement situationnel, tant de la part des praticiens que des académiciens (Schmitt, 2012). Comme ils sont de plus en plus utilisés par les organisations pour sélectionner leur personnel, il est essentiel de s'attarder à l'évaluation de leur utilité et de leur validité en sélection de personnel (Conger et Riggio, 2007; Gatewood et al., 2008; Joiner, 2002).

## 2. LE TEST DE JUGEMENT SITUATIONNEL

---

Malgré que la première utilisation des tests de jugement situationnel (TJS) remonte aux années 1920, ils ont gagné en popularité depuis les 20 dernières années (Joiner, 2002; McDaniel et al., 2001). En effet, lorsqu'ils sont comparés avec d'autres techniques d'évaluation pour ce type de poste, on note que les TJS sont relativement peu coûteux puisqu'ils peuvent être administrés et corrigés facilement lors de processus impliquant un haut volume de candidatures, ne nécessitant aucune intervention humaine lors de leur correction. Par le fait même, ils permettent généralement d'augmenter la standardisation des évaluations en évitant certains biais et erreurs de mesures associés au degré d'accord entre les correcteurs.

En outre, ils permettent d'évaluer de façon efficiente et intégrée différentes dimensions reliées au leadership, aux relations humaines, ainsi que d'autres capacités et techniques de vente ou de service à la clientèle (Christian, Edwards et Bradley, 2010). D'ailleurs, on note qu'ils sont souvent mis à contribution lors de l'évaluation de certains comportements qui entrent en ligne de compte dans la gestion de situations plus complexes, par exemple dans le cadre de situations de nature interpersonnelle, puisque les mises en situation présentées dans un TJS vont contenir habituellement beaucoup plus d'éléments contextuels que d'autres tests standardisés permettant ainsi des observations plus nuancées (Christian et al., 2010).



Par ailleurs, comme ils sont étroitement reliés au contenu de l'emploi, ils donnent lieu ainsi à une appréciation plus favorable de la part des candidats qui les perçoivent comme un moyen d'évaluation juste et équitable (Banki et Latham, 2010; Bauer et al., 1998). En effet, plusieurs études ont démontré qu'ils tendent à générer des attitudes plus favorables chez les candidats, notamment chez certains groupes visés par l'égalité en emploi (Chan et Schmitt, 1997; Weekley et Jones, 1997). Il s'agit d'un avantage non négligeable étant donné que des réactions positives de la part des candidats peuvent entraîner des bénéfices importants pour une organisation, tels qu'une capacité accrue à attirer de nouvelles candidatures et à conclure des contrats d'embauche (Truxillo et Bauer, 2011). Finalement, un de leurs avantages non négligeables est que les TJS sont beaucoup moins sujets aux impacts négatifs ou « *adverse impacts* » que le sont les tests d'intelligence, par exemple (Clevenger et al., 2001; Motowidlo et Tippins, 1993; Nhung T. Nguyen et al., 2005). Par exemple, une étude de Schmitt et ses collègues (2009) a démontré que l'utilisation de tests de jugement situationnel dans le cadre de processus d'admission d'étudiants à des programmes universitaires permettait d'augmenter significativement le taux d'acceptation des candidats provenant de minorités culturelles et ainsi d'améliorer l'équité entre tous les candidats (Schmitt et al., 2009).

Or, au-delà de la valeur pratique d'un processus de sélection, c'est par le biais de l'application des notions tirées de la « psychométrie » que l'on peut identifier quantitativement et prédire les différences de comportement chez les individus au travail par le biais d'instruments valides et défendables (Edenborough, 2007). Plus spécifiquement, la psychométrie fait référence à la théorie derrière la construction des tests psychologiques et à l'évaluation statistique de leurs qualités métrologiques (Guion, 2010). Cependant, selon Guion (2010), on devrait également y inclure un large éventail de procédures d'évaluation, sans se limiter à l'application spécifique de tests de mesures psychologiques. En effet, toujours selon cet auteur, les exigences dites « psychométriques » devraient s'appliquer dès qu'il y a une action de mesurer tout phénomène, aussi éloigné ce dernier soit-il d'un test psychologique (Guion, 2010).

Par conséquent, les exigences psychométriques associées à la mesure d'un phénomène doivent d'abord être respectées par le moyen ou l'instrument utilisé pour identifier les différences individuelles et prédire le phénomène, c'est-à-dire le prédicteur. Dans le cadre de ce projet de recherche doctorale, le prédicteur qui fait l'objet d'attention est donc le test de jugement situationnel. Ensuite, ce dernier est utilisé en gestion des ressources humaines afin de prédire un phénomène. C'est ce que l'on désigne par

critère. Une première définition d'un critère correspond à une valeur à atteindre, fixée au départ, ou un comportement précis à atteindre, lui aussi fixé au départ. En sélection de personnel, on s'entend généralement pour dire que la notion de critère représente une mesure ou une définition de ce que représente le succès dans l'emploi visé (Gatewood et al., 2008). Voyons maintenant, dans la prochaine partie, ce qui caractérise le prédicteur, c'est-à-dire les tests de jugement situationnel, tant d'un point de vue descriptif que psychométrique. Le critère, c'est-à-dire le rendement en emploi, fera quant à lui l'objet d'attention au chapitre suivant.

---

## 2.1 SA DESCRIPTION ET SES CARACTERISTIQUES

---

Dans sa définition la plus large, le test de jugement situationnel inclut tout test élaboré en vue de mesurer le jugement dans des contextes reliés au travail (Weekley et Ployhart, 2006). Ce type d'exercice présente habituellement une description de situations problématiques qui peuvent être très réalistes en fonction du domaine de l'emploi ou approximatives. Les candidats doivent alors choisir une réponse parmi les différentes options qui lui sont proposées. Il s'agit habituellement de tests de type « papier-crayon », malgré que les mises en situation peuvent également être présentées sous format audio ou vidéo (Lievens et De Soete, 2012). Les questions et les choix de réponses peuvent varier en fonction de leur représentativité avec l'emploi ciblé, c'est-à-dire comment le format de la tâche simulée ressemble à la façon dont ces situations surviennent dans l'emploi. Ils varient aussi en fonction de leur longueur, de leur complexité, de leurs interrelations et selon le type de directives ou d'instructions qui sont fournies aux candidats (Lievens, Peeters et Schollaert, 2008; McDaniel et al., 2001; McDaniel et Nguyen, 2001).

Aux fins de processus de sélection, les TJS sont le plus souvent contextualisés par domaine d'activités (par exemple, le service à la clientèle, la vente, la supervision de personnel, etc.) et, la plupart du temps, ils ne visent pas la mesure d'un construit homogène. En effet, comme il le sera exposé plus loin, on les qualifie habituellement d'instruments de mesure « multidimensionnels », c'est-à-dire qu'ils sont élaborés pour évaluer une variété de construits différents, au même titre que l'entrevue de sélection et le centre d'évaluation, par exemple (Ployhart et MacKenzie Jr., 2011).

Les questions d'un TJS peuvent s'inspirer de deux approches différentes (McDaniel et al., 2001). La première s'oriente davantage vers l'aspect situationnel de la problématique et exige que le candidat

choisisse la réponse la plus appropriée en fonction d'un scénario et d'une liste de réponses possibles. Il s'agit de la méthode du choix forcé que l'on retrouve traditionnellement et le plus fréquemment lors de l'utilisation de n'importe quel test à choix multiple. Le deuxième format de question utilise une échelle de type Likert et demande au candidat d'indiquer son niveau d'accord avec des énoncés ou d'évaluer la pertinence de chacun des comportements présentés. De plus, alors que certains exercices vont demander aux candidats la façon dont ils se comporteraient par rapport à la situation et d'identifier l'action ou l'approche qu'ils adopteraient (ou éviteraient) en fonction du contexte (*Behavioral Tendency Instructions*), d'autres questions vont exiger que les candidats évaluent l'efficacité des choix de réponses possibles ou qu'ils identifient la meilleure (ou la pire) façon de faire face à la situation (*Knowledge Instructions*) (McDaniel et al., 2001).

À titre indicatif, le tableau 2 présente un exemple de question pour chacun de ces formats de directives. Toutes deux citées dans l'ouvrage collectif de Weekley et Ployhart (2006) portant sur les TJS, la première question est tirée d'un examen utilisé dans le cadre du processus de sélection pour les postes d'agent secret au FBI, alors que la deuxième provient du « *Army Judgment Test* ».

**Tableau 2**  
**Exemples de questions illustrant les deux types de directives associées au TJS**  
**(Weekley et Ployhart, 2006, p. 27 & 39, traduction libre)**

<b>Directives de nature comportementale (<i>Behavioral Tendency Instructions</i>)</b>	
Pendant que vous magasinez, vous apercevez un homme en train de commettre un vol. Que faites-vous ?	
a)	Quitter le magasin le plus rapidement possible et appeler la police.
b)	Essayer d'appréhender moi-même le voleur.
c)	Suivre le voleur et appeler la police aussitôt que le voleur semble faire une halte quelque part.
d)	Rien, car vous ne voulez pas être impliqué dans cette affaire.
<b>Directives reliées aux connaissances (<i>Knowledge Instructions</i>)</b>	
Durant une bataille, un soldat en mission urgente réalise qu'il doit traverser une rivière d'une largeur d'environ 40 pieds. Suite à un blizzard, la rivière a gelé et la présence de neige empêche le soldat de connaître l'épaisseur de la glace. Il voit deux planches d'environ 10 pieds près de l'endroit où il voudrait traverser. Il sait aussi qu'il y a un pont 2 miles plus bas, en aval. Dans ces circonstances, il doit :	
a)	Marcher jusqu'au pont et traverser à cet endroit.
b)	Traverser en courant rapidement sur la glace.
c)	Faire un trou dans la glace près de la rive pour en évaluer l'épaisseur.
d)	Traverser à l'aide des planches, en les plaçant alternativement l'une devant l'autre et en marchant dessus.
e)	Traverser en rampant lentement sur la glace.

Par ailleurs, dans le cadre de leur méta-analyse sur la validité des TJS, McDaniel, Hartman, Whetzel et Lee Grubb III (2007) ont dressé une liste des différentes formulations ayant emprunté l'un ou l'autre de ce format de questions. Le tableau 3 présente cette typologie.

**Tableau 3**  
**Typologie du format de questions des TJS en fonction du type de directive**  
**(McDaniel et al., 2007, p. 70, traduction libre)**

<b>Directives de nature comportementale (<i>Behavioral Tendency Instructions</i>)</b>
1. Qu'est-ce que vous feriez ?
2. Qu'est-ce que vous feriez et qu'est-ce que vous éviteriez de faire ?
3. Évaluez et mettez en ordre ce que vous seriez le plus susceptible de faire.
4. Selon une échelle de Likert, évaluez votre tendance à privilégier cette option.
<b>Directives liées aux connaissances (<i>Knowledge Instructions</i>)</b>
5. Qu'est-ce que vous devriez faire ?
6. Quelle est la meilleure réponse?
7. Quelle est la meilleure réponse et quelle est la pire option?
8. Quelle est la première et la deuxième meilleure réponse?
9. Évaluez l'efficacité des choix présentés.
10. Quelle est la première, la deuxième et la troisième meilleure réponse?
11. Quelle est l'importance d'effectuer chacun de ces choix de réponse?

La distinction entre les deux approches de questions est reliée à l'évaluation des concepts de performance typique par rapport à la performance maximale (*typical vs. maximal performance*). Dans les tests visant la mesure de la performance maximale, tels que les tests de connaissances, les tests d'habiletés cognitives, les échantillons de travail et les TJS avec des questions liées aux connaissances, il est possible d'évaluer comment un candidat performera en emploi lorsqu'il s'appliquera à faire de son mieux. En effet, dans ces contextes d'évaluation qui font référence à l'habileté du candidat à identifier la bonne (ou la mauvaise) réponse, ce dernier voudra démontrer de façon claire ses connaissances et ses habiletés cognitives (McDaniel et al., 2007; Nhung T. Nguyen et al., 2005). Alors que par d'autres prédicteurs, tels que les tests de personnalité et les TJS avec des questions de nature comportementale, les candidats doivent indiquer la façon dont ils se comportent habituellement dans le cadre de situations reliées au travail. C'est pour cette raison que ces contextes d'évaluation produisent des mesures de la performance typique (Lievens, Sackett et Buyse, 2009; McDaniel et al., 2007).

Pour conclure, il est intéressant de se reporter au résumé des distinctions établies entre le centre d'évaluation et le test de jugement situationnel afin de comprendre les principaux changements qui ont été introduits dans l'évaluation du personnel (voir tableau 4).

**Tableau 4**  
**Comparaisons entre le test de jugement situationnel et le centre d'évaluation**  
**(Lievens et al., 2008, p. 429, traduction libre)**

	<b>Test de jugement situationnel</b>	<b>Centre d'évaluation</b>
<b>Type de simulation</b>	Simulation de type « <i>Low-fidelity</i> » mesurant une variété de construit	Simulation de type « <i>High-fidelity</i> » mesurant une variété de construit
<b>Stimulus</b>	Contextualisés et reliés à l'emploi Directives écrites ou vidéo standardisées	Contextualisés et reliés à l'emploi Contenu et directives standardisés
<b>Réponse</b>	Réponses écrites « <i>Self-report</i> » À choix multiple	Observations comportementales « <i>Other-report</i> », i.e les évaluateurs Format ouvert
<b>Correction</b>	Déterminée à l'avance par des experts ou de manière empirique	Observation directe (ou vidéo) et notée par des évaluateurs formés
<b>Utilité</b>	« <i>Select-out</i> » <sup>5</sup> Utilisé pour de gros bassins (Internet)	« <i>Select-in</i> » <sup>6</sup> Vise des petits groupes de candidats

---

## 2.2 LE DEVELOPPEMENT D'UN TEST DE JUGEMENT SITUATIONNEL

---

Lors du développement d'un TJS, on distingue généralement deux méthodes, soit les approches déductive et inductive (Weekley et al., 2006). En premier lieu, l'approche déductive consiste à générer les mises en situation à partir d'une revue de la littérature, d'un modèle théorique sous-jacent, d'un profil de compétences ou d'une série d'attributs déterminés à la suite d'une analyse de tâches. Par exemple, Stevens et Campion (1999) ont élaboré un test situationnel à choix multiple visant la mesure de la capacité à travailler en équipe qui a été déployée selon cinq critères dont les fondements s'appuient théoriquement sur une revue exhaustive de la littérature à ce sujet. Parallèlement, c'est également le cas de l'étude de Beldow et Frese (2009) qui ont développé un TJS basé sur la littérature portant sur le concept d'initiative personnelle. Ainsi, cette approche de développement centrée sur la mesure d'un construit spécifique permet alors de cibler un domaine de performance particulier à travers les questions qui composent le TJS.

<sup>5</sup> Visant à choisir les candidats qui seront éliminés du processus de sélection.

<sup>6</sup> Visant à choisir les candidats qui seront retenus pour la poursuite du processus de sélection.

Pour sa part, l'approche inductive repose soit sur des informations recueillies auprès des experts de contenu (SME pour *Subject-Matter Expert*), soit sur une compréhension de l'environnement de travail acquise par le biais de la documentation disponible (par exemple, rapports d'intervention, recueil de plaintes, etc.). Cette approche a été notamment celle adoptée par Motowidlo et ses collègues en 1990 pour développer leur simulation de type « *low-fidelity* » afin d'évaluer le rendement global dans un poste de gestion (Motowidlo, Dunnette et Carter, 1990). D'ailleurs, leur recherche est régulièrement citée comme ayant eu un impact majeur sur la popularité des TJS et la majorité des écrits sur le sujet vont généralement s'inspirer de la méthode adoptée par ces auteurs (Lievens et al., 2008; Ployhart et MacKenzie Jr., 2011).

Comme cette approche de développement est beaucoup plus centrée sur le contenu du travail, il s'agit d'abord de procéder à un examen approfondi des différentes descriptions d'emploi pour le poste à pourvoir. En fait, en analysant les tâches et en dressant la liste des principales activités associées à la fonction, le concepteur développe une meilleure compréhension du poste, c'est-à-dire quel est le pourcentage du temps de travail qui est alloué à chacune des tâches à exécuter, quelle est son importance parmi les autres activités à réaliser, quelle est la difficulté d'apprentissage liée à cette tâche, etc. (Gatewood et al., 2008).

À la suite de cette analyse, des rencontres avec des titulaires de postes et les supérieurs immédiats de cette fonction permettent au concepteur de recueillir des exemples de situations reliées à l'emploi. Plus précisément, ces informations sont habituellement colligées selon la méthode des incidents critiques mise de l'avant par Flanagan (1954) qui consiste à répertorier des observations directes de comportements au travail qui sont significatifs et reliés directement aux dimensions à évaluer :

« By an incident is meant any observable human activity that is sufficiently complete in itself to permit inferences and predictions to be made about the person performing the act. To be critical, an incident' must occur in a situation where the purpose or intent of the act seems fairly clear to the observer and where its consequences are sufficiently definite to leave little doubt concerning its effects. » (Flanagan, 1954, p. 327)

Cette méthode, qui implique un ou plusieurs groupes d'experts composés de superviseurs ou d'employés occupant le poste, est de loin la plus couramment utilisée, et ce, bien qu'elle soit beaucoup plus dispendieuse et compliquée à mettre en œuvre (Weekley et al., 2006). En général, lors de la collecte des incidents critiques, les experts doivent décrire des situations qui sont représentatives du type de problèmes habituellement associés à l'emploi, les comportements qui y sont reliés ainsi que les antécédents d'une situation de ce type à laquelle ils ont déjà été confrontés (Weekley et Jones, 1999). Par surcroît, ces discussions visent à connaître quel était le rôle des personnes et des intervenants concernés par le problème, la manière dont ils ont réagi et quels ont été les résultats de leurs actions. Parfois, les discussions sont très peu structurées, alors que dans d'autres cas, les experts doivent discuter d'événements spécifiques en lien avec les KSAOs identifiés comme devant faire l'objet de l'évaluation (McDaniel et Nguyen, 2001). Bien que les experts de contenu soient habituellement des titulaires de poste ou des superviseurs, lors du développement d'un TJS pour le personnel d'encadrement, il pourrait être bénéfique, par exemple, de recueillir des incidents critiques directement auprès des employés (Weekley et al., 2006). En effet, ces derniers ont sûrement été témoins de plus d'un exemple de situations faisant référence à des comportements de gestion efficaces ou inefficaces qui demeurent inconnus des membres de la direction. Comme Catano (2010, p. 119) le fait remarquer : « a good rule of thumb in choosing SMEs is that the closer the SME is to the job, the better the resultant information will be ». Dans tous les cas, les SMEs doivent être représentatifs de la population visée par le test en termes d'âge, de genre, d'origines ethniques, etc.

Les informations qui caractérisent le domaine d'activité du poste à doter sont ensuite transformées en scénarios problématiques par les concepteurs. Outre leur degré de difficulté à être évaluées, les situations sont également sélectionnées pour leur contribution importante au succès en emploi. Il faut néanmoins demeurer prudent en liant directement les événements réels du travail avec les comportements à évaluer. En effet, bien que cette stratégie soit très efficace pour s'assurer du degré de représentativité du test, les mises en situation peuvent parfois faire appel à des connaissances spécifiques reliées à l'emploi ne pouvant être acquises autrement que par le fait d'avoir occupé cet emploi précédemment. Cet élément pourrait alors créer en retour un biais envers certaines catégories de postulants. Ainsi, il est fortement recommandé de restreindre le domaine de contenu aux connaissances et habiletés qu'une personne doit posséder avant d'entrer en fonction et d'éliminer les éléments qu'une personne peut apprendre par le biais d'une courte période d'adaptation dans le poste (Pettersen, 2000). Dans le même sens, lorsqu'il y a des candidats internes et externes à l'organisation, il

faut éviter le fait que certaines des mises en situation présentées dans le TJS se sont réellement produites. En effet, comme leur dénouement est généralement connu du personnel en place, ces derniers seraient fortement avantagés en comparaison avec ceux de l'externe. Il peut également arriver que certains incidents critiques soient écartés au moment du développement des scénarios parce qu'ils sont trop longs, complexes, inappropriés ou qu'ils soulèvent des questionnements sur le plan légal (par exemple, discuter d'une altercation physique entre deux collègues, faire référence à des valeurs d'une certaine religion, etc.) (McDaniel et Nguyen, 2001).

En somme, selon Weekley et ses collègues (2006), une méthode d'élaboration mixte demeure sans doute la manière la plus efficace et intégrée de mettre sur pied un TJS, c'est-à-dire développer d'abord un cadre conceptuel basé sur la théorie et l'analyse de l'emploi, colliger ensuite des incidents critiques d'une variété de sources puis les intégrer à ce modèle.

Ces étapes d'élaboration permettent également au concepteur de générer des comportements et actions efficaces et d'autres inefficaces qui représentent des façons différentes de faire face aux scénarios élaborés. La série de choix de réponses proposés peut comprendre jusqu'à 12 possibilités de réponse (Weekley et al., 2006). Alors que certains concepteurs vont eux-mêmes élaborer ces choix de réponses en fonction des informations recueillies (par exemple, Stevens et Campion, 1999), certains autres, comme Motowidlo et ses collègues (1990), vont suggérer de remettre les scénarios élaborés à un autre groupe d'experts de contenu en leur demandant de proposer différentes façons de faire face à la situation. En effet, ces personnes qui connaissent très bien l'emploi ont habituellement beaucoup plus de facilité à identifier les meilleures réponses tout en étant en mesure d'identifier des approches beaucoup moins appropriées afin de régler le scénario problématique (Lievens et al., 2008). Ces options peuvent alors être révisées par le concepteur afin d'éviter les redondances et de s'assurer de leur pertinence en fonction du domaine d'activités ou du construit à mesurer. Les scénarios peuvent également être soumis à des personnes sans aucune expérience. Cette façon de faire a l'avantage de susciter une gamme de comportements à l'efficacité très variée (McDaniel et Nguyen, 2001). Cependant, comme il s'agit souvent d'un travail de réflexion et de rédaction assez exigeant, il n'est pas toujours aisé d'obtenir la collaboration de plusieurs personnes (avec ou sans expérience) pour l'élaboration d'un tel instrument de mesure. Dans tous les cas, il est clair que des experts de contenu doivent revoir l'ensemble des choix de réponse afin de s'assurer de leur réalisme et de leur pertinence. En général, c'est également lors de cette étape que sera discutée l'identification des bonnes réponses. Objet



principal de cette discussion, l'établissement de la grille de correction sera donc présenté de façon détaillée dans le cadre de la prochaine section.

---

### 2.3 LA GRILLE DE CORRECTION

---

Lorsque tous les choix concernant les scénarios et les choix de réponse ont été faits et que le TJS a été développé, le concepteur doit alors confectionner une clé de correction. En effet, comme les TJS sont habituellement des exercices à choix multiple, leur correction s'effectue de manière très standardisée et mécaniquement, où chaque réponse du candidat est cotée en fonction d'une grille. Cette étape est critique puisque la grille de correction demeure la principale source des prédictions concernant le rendement futur en emploi. Or, il n'y a habituellement pas qu'une seule bonne réponse objective dans un TJS et, la plupart du temps, plusieurs des options de réponses proposées sont plausibles (Lievens et De Soete, 2012). Ce qui n'est pas le cas lors de l'utilisation d'un test de connaissances, par exemple, où il existe une seule et unique bonne réponse. En fait, les TJS reflètent habituellement des aspects pratiques et sociaux du travail qui sont beaucoup plus complexes à évaluer. Ainsi, la question devient alors de savoir quel est le meilleur scénario de réponse plutôt que d'identifier quelle est la bonne réponse. Dans le même sens, comme il s'agit habituellement d'une mesure de type « *self-report* », visant l'évaluation de caractéristiques difficilement observables comme des traits de personnalité ou d'autres habiletés, la littérature concernant les procédures de correction de ce type d'exercice nous renseigne également sur les meilleures pratiques à cet égard.

Il existe donc plusieurs façons de déterminer l'efficacité des choix réponses suggérés dans un TJS et d'effectuer l'attribution des pointages pour chacun d'entre eux. Dans le même sens, il ne semble pas y avoir de consensus dans la littérature concernant la meilleure stratégie pour élaborer cette grille de pointage (Bergman, Drasgow, Donovan et al., 2006; Krokos, Meade, Cantwell et al., 2004; Weekley et al., 2006). Par ailleurs, les enjeux que l'on retrouve concernant les procédures de correction des TJS sont similaires à ceux que l'on retrouve dans la littérature portant sur l'instrument de données biographiques (« *biodata questionnaire* ») où le consensus concernant la meilleure façon de faire n'a d'ailleurs pas encore été atteint non plus (Bergman et al., 2006).

Dans la littérature portant sur les tests de jugement situationnel, il existe trois catégories de méthodes de correction : (1) théorique, (2) basée sur les experts de contenu, et (3) empirique (Weekley et al., 2006). Par ailleurs, la littérature portant sur les données biographiques nous renseigne également sur différentes méthodes de grilles de correction, c'est-à-dire : (1) rationnelle, (2) empirique, (3) factorielle, et (4) celle des sous-regroupements (Bergman et al., 2006; M. D. Mumford et Whetzel, 1997). À cette liste, Bergman et ses collègues (2006) ajoutent également la méthode hybride. La méthode empirique est donc une approche commune aux deux volets de littérature, bien que beaucoup moins utilisée dans les études portant sur les TJS (Krokos et al., 2004). Dans le même sens, comme c'est le cas pour la méthode des sous-regroupements qui vise à établir des schémas de réponses entre les candidats, la méthode qui consiste à établir la grille de correction par analyse factorielle et corrélations entre les items a connu très peu de publications dans le domaine des TJS (Krokos et al., 2004). En fait, les grilles de correction des questionnaires de données biographiques visent souvent la construction d'échelles en lien avec un construit spécifique. Or, compte tenu du caractère multidimensionnel et des faibles caractéristiques psychométriques des sous-échelles dans les TJS, la plupart du temps, il s'agit d'un score global au TJS qui sera communiqué au candidat (Weekley et al., 2006). Enfin, la méthode rationnelle, quant à elle, se retrouve également dans la littérature portant sur les TJS, mais dans ces écrits, elle fait davantage référence à la méthode théorique et au recours aux experts.

### 2.3.1 APPROCHE THEORIQUE

---

D'abord, en suivant la méthode théorique, les options de réponses sont construites afin de refléter une théorie et il s'agit exclusivement de cette dernière qui est utilisée pour démontrer la pertinence de la réponse en lien avec le scénario présenté. Par le biais de cette approche, il s'agit donc de la mise en relation rationnelle des choix de réponses avec les caractéristiques souhaitées et leurs différences sur le rendement à partir de l'adoption d'un modèle conceptuel. Dans cette perspective, lorsque les candidats choisissent des énoncés qui reflètent le cadre théorique, ils se voient alors attribué des points. Lorsqu'il s'agit d'énoncés qui vont à l'encontre de la théorie ou qui sont jugés étant non pertinents en regard aux attentes théoriques, les candidats n'obtiennent aucun point pour ces réponses.

Cette approche a pour avantage d'offrir au concepteur une relation très étroite entre le contenu des questions et les exigences reliées au rendement en emploi tout en s'assurant de la validité du contenu et de l'interprétation directe des antécédents de la mesure. Par ailleurs, les grilles de correction

établies selon un modèle théorique sont plus facilement généralisables d'un environnement de travail à un autre (Bergman et al., 2006). Cependant, cette technique d'élaboration requiert une compréhension sophistiquée du domaine d'activité et des antécédents liés à un bon rendement dans le poste faisant l'objet d'attention, ce qui n'est pas toujours le cas dans plusieurs disciplines. Par ailleurs, à certaines occasions, les bonnes réponses et leur pertinence théorique peuvent être plus facilement repérables par les candidats ce qui peut en retour augmenter le risque de falsification de leur part (Bergman et al., 2006).

### 2.3.2 APPROCHE DES EXPERTS

---

La méthode la plus courante pour établir une grille de correction dans la littérature sur les TJS repose sur l'implication d'experts (Bergman et al., 2006). En effet, selon Motowidlo, Hooper et Jackson (2006) ainsi que McDaniel et Nguyen (2001), les TJS sont plus souvent corrigés en comparant les réponses des personnes candidates avec les jugements émis par les experts de contenu. Au-delà des employés « experts » ou des superviseurs « experts de contenu », comme il l'a été mentionné précédemment, il peut arriver que des experts sollicités soient d'autres intervenants en lien avec le poste. À titre d'exemple, Weekley et Jones (1997) ont eu recours aux membres de la famille lors de l'élaboration de la grille de correction de leur TJS pour évaluer les personnes au poste d'infirmière à domicile.

Par le biais de cette méthode, les experts ont alors pour rôle de s'assurer que la bonne réponse est indiscutablement la façon la plus efficace de faire face à la situation, alors que les mauvaises réponses, bien que plausibles, représentent une façon moins efficace d'y faire face. En général, on soumet ainsi aux experts identifiés l'ensemble des choix de réponses en leur demandant de porter un jugement sur leur efficacité. La bonne réponse ou la pondération de chacune des options de réponse est ensuite déterminée par consensus ou par d'autres méthodes de calcul (par exemple, la cote d'efficacité moyenne qui a été attribuée à chacune des options, l'écart-type qui représente le niveau d'accord entre les experts, etc.) (McDaniel et Nguyen, 2001; McDaniel, Pstotka, Legree et al., 2011). De cette manière, les personnes candidates obtiendront des résultats plus élevés lorsqu'elles choisiront des actions ou des approches qui seront en concordance avec les jugements des experts de contenu. Par exemple, dans le cadre de leur recherche, Motowidlo, Dunette et Carter (1990) ont demandé à un groupe de cadres exécutifs d'expérience d'évaluer l'efficacité de chacun des choix de réponses soumis afin d'établir leur

grille de correction. Dans le même sens, dans le cas de l'étude de Weekley et Jones (1997), les « clients » devaient indiquer le type d'actions qu'ils étaient le plus susceptibles d'apprécier chez les infirmières en fonction des scénarios présentés. Cette approche a l'avantage de cibler plus facilement dans la grille de correction du TJS les comportements spécifiques qui vont être en adéquation avec la vision stratégique de l'organisation et qui vont lui permettre d'atteindre ses objectifs fixés.

Cependant, les jugements des experts ne sont pas absolus et ils peuvent varier en fonction de l'évaluation propre à chaque personne de l'importance des critères ou de leur compréhension du construit mesuré. Ainsi, il peut parfois survenir des divergences importantes dans l'a priori de leurs appréciations (Motowidlo et al., 1990). Par contre, leurs jugements ne sont pas totalement arbitraires non plus. En général, on élimine les choix de réponses pour lesquels les experts de contenu divergent grandement d'opinion au profit de ceux pour lesquels il y a une convergence dans les jugements émis concernant leur efficacité (McDaniel et Nguyen, 2001).

Dans le même sens, malgré que les recherches à ce sujet soient peu nombreuses, on observe généralement le fait que différents groupes d'experts vont produire des grilles de correction différentes (Weekley et al., 2006). Par exemple, dans l'étude de Weekley et Jones (1997), les résultats obtenus par le biais de la grille de correction établie par les clients corrélaient seulement à 0,48 avec les résultats produits par la grille de correction élaborée de manière empirique. Dans cette perspective, choisir adéquatement le groupe d'experts à consulter est primordial en fonction des objectifs poursuivis. Par exemple, recourir à des cadres supérieurs pour l'établissement de la grille de correction permet de s'assurer que les valeurs et la culture organisationnelle se refléteront chez les futurs employés qui auront réussi avec succès le TJS. En contrepartie, le recours à des titulaires de poste jugés comme performants à titre d'experts de contenu permet de s'assurer que le TJS reflète les normes et standards de rendement actuels.

Pour conclure, il va sans dire qu'il est souvent très fastidieux et malaisé de réunir un groupe d'experts afin de procéder à ce type d'exercice. Ainsi, le temps et les coûts qui sont associés à cet exercice limitent souvent le nombre et le niveau hiérarchique des personnes pouvant y être mis à profit.

D'un autre regard, une clé de correction peut également être construite à partir d'une méthode empirique. Dans ce cas, les bonnes options de réponses sont identifiées à partir d'analyses statistiques. Cette procédure est donc basée sur la capacité du mode empirique à discriminer les membres d'un groupe en fonction de leur relation avec un critère attribué à un groupe de référence (M. D. Mumford et Owens, 1987). Cette méthode est souvent qualifiée de « non théorique » étant donné que son objectif n'est pas de comprendre les construits qui contribuent à la prédiction du phénomène, mais il s'agit plutôt de la prédiction démontrée à travers un critère qui en établit l'efficacité prédictive (M. D. Mumford et Whetzel, 1997).

Dans ce cas, on administre généralement l'instrument à un groupe de personnes et les options de réponses choisies font alors une distinction entre les personnes qui ont eu des résultats différents à un autre critère. Il existe différentes méthodes pour y parvenir. D'abord, il est possible d'analyser statistiquement les relations de manière verticale (relations entre les individus), horizontale (relations entre des variables), de manière corrélacionnelle ou par rapport à la moyenne de l'échantillon. Dans la majorité des cas, les options de réponses qui sont généralement choisies par les personnes qui ont des résultats supérieurs au critère identifié sont alors indiquées comme étant la bonne réponse. Inversement, les options de réponse le plus souvent choisies par les personnes qui performant moins bien sur le critère sont alors identifiées comme étant la mauvaise réponse. À titre d'exemple, c'est l'approche utilisée par Dalessio (1994) qui visait à prédire le taux de rétention des agents d'assurance à l'aide d'un test de jugement situationnel. Plus spécifiquement, par le biais de la méthode horizontale, ce dernier a calculé pour chacun des items du test le pourcentage d'agents qui demeurait en poste après un an d'exercice en comparaison avec ceux qui mettaient fin à leur emploi. En fait, la réponse choisie par le plus fort pourcentage d'agents toujours en emploi était alors considérée comme la meilleure réponse alors que la réponse avec le plus faible pourcentage d'agents en poste était considérée comme la pire réponse.

À titre indicatif, une autre façon de mettre en œuvre cette méthode consiste à comparer les réponses données par des experts en les mettant en relation avec les réponses données par des personnes possédant peu ou pas d'expérience par rapport au poste (Bergman et al., 2006). Dans cette perspective, au lieu de donner leur évaluation de l'efficacité de chacun des choix de réponse, chacun des

deux groupes complète le test en identifiant l'énoncé qu'ils considèrent comme étant la bonne réponse. Par la suite, les énoncés qui sont le plus souvent choisis par les experts sont identifiés comme étant les bonnes réponses (peu importe s'ils ont été choisis ou non par les personnes avec moins d'expérience), alors que les réponses choisies par les personnes avec peu d'expérience (et non choisies par les experts) sont identifiées comme étant incorrectes (Bergman et al., 2006).

Or, la nature et l'efficacité de n'importe quelle démarche empirique dépend de la qualité de la définition et de la mesure du critère utilisé pour définir l'appartenance au groupe ainsi que de la pertinence du groupe de référence par rapport à l'application prévue des données (M. D. Mumford et Owens, 1987). Par exemple, en tenant pour acquis qu'un test est construit pour évaluer l'orientation vers la clientèle, l'utilisation d'un critère comme le rendement dans les tâches pourrait mener à une clé de correction bien différente que l'utilisation de la mesure d'un critère comme le rendement par rapport au service à la clientèle offert.

Par ailleurs, la méthode empirique nécessite habituellement de larges échantillons, raisonnablement bien définis et relativement homogènes afin de pouvoir en généraliser les résultats (Krokos et al., 2004). En effet, différents groupes de référence peuvent produire des différences importantes dans la nature des grilles de correction qui en découlent et la méthode empirique est souvent critiquée pour la variance qu'elle amène selon la spécificité des échantillonnages (Allworth et Hesketh, 1999). Bien que England (1971) recommande un échantillon comprenant un minimum de 75 participants lors de l'élaboration d'une grille de correction basée sur une méthode empirique, il est également possible de s'entendre sur le fait que : « *when deriving weights that reflect the relationship between the item response options and the criterion for the population as a whole, the larger the sample used to derive these weights, the better* » (L. Hough et Paullin, 1994). Pour la même raison, la plupart du temps, une méthode empirique doit être soumise à une validation croisée ou « *cross validation* », c'est-à-dire qu'il devient nécessaire d'estimer la fiabilité d'un modèle à l'aide d'un nouvel échantillon afin de voir si un échantillonnage indépendant produit des résultats similaires (Krokos et al., 2004).

Dans cette perspective, une grille de correction élaborée empiriquement peut parfois être plus difficilement généralisable ayant été développée pour prédire un critère spécifique en lien avec un emploi en particulier. Finalement, il arrive parfois que des conclusions établies exclusivement de façon empirique apparaissent comme irrationnelles et très peu valides aux yeux des candidats et des autres

intervenants non initiés aux méthodes statistiques (Cucina, Caputo, Thibodeaux et al., 2012). En effet, comme le soulignent Cucina et ses collègues (2012, p. 423) : « *Explaining how empirical keys are created to a layperson is a difficult task. (...) Thus, a purely empirical approach (although defensible) could require work to educate opposing parties* ».

#### 2.3.4 APPROCHE HYBRIDE

---

Pour terminer, les différentes procédures de correction présentées précédemment ne sont pas nécessairement mutuellement exclusives. En fait, il est fort probable qu'une combinaison judicieuse de plusieurs stratégies d'élaboration mène à de meilleures mesures de la performance future en emploi (M. D. Mumford, 1999). De cette manière, en adoptant la méthode hybride, il s'agit en fait de combiner différentes grilles de correction ayant été élaborée de façon indépendante (Bergman et al., 2006). Une grille de correction hybride comprend donc habituellement des informations provenant de sources empiriques et des informations de sources rationnelles (Cucina et al., 2012).

Dans les faits, il existe plusieurs façons d'opérationnaliser ce type de démarche. Selon Cucina et ses collègues (2012) qui se sont intéressés à l'élaboration des grilles de correction liées aux inventaires de données biographiques, il est possible d'adopter différentes méthodes statistiques lorsque l'on veut établir une pondération par item ou par option de réponse de manière hybride. Par exemple, par le biais de la démarche « *Hybrid Approximation Weights* », la pondération est alors associée à la somme de celle établie via l'approche empirique avec celle qui découle de l'approche rationnelle (Bergman et al., 2006). « *Because the rational and empirical weights are standardized and unit weighted, the rational and empirical weights have an equal amount of influence on the resulting hybrid weight* » (Cucina et al., 2012, p. 394). Dans le même sens, deux grilles de correction pourraient être combinées de façon à ce qu'un résultat positif obtenu à un item annule un score considéré négatif dans la deuxième grille (Bergman et al., 2006). On pourrait également désigner une de ces grilles de correction comme étant la clé primaire et en identifier une autre comme étant secondaire. Après une première entrée de données avec la grille de correction primaire, on pourrait alors utiliser la grille de correction secondaire en substituant seulement l'ensemble des 0 obtenus. En bref, comme le rappellent Bergman et ses collègues (2006), il est possible d'implanter une grille de correction hybride en combinant ensemble n'importe quel type de traitements statistiques standards.

Par contre, les concepteurs vont habituellement établir de façon empirique la grille de correction en vérifiant par la suite s'il y a convergence entre cette grille et le modèle théorique privilégié ou le jugement des experts (Cucina et al., 2012). Dans ce cas, si les conclusions tirées par le biais de l'approche empirique ne sont associées à aucun fondement théorique, les concepteurs modifient alors grille en conséquence soit en adaptant la pondération soit en éliminant complètement l'item. Comme le soulèvent Hough et Paullin (1994), cette façon de faire, mieux connue sous l'appellation « *construct-based rational scoring* », est souvent adoptée lors de l'élaboration des grilles de correction des inventaires de données biographiques. « *With this type of scoring, the exact nature of the relationship between the items and the construct (i.e. the scoring) is determined empirically though the theoretical link between predictor and criteria remains* » (Krokos et al., 2004, p. 14). Dans tous les cas, des questions importantes doivent cependant être prises en compte lors de la conception d'une grille de correction hybride. À l'évidence, « *there are theoretical (e.g., is combining these keys theoretically justified?) and practical (e.g., is it possible to implement key without cross-validation?) issues to consider* » (Bergman et al., 2006, p. 226).

D'un point de vue opérationnel, il est fort probable qu'un instrument de sélection développé par le biais d'une approche hybride soit plus facilement défendable. En effet, une approche hybride permet de s'assurer que l'identification des bonnes réponses puisse être expliquée et interprétée, tout en maximisant l'effort de prédiction apporté par les méthodes empiriques (Cucina et al., 2012; M. D. Mumford, 1999).

### 2.3.5 L'ATTRIBUTION DES POINTAGES

---

Une fois que les options de réponses ont été déterminées et qu'ils ont fait l'objet d'une discussion sur le plan de leur efficacité, il faut alors identifier la façon dont les bonnes et mauvaises réponses identifiées sont cotées afin d'en arriver à un résultat global au TJS. En effet l'attribution des pointages est essentielle afin de vérifier le degré de réussite de chaque candidat en vue de prendre une décision de sélection. De façon générale, il existe deux méthodes qui sont habituellement adoptées par les concepteurs de TJS dans la littérature selon le type de directive soumis aux candidats (Weekley et al., 2006).



D'abord, lorsque le TJS a été élaboré en fonction de la méthode du choix forcé et que l'on demande aux candidats de choisir la réponse la plus appropriée en fonction d'une liste, on aborde la correction de manière dichotomique (« *single-best-answer method* ») (Muros, 2008). Il s'agit d'une façon très simple qui consiste à attribuer 1 point à chaque fois que le candidat a choisi la réponse qui a été identifiée comme la meilleure façon de faire face à la situation et 0 point lorsqu'il a choisi la réponse identifiée comme étant incorrecte (St-Sauveur, Girouard et Goyette, 2014). C'est une approche que l'on adopte également parfois lorsque le candidat doit se positionner en fonction d'une échelle. En effet, il est possible de lui attribuer 1 point lorsqu'il a identifié correctement la meilleure option de réponse comme étant « Très efficace » ou « Efficace » et 0 point dans le cas où il l'a choisi comme étant « Très inefficace » ou « Inefficace ». Cette approche permet d'éviter, en cas de litige, le fait de devoir justifier concrètement la différence souvent arbitraire entre ces deux ancrages et pourquoi une option est « Efficace » alors que l'autre serait « Très efficace ».

Or, comme les réponses incorrectes ne sont pas toutes également mauvaises, une autre façon de faire consiste à demander aux candidats, pour chacune des questions, quelle est la meilleure option de réponse et quelle est la pire façon de faire face à la situation (« *best-and-worst-answer method* ») (Weekley, Ployhart et Harold, 2004). Cette méthode adoptée, entre autres, par Motowidlo, Dunette et Carter (1990) est connue sous l'appellation « *partial credit scoring* » et a eu beaucoup d'influence dans la littérature sur les TJS (Muros, 2008; Ployhart et MacKenzie Jr., 2011). Avec cette approche de correction, chaque question génère un score allant de -2 à +2. En effet, identifier correctement la meilleure et la pire réponse résulte en un score de 2 points. Par contre, identifier la bonne réponse comme étant la pire et vice-versa donne alors un score de -1 au candidat ou -2 lorsque les deux réponses sont complètement inversées. Finalement, le fait d'avoir choisi les autres options équivaut à un score de 0. Cette façon de procéder a pour avantage de générer plus de variance dans les résultats entre les candidats. Dans le même sens, en plus d'être relativement simple à développer, il semble que cette méthode permette d'évaluer deux types de capacités présentant ainsi une vision plus complète et plus large de la capacité d'une personne à résoudre des problèmes en situation de travail (St-Sauveur et al., 2014). En effet, selon Weekly et ses collègues (2006), alors que l'option de réponse qui correspond à ce que la personne a identifié comme étant la meilleure façon de faire face à la situation est fortement corrélée avec les traits de personnalité, l'option de réponse étant identifiée à la pire façon de faire face à la situation est fortement corrélée avec les habiletés cognitives. Enfin, en pénalisant les erreurs de jugement d'une

personne, cette méthode a l'avantage d'être plus représentative de la performance globale d'un individu au travail (St-Sauveur et al., 2014).

D'autres concepteurs prévoient plutôt que le candidat doit mettre en rang chacune des options de réponses en identifiant la façon la plus efficace de faire face à la situation jusqu'à celle étant la moins efficace (« *rank-ordering of response options method* ») (Weekley, Harding, Creglow et al., 2004). Ainsi, lorsque le candidat doit classer les choix de réponses selon leur efficacité, il est alors possible d'allouer un pointage qui correspond à la moyenne du degré de la relation (c'est-à-dire l'indice de corrélation de Spearman) établie entre leur classification et celle recommandée par les experts (Weekley, Harding, et al., 2004). De cette manière, plus la relation observée est forte, plus le candidat a adopté une classification qui se rapprochait de celle des experts, alors que lorsque l'indice de corrélation est négatif, le candidat a alors porté une appréciation qui était très différente de celle des experts (Weekley et al., 2006).

En ce qui concerne la deuxième méthode, elle s'applique lorsque les candidats doivent utiliser une échelle de type Likert afin d'évaluer leur tendance à adopter les comportements suggérés ou l'efficacité des options de réponses présentées. Dans ce cas, il est possible d'attribuer un pointage aux candidats en fonction de la comparaison entre l'efficacité établie par les experts et celle identifiée par ces derniers. Pour y parvenir, on peut alors retenir la cote d'efficacité attribuée à l'option de réponse obtenue par consensus ou alors utiliser la moyenne de l'ensemble des cotes d'efficacité recueillies. Dans ce dernier cas, la bonne réponse devient alors la moyenne et les réponses des candidats qui attribuent des cotes d'efficacité qui s'en éloignent reçoivent alors des résultats plus faibles.

Par exemple, il s'agit de l'approche adoptée par Chan et Schmitt (2002) qui, après avoir demandé aux candidats d'évaluer l'efficacité de chacune des options de réponse proposées en fonction d'une échelle en 6 points, ont attribué un score de 1, de 2 ou de 3 points pour chaque option de réponse en fonction du pourcentage de concordance entre l'appréciation du candidat et celle des experts (Chan et Schmitt, 2002). C'est l'approche également adoptée par Legree et de ses collègues (2005) qui eux ont procédé en soustrayant chacune des différences obtenues entre la moyenne d'efficacité attribuée par les experts et celle identifiée par les candidats. Par exemple, si la moyenne obtenue pour une option de réponse est 1,5 point, un candidat qui lui attribue un score de 2 sur l'échelle de Likert se voit alors pénalisé de 0,5 point. Ainsi, plus un candidat se rapproche d'un score de 0 (le plus haut résultat pouvant

être atteint), plus il sera considéré comme performant au test (Legree et al., 2005). Certains auteurs ont également procédé en cumulant des scores positifs par le biais d'une mesure de distanciation (« *distance-measure* ») dont le résultat est égal à la somme au carré des différences entre les réponses du candidat et les cotes d'efficacité établie (*square deviation*) (Krokos et al., 2004). Par contre, il est nécessaire de garder à l'esprit qu'avec une telle procédure d'attribution de pointages, les candidats pourraient avoir tendance à adopter une stratégie de réponse qui vise à éviter les ancrés extrêmes de l'échelle. Cette façon de faire pourrait donc leur permettre de maximiser le nombre de points obtenus, leur réponse présentant toujours un petit écart avec la bonne réponse attendue, améliorant ainsi artificiellement leur score obtenu à l'exercice (Cullen, Sackett et Lievens, 2006).

Sur le plan des avantages associés à cette procédure de correction, comme chaque option de réponse est associée à une pondération, il va sans dire que cela permet d'accumuler beaucoup plus de données permettant de contribuer à démontrer la validité et la fidélité du test (McDaniel et Nguyen, 2001). Dans le même sens, comme chaque réponse est indépendante l'une de l'autre, les résultats obtenus ne donnent pas de mesure de type ipsatif. En d'autres mots, le score obtenu à une variable ne dépend pas du score obtenu à une autre ce qui rend les comparaisons entre les individus beaucoup plus faciles. En fait, lorsqu'un test donne une mesure ipsative, l'élévation d'un score à une échelle va de pair avec une diminution des scores aux autres échelles ce qui permet seulement une comparaison de la maîtrise de ces échelles chez cet individu en particulier.

Finalement, parmi les avantages, on remarque également que cela sera beaucoup plus difficile de retenir l'ordre des réponses identifiées suite à la passation d'un TJS ce qui, non seulement, permet plus facilement d'administrer le même examen à plusieurs reprises à un même candidat, mais augmente également la validité apparente du test en laissant aux candidats la perception que le test est plus difficilement falsifiable (St-Sauveur et al., 2014).

Cependant, cette approche comprend également des inconvénients. En fait, lorsqu'un test utilise une échelle de Likert, en plus d'être beaucoup plus complexe à développer, il peut devenir très ardu en cas de litige de justifier rationnellement la différence entre une option valant 2 points et identifiée comme étant « Très efficace » alors qu'une autre très similaire équivaut à un score de 1 point puisqu'elle a seulement été jugée comme étant « Efficace ». D'un point de vue organisationnel, devoir établir la distinction entre ces deux significations peut effectivement causer certains préjudices. Par ailleurs, tel

que mentionné précédemment, un candidat pourrait également s'employer à éviter de choisir les ancrages de l'échelle qui se retrouvent aux deux extrêmes de façon à améliorer artificiellement son score au test.

Dans tous les cas, lorsqu'on s'intéresse à l'efficacité relative d'une grille de correction et des procédures de correction d'un TJS, il est essentiel de se tourner vers les critères permettant d'évaluer les tests, les pratiques de testing et les effets de leur utilisation qui ont été compilés et codifiés dans « *Standards for Educational and Psychological Testing* » (1999). Ces critères ont été élaborés grâce aux efforts de trois organisations : *American Educational Research Association* (AERA), *American Psychological Association* (APA), et *National Council on Measurement in Education* (NCME). Il existe d'ailleurs un consensus au sein de la communauté scientifique et des praticiens qui jugent que les meilleures pratiques en évaluation et en sélection du personnel découlent directement de l'application de ces standards, principes et lignes directrices (Pettersen, 2000).

Les exigences psychométriques devant être suivies en contexte de sélection de personnel consistent à valider l'outil d'évaluation, à mesurer son degré de fidélité et à prendre en compte ses considérations éthiques et légales. Par ailleurs, comme le font remarquer Putka et Sackett (2010, p. 9), « *reliability and validity are concepts that provide scientific foundation upon which we construct and evaluate predictor and criterions measures of interest in personnel selection* ». De cette manière, ces deux critères doivent être définis et mis en relief avec les caractéristiques propres au TJS afin d'évaluer l'efficacité relative de sa grille de correction.

---

## 2.4 LE CONCEPT DE VALIDITE

---

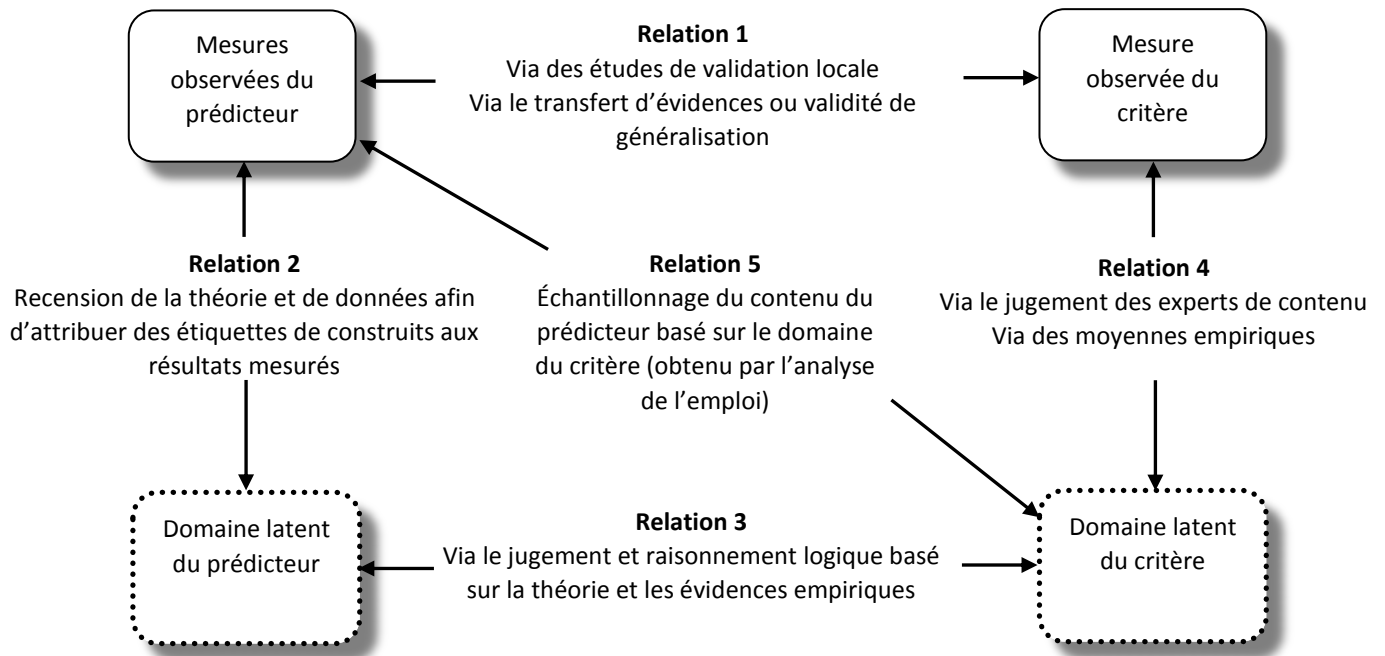
Au-delà de la démarche liée à son développement et sa correction, il est essentiel que le TJS soit conforme aux paramètres psychométriques définis dans la littérature scientifique et professionnelle. D'un point de vue théorique, la validité est la qualité première et le principal aspect à considérer lorsque vient le temps de déterminer et d'évaluer la qualité d'un instrument de mesure (Guion, 1998). La validité indique jusqu'à quel point l'instrument parvient à mesurer ce qu'il est censé mesurer ou à prédire ce qu'il est censé prédire (Pettersen, 2000).

De façon plus spécifique, la validité évalue jusqu'à quel point les inférences faites à partir des résultats recueillis à un instrument de mesure sont exactes et fondées dans la situation particulière où elles sont appliquées (Nunnally, 1978; Pettersen, 2000). En effet, si les inférences ne sont pas valides, les résultats ne peuvent pas servir à décrire le répondant (inférence descriptive) ni à prédire son comportement (inférence relationnelle). De cette manière, la validité se rattache à l'accumulation de tous les éléments de preuve concernant un instrument de mesure, et ce, peu importe la démarche de validation utilisée (Pettersen, 2000). C'est pourquoi on ne discute plus de différents types de validité dans la littérature, mais plutôt de différents types d'évidences permettant d'attester de la validité des inférences (Putka et Sackett, 2010).

En bref, pour vérifier la validité de ces inférences, on propose différentes approches ou stratégies de validation interreliées : la validité de construit et la validité de contenu, que l'on regroupe en gestion des ressources humaines dans la catégorie « *evidence based on test content* », et la validité critériée que l'on nomme aussi « *evidence based on relations to other variables* » (Catano, 2010; Gatewood et al., 2008). Les inférences qui sont réalisées dans le cadre d'un processus de sélection impliquent donc autant les deux stratégies de validation. Les conclusions que l'on en tire sont multidimensionnelles, c'est-à-dire qu'elles touchent autant au contenu de la procédure qu'à la relation avec une autre variable. Cependant, la faisabilité et la pertinence de conduire une ou l'autre de ces stratégies de validation peuvent parfois varier en fonction du contexte organisationnel (le nombre de candidats, le caractère nouveau de l'emploi, etc.) (Tippins, Papinchock et Solberg, 2010). Malgré tout, comme le rappellent Putka et Sackett (2010, p. 40), « *one's obligation is simply to provide evidence in support of the inferences one wishes to draw* ».

À cet effet, l'adaptation du modèle de Binning et Barrett (1989) par Putka et Sackett (2010) permet de visualiser le caractère unitaire de la notion de validité en sélection de personnel. En effet, cette illustration présente l'ensemble des stratégies de validation visant autant des inférences de nature relationnelle que la prédiction du rendement en emploi – voir figure 3. Les prochaines sections exposeront les grandes lignes de ces relations d'un point de vue théorique.

**Figure 3**  
**Différentes stratégies de validation en sélection de personnel**  
**(Putka et Sackett, 2010, p. 41, traduction libre)**



#### 2.4.1 STRATEGIE DE VALIDATION BASEE SUR LE CONTENU

En premier lieu, dans le monde du travail, l'inférence primordiale à établir lorsque l'on utilise un instrument de mesure est son lien avec l'emploi. Dit autrement, pour un processus de sélection valide, il est clair que l'ensemble des exercices qui le composent doit refléter les types et le niveau de complexité des activités reliées à l'emploi à pourvoir, c'est ce que l'on désigne par validité de contenu – voir relation 5 de la figure 3. Le réalisme et la représentativité d'un outil d'évaluation et d'un processus de sélection sont donc vérifiés par le degré de relation entre son contenu (mises en situation, questions, tâches, procédures d'administration, de correction, etc.) et le contenu du poste visé (activités, tâches, rôle et responsabilités, etc.) (Arthur Jr et Villado, 2008; Gatewood et al., 2008). Dans ce contexte, un outil d'évaluation est valide si son contenu est représentatif des tâches dont doit s'acquitter le candidat dans le poste visé et si le contenu des réponses recherchées à l'outil est représentatif du domaine de performance requis par l'emploi.

La validité de contenu est donc intimement liée à l'analyse de fonction et au profil d'exigences qui en découle (Bourhis, 2007). De cette manière, les moyens d'évaluation doivent d'abord être adaptés aux déterminants de cet emploi et la gestion des activités de sélection consiste d'abord à analyser le travail à effectuer sur le plan des tâches et des comportements à adopter afin de proposer un certain nombre de caractéristiques requises pour exécuter correctement le travail (Durivage, 2004). L'objectif de l'analyse de la fonction est d'en arriver à circonscrire et à mieux comprendre l'ensemble des éléments d'un poste de travail. On met ainsi en place un processus systématique de recueil, de documentation et d'analyse visant à identifier les éléments critiques de l'emploi. Une combinaison de méthodes de collecte de données différentes mène à dresser un portrait le plus complet possible de ce qui est requis pour travailler dans ces fonctions. Habituellement, on débute l'analyse de fonction par la description d'emploi qui comprend les exigences du poste, la façon dont le titulaire doit s'y prendre pour parvenir à les remplir ainsi que la justification derrière certaines procédures de travail (Catano, 2010). Par contre, avec les environnements de travail turbulents des dernières décennies, il est toujours recommandé de passer du temps avec les employés qui occupent des fonctions similaires afin de voir ce qu'il est en réellement sur le terrain. On observe et on porte attention aux tâches en les traduisant en comportements et caractéristiques personnelles requises. On peut également consulter des experts de contenu ou « SME » pour *Subject-Matter Experts* qui connaissent très bien l'emploi et les façons dont il doit être réalisé. Que l'on procède par entrevue, observation, questionnaire structuré ou par le biais de la méthode des incidents critiques, il est clair que l'analyse de l'emploi est beaucoup plus qu'une simple méthodologie et qu'elle fait référence à une multitude de techniques afin d'en arriver à dresser le portrait le plus fidèle possible du contenu de l'emploi.

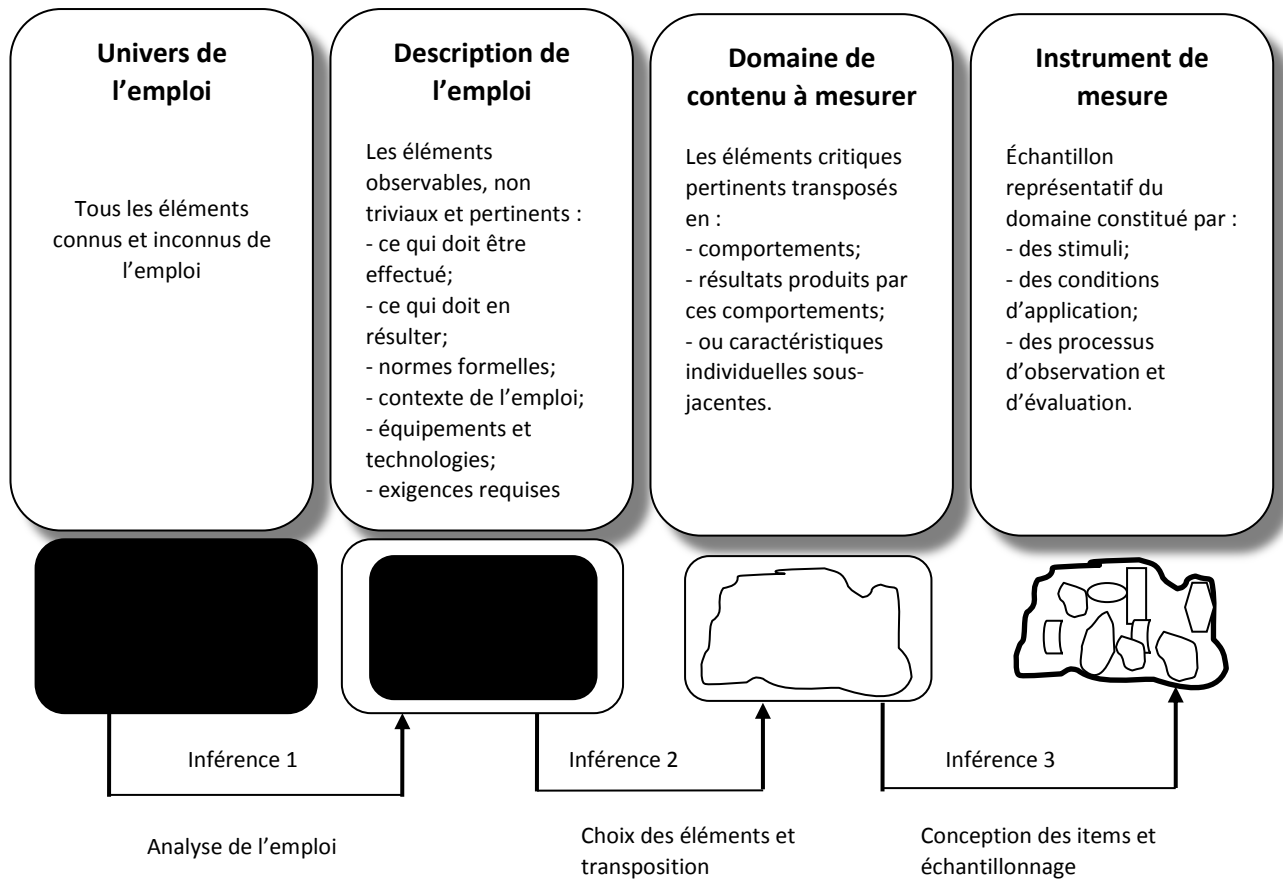
Par ailleurs, comme indiqué précédemment, la majorité des auteurs vont plaider aujourd'hui en fonction d'une analyse de l'emploi beaucoup plus stratégique et le résultat d'un effort multidisciplinaire partant du plus général (c'est-à-dire analyser et comprendre le contexte et la stratégie organisationnelle) au plus spécifique (analyser ce que font les travailleurs réellement) plutôt que l'inverse (Pearlman et Sanchez, 2010). Dans le monde du travail du XXI<sup>e</sup> siècle, l'analyse de tâches ou de poste traditionnelle est dorénavant remplacée par une analyse beaucoup plus globale de l'emploi qui inclut, non seulement le contenu de l'emploi (tâches, responsabilités, résultats), les attributs nécessaires pour effectuer correctement ce travail (KSAOs), mais également le contexte dans lequel l'emploi évolue (conditions physiques et psychologiques, contexte organisationnel global et environnement externe) (Pearlman et

Sanchez, 2010). En effet, quand on pense aux nouvelles formes d'organisations du travail, il est clair que ces transformations sollicitent de nouvelles formes d'analyse des emplois qui vont incorporer des mesures d'une grande variété de facteurs, notamment ceux reliés au contexte du travail, particulièrement ceux orientés sur la vision, la mission, la stratégie, la structure, la culture et les valeurs organisationnelles. En effet, par opposition à l'organisation tayloriste du travail, les nouvelles formes d'organisations du travail chercheraient à regrouper les tâches inhérentes à un poste, à accorder plus d'autonomie et de flexibilité aux employés ainsi qu'à leur confier davantage de responsabilités (Chênevert, Dubé et Déom, 2008). En termes concrets, on observe donc de plus en plus dans le monde du travail d'aujourd'hui une rotation entre les postes, l'élargissement des tâches, une réingénierie des processus, des cercles de qualité et des équipes de travail (Chênevert et al., 2008). Alors que la littérature en GRH continue de présenter des méthodes d'analyse traditionnelles reliées aux tâches spécifiques et aux caractéristiques particulières des emplois, comme Perlman et Sanchez (2010, p. 89) l'indiquent, « *definition, measurement, and mapping of the work environment – (...) – at multiple levels of analysis is the next major frontier in work analysis* ».

À ce sujet, le modèle de Pettersen (2000) illustre de façon intéressante le processus associant l'analyse de l'emploi à l'opérationnalisation d'un processus de sélection par le biais d'un instrument de mesure. Ce schéma peut ainsi servir de guide pour mettre en œuvre une stratégie de validation basée sur le contenu lors de l'élaboration d'un nouveau prédicteur (voir figure 4).



**Figure 4**  
**Processus de définition du contenu d'un instrument de mesure à partir de l'emploi**  
 (Pettersen, 2000, p. 235)



En somme, la validité d'un processus de sélection passe par la validité du processus d'analyse de l'emploi. À cet effet, plusieurs auteurs vont discuter de la difficulté à recueillir des informations valides et précises à partir de l'analyse de l'emploi (Catano, 2010; Harvey et Lozada-Larsen, 1988). Par exemple, il semble que différents groupes d'experts pourraient conduire à des résultats différents au niveau de l'analyse des emplois et la quantité d'informations transmises aux analystes au préalable pourrait en retour en affecter la précision (Harvey et Lozada-Larsen, 1988). Cependant, comme le rappelle Catano (2005), il existe une multitude de techniques pour réaliser une analyse de poste adéquate et le choix des techniques mises à contribution doivent être basées sur une évaluation des objectifs du processus d'analyse, des ressources disponibles ainsi que des besoins de l'organisation. Dans tous les cas, la position de Perlman et Sanchez (2010) sera retenue à l'effet que les leaders et les intervenants RH dans les organisations demeurent la courroie de transmission permettant de lier la sélection de personnel (et autres pratiques RH) avec le contenu pertinent à l'emploi.

Dans une autre perspective, la validation basée sur le construit mesuré tente de démontrer si les résultats obtenus à un instrument de mesure, par exemple, reflètent réellement le construit devant être mesuré (Pettersen, 2000). Celle-ci indique jusqu'à quel point le prédicteur permet de mesurer le concept (le construit) qu'il est censé représenter. Dans la figure 3, les relations 2 et 4 représentent cette forme de validation en mettant en lumière le lien entre un construit et une mesure. Traditionnellement, un construit a été défini comme une caractéristique psychologique ou un objet abstrait qui ne peut pas être observé directement, mais qui se manifeste à travers divers comportements. Par exemple, dans les tests psychologiques, le construit fait référence à un postulat que l'on attribue à une personne et que l'on assume qu'il sera reflété selon la performance aux tests (Cronbach et Meehl, 1955). On pourrait alors parler du construit de l'intelligence, de l'estime de soi, de l'anxiété, du leadership, la motivation, etc. Dans tous les cas, c'est une construction théorique qu'il est nécessaire de définir afin de le mesurer. À cet égard, la validation d'un instrument par le construit devrait être impliquée dès que ce dernier doit être interprété comme une mesure d'un attribut ou d'une qualité qui n'est pas opérationnellement définie (Cronbach et Meehl, 1955). Cependant, dans le domaine de la gestion des ressources humaines, la notion de validité de construit devient généralement implicite lorsque l'on mène une stratégie de validation basée sur le contenu puisque dans les deux cas on s'intéresse à des comportements qui sont reliés au rendement en emploi et qui doivent être définies pour faire l'objet d'une mesure.

Pour conclure, les jugements de similarités entre le contenu de l'exercice et le poste sont les types d'évidences les plus utilisés pour démontrer la validité des instruments utilisés en sélection de personnel. En fait, comme le fait remarquer Pettersen (2000), l'élaboration d'instruments de mesure en sélection de personnel est souvent du ressort des professionnels en ressources humaines et la validation basée sur le contenu est la seule approche qui soit à la fois rigoureuse, aisément défendable en cas de litiges et accessible pour mener à bien cette responsabilité. Dans le même sens, la plupart des études démontrent qu'il est beaucoup plus facile pour une entreprise de se défendre en cas de contestation si elle a mené une démarche de validation basée sur le contenu, tant parce qu'elle retient davantage l'attention des magistrats que parce qu'elle est plus facilement acceptée par les utilisateurs et les candidats (Bourhis, 2007). Cependant, la validité de contenu est subordonnée à la validité critériée dans la mesure où même si l'on indique que l'instrument est représentatif du rendement à offrir en emploi sur le plan du contenu, les organisations devraient néanmoins conduire une procédure de validation critériée pour s'assurer que c'est effectivement le cas (Cook, 2009).

#### 2.4.2 LA VALIDITE DE CONTENU DU TEST DE JUGEMENT SITUATIONNEL

---

Bien que les TJS soient reconnus comme des tests situationnels de la catégorie « *low-fidelity simulation* », la validité basée sur le contenu de l'emploi constitue habituellement la principale source de leur validité (McDaniel et Nguyen, 2001). En effet, comme il l'a été expliqué précédemment, le développement de ce prédicteur repose essentiellement sur la collecte d'incidents critiques ou d'un modèle qui présente l'ensemble des déterminants du rendement en emploi pour le poste à pourvoir (Weekley et al., 2006). De cette manière, étant donné que le contenu du TJS et le développement des questions et des réponses sont basés sur une analyse des tâches et sur le jugement des experts de contenu, cette façon de faire contribue sans contredit à la validité de l'instrument sur le plan de son contenu (Ployhart et MacKenzie Jr., 2011).

Dans le même sens, les questions de ce type d'exercice sont généralement développées selon une méthodologie orientée vers le critère plutôt que sur le construit. Par exemple, lorsqu'utilisées dans le cadre de processus de sélection de personnel, les situations des TJS n'ont qu'une seule chose en commun, elles sont importantes pour le succès en emploi (Kanning, Grewe, Hollenberg et al., 2006). En effet, les questions à l'intérieur d'un même TJS sont aussi hétérogènes que le sont les situations réelles au travail et leur solution requiert habituellement une variété de compétences. Dans cette optique, les avis sont divergents sur la signification des résultats obtenus à ce type de test et leur validité de construit demeure incertaine (Ployhart et Ehrhart, 2003). Par ailleurs, comme il n'est pas possible d'établir la validité de construit d'une méthode (on peut seulement établir si les items utilisés à l'intérieur de cette méthode sont reliés aux construits à mesurer), il est donc essentiel de se reporter aux dimensions devant faire l'objet de la mesure dans ce cas en particulier afin d'établir validité de construit d'un TJS.

Or, plusieurs auteurs sont plutôt d'avis que cet instrument représente une méthode permettant de mesurer une variété de construits hétérogènes (*Measurement Method*), comme il s'agit du cas des entrevues, des tests de connaissances, des centres d'évaluation et des exercices de simulation (McDaniel et al., 2007; McDaniel et Nguyen, 2001; McDaniel et Whetzel, 2005; O'Connell, Hartman, McDaniel et al., 2007). En fait, la variété des construits qui sont reliés aux TJS et leur nature multidimensionnelle impliquent, selon ces auteurs, qu'il est difficile de cibler un construit spécifique au détriment des autres.

Dans le même sens, même si les TJS sont développés pour évaluer un construit spécifique, il semble qu'ils continueront d'évaluer les autres construits.

La plupart des écrits s'inscrivant dans ce courant se sont donc intéressés aux différents facteurs qui corréleront positivement avec les TJS. Par exemple, McDaniel et ses collègues (2006) ont compilé les données se rapportant aux différents construits qui sont évalués par les TJS, et ce, selon la nature des directives qui sont soumises aux candidats par le biais de la méthode de la méta-analyse. Le tableau 5 présente les résultats de cette étude de McDaniel et ses collègues (2006).

**Tableau 5**  
**Résultats des corrélations obtenues entre différents facteurs et les TJS**  
**en fonction du type de directive**  
**(McDaniel et al., 2006, p. 191, traduction libre)**

<b>Corrélations des facteurs avec les TJS</b>	<b>N</b>	<b>Nb de recherches</b>	<b><i>p</i></b>
<b>Habiletés cognitives</b>	22 553	62	0,39
Directive de nature comportementale	5 263	21	0,23
Directive reliée aux connaissances	17 290	41	0,43
<b>Agréabilité</b>	14 131	16	0,33
Directive de nature comportementale	5 828	11	0,53
Directive reliée aux connaissances	8 303	5	0,20
<b>Conscience</b>	19 656	19	0,37
Directive de nature comportementale	5 902	11	0,51
Directive reliée aux connaissances	13 754	8	0,33
<b>Stabilité émotionnelle</b>	7 718	14	0,41
Directive de nature comportementale	5 728	10	0,51
Directive reliée aux connaissances	1 990	4	0,11

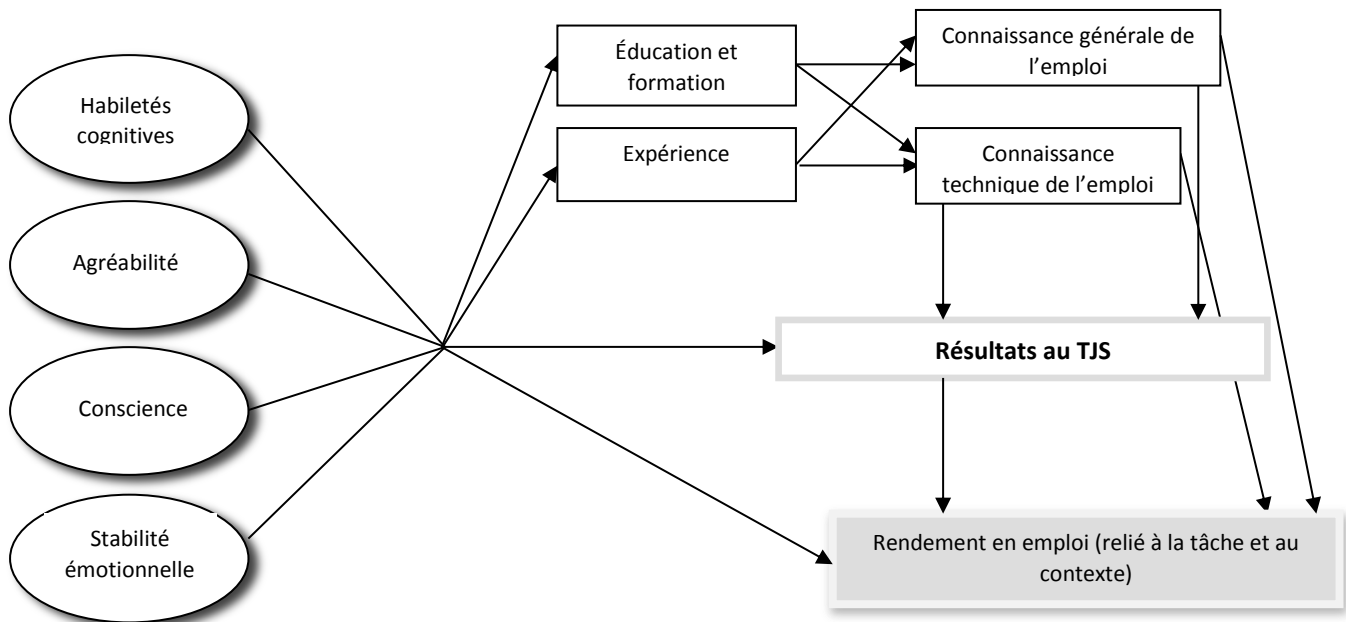
En résumé, les TJS sont corrélés positivement avec les habiletés cognitives, l'agréabilité, la conscience et la stabilité émotionnelle. Qui plus est, alors que les questions comprenant des directives de nature comportementale (*would do*) sont davantage corrélées avec les dimensions relatives à la personnalité, les questions avec des directives reliées aux connaissances (*should do*), sont davantage corrélées avec les habiletés cognitives. Ces auteurs font également remarquer que la validité de construit des TJS sera également influencée par le type de test. En effet, plus la longueur, la complexité et l'interrelation entre les questions seront élevées, plus les TJS seront influencés par le facteur cognitif. À l'inverse, lorsque le TJS sera administré par le biais de vidéos, le facteur cognitif aura beaucoup moins d'influence sur le plan des résultats compte tenu de la diminution des exigences relativement à la lecture

des scénarios. Ces conclusions démontrent qu'il est donc possible de changer la validité de construit des TJS en modifiant la nature des questions soumises aux candidats (McDaniel et al., 2006).

D'autres auteurs ont également démontré que les résultats au TJS corrélaient positivement avec l'expérience à 0,23 (Weekley et Jones, 1999). Par contre, il n'est pas certain que cette relation soit aussi simple, l'expérience étant également un construit multidimensionnel. En fait, selon Weekley et Jones (1999), une mesure adéquate de l'expérience lorsque l'on s'intéresse au jugement en emploi est plus qu'une mesure du nombre d'années d'expérience auprès d'un employeur. Il s'agit plutôt d'un champ vaste incluant une combinaison du nombre d'années dans l'industrie, le nombre d'années à temps plein sur le marché du travail, la durée du dernier emploi à temps plein, la plus longue durée avec n'importe quel employeur et le nombre d'employeurs à ce jour. Finalement, d'autres auteurs ont compilé des évidences à l'effet que les résultats au TJS pouvaient également être influencés par les connaissances propres à l'emploi (Clevenger et al., 2001). Cependant, selon Weekley et Ployhart (2005), on parle davantage d'une évaluation des connaissances générales plutôt que des connaissances spécifiques à un emploi.

Le modèle conceptuel présenté à la figure 5 permet de visualiser l'ensemble des facteurs pouvant influencer la validité d'un TJS tout en résumant très bien le caractère multidimensionnel du construit évalué. Élaborée par McDaniel et ses collègues (2006) cette figure démontre que les résultats obtenus au TJS sont influencés directement par certaines caractéristiques individuelles et indirectement par leur effet sur l'expérience, l'éducation, la formation et les connaissances. De façon plus précise, les quatre caractéristiques individuelles qui influencent les résultats obtenus au TJS sont les habiletés cognitives d'une personne, son agréabilité, son caractère consciencieux ainsi que sa stabilité émotionnelle. Celles-ci viennent alors influencer les connaissances reliées à l'emploi apprises par l'individu, et ce, à travers les effets directs qu'auront ses apprentissages réalisées par le biais de son éducation et de sa formation. Dans le même sens, une personne intelligente et responsable aura également plus de chances d'acquérir davantage de connaissances (générales et techniques) reliées au travail. Enfin, le rendement en emploi devient alors une prédiction de ces quatre caractéristiques individuelles, des résultats obtenus au TJS et de ses connaissances reliées à l'emploi (McDaniel et al., 2006).

**Figure 5**  
**Modèle conceptuel des facteurs influençant la validité de construit des TJS**  
 (McDaniel et al., 2006, p. 195, traduction libre)



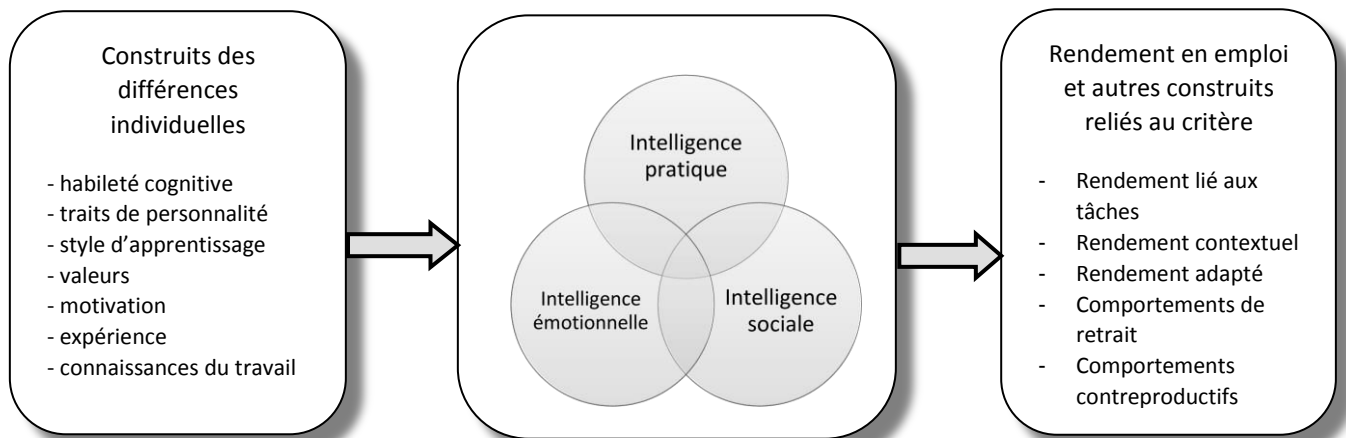
Cependant, un autre volet de la littérature portant sur la validité de construit des TJS repose sur la croyance implicite qu'il existe un construit spécifique relié au jugement situationnel. Certaines études en ce sens ont donc cherché à comprendre ce construit unique. Par exemple, certains auteurs ont avancé que les TJS mesurent un nouveau construit qui dépasse la mesure traditionnelle du facteur *g* et de l'expérience, c'est-à-dire une sorte d'intelligence pratique qui permet à une personne d'intervenir adéquatement dans les situations qui surviennent quotidiennement au travail (Sternberg, Wagner, Williams et al., 1995). À partir de cette conception, une des explications du pouvoir de prédiction du TJS est qu'il mesure un aspect particulier de l'intelligence qui n'a rien à voir avec le volet académique (*street smart vs. book smart*), mais plutôt « *a combination of creative, analytical, and practical skills* » faisant référence à « *the ability to adapt to, shape, and select real-world environments* » (Stemler et Sternberg, 2006, p. 109). Dans le même sens, d'autres auteurs ont centré leurs recherches sur la mesure de l'intelligence émotionnelle (« *the ability to monitor one's own and others' emotions, to discriminate among them, and to use the information to guide one's thinking and actions* »<sup>7</sup>) ou de l'intelligence sociale (« *the ability to understand and manage men and women, boys and girls – to act wisely in human relations* »<sup>8</sup>) par le biais des TJS (Chan, 2006; O'Sullivan, 2007; Schulze, Wilhem et Kyllonen, 2007).

<sup>7</sup> Salovey, P. et Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, 9, 185-211.

<sup>8</sup> Thorndike, E. L. (1920). Intelligence and its uses. *Harper's Magazine*, 140 (227-235).

Dans tous les cas, plusieurs critiques ont été émises à l'effet que ces trois formes d'intelligence ne sont pas nécessairement structurellement indépendantes les unes des autres (Lievens et Chan, 2010). Tirant ses origines de travaux élaborés par Chan et Schmitt (2005), Lievens et Chan (2010) ont récemment proposé un modèle permettant de conceptualiser le construit évalué par un TJS (voir figure 6 11). Ce dernier stipule que les trois types d'intelligence, c'est-à-dire les critères faisant l'objet de l'évaluation dans un TJS, agissent à titre de variable modératrice dans la relation prédictive reliant les caractéristiques individuelles avec le rendement en emploi. Dans le même sens, le modèle met l'accent sur le fait que les KSAO, qui eux sont de nature plutôt unidimensionnelle, jouent également un rôle dans la détermination de ces trois types d'intelligence. Par ailleurs, bien qu'elles soient distinctes, ces formes d'intelligence sont interreliées entre elles, chacune ayant un espace formant un construit commun et chacune une part de construit unique.

**Figure 6**  
**Modèle conceptuel pour examiner l'intelligence pratique, émotionnelle et sociale**<sup>9</sup>  
 (Lievens et Chan, 2010, p. 350, traduction libre)



<sup>9</sup> Adapté par les auteurs de Chan, D. et Schmitt, N., Situational judgment tests, in A. Evers, O. Smit-Voskuil, & N. Anderson, Eds., *Handbook of personnel selection*, 219-242, Blackwell, Oxford, United Kingdom, 2005.

En sélection de personnel, la relation entre deux variables revêt un intérêt considérable (Catano, 2010). En effet, par exemple, un responsable de ressources humaines sera intéressé à vérifier comment les résultats obtenus à un instrument d'évaluation et de sélection seront associés à un rendement en emploi différent. De cette façon, il s'agit généralement de mettre en relation le rendement au travail (le critère) avec les résultats aux instruments de sélection (prédicteurs) – voir relation 1 de la figure 3.

Communément appelée la validité critériée, cette forme de validation désigne donc la capacité d'un instrument à prévoir un phénomène futur. Dans le cas de la sélection, cette stratégie de validation sert à vérifier s'il existe une corrélation entre les résultats individuels obtenus au processus de sélection et un critère, c'est-à-dire le rendement au travail. Cette stratégie vérifie jusqu'à quel point les différences individuelles, telles que mesurées par l'instrument, permettent de prédire le comportement d'un individu au travail. De cette manière, on se réfère de nouveau à l'analyse de l'emploi qui apporte de l'information permettant de mieux comprendre le critère, c'est-à-dire le contenu de l'emploi (activités, comportements et résultats), mais également le contexte qui permet de fixer les attentes et les normes de rendement. Compte tenu du caractère multidimensionnel du rendement en emploi et des défis reliés à la mesure de ce critère lorsque l'on conduit une stratégie de validation en sélection de personnel, cette notion sera traitée de façon détaillée dans la prochaine section.

Lorsque l'on s'intéresse à une stratégie de validation basée sur la relation avec d'autres variables, on peut parler de validité concomitante ou de validité prédictive. La principale différence entre ces deux types est de nature temporelle (Cook, 2009). Dans le cadre d'une stratégie prédictive, la vérification de la validité se fait en comparant la relation entre les résultats des sujets à l'outil d'évaluation et leur performance future au travail. Dans le premier cas, les inférences sur la prédiction de la performance au travail du candidat peuvent être émises dès la première administration de l'instrument de mesure, les données ayant trait à un groupe d'employés travaillant déjà dans l'emploi étudié (Pettersen, 2000). Bien que le schème concomitant offre l'avantage de faire connaître assez rapidement les résultats concernant la validité de l'instrument étudié, il faut également demeurer prudent étant donné que l'utilisation de cet échantillon peut également présenter certains biais en sélection de personnel (Pettersen, 2000). En effet, les employés déjà en place représentent une population distincte de celles des candidats qui aspirent à occuper un poste en particulier en termes



d'expérience, d'ancienneté et de motivation, par exemple (Arvey et Faley, 1992). Dans le même sens, le schème concomitant entraîne une réduction de la variance compte tenu que des employés qui ne correspondaient pas aux attentes du poste (sur ou sous-qualifiés) ont quitté l'organisation et ne font pas partie de l'étude de validation (Arvey et Faley, 1992). Ainsi, les employés en place représentent un échantillon plus homogène que la population générale des candidats susceptibles de compléter l'instrument d'évaluation. Finalement, comme le rappelle Pettersen (2000) : « Forts de l'expérience acquise à l'emploi étudié, les employés en place pourront être avantagés avec tous les instruments de sélection qui mesurent des caractéristiques se développant avec l'apprentissage » (Pettersen, 2000, p. 99).

Or, lorsque des stratégies de validation peuvent difficilement être conduites dans le contexte spécifique du processus de sélection en cours, on se base souvent sur la validité de généralisation. En effet, depuis les dernières décennies, la recherche dans le domaine de la sélection de personnel a fait la démonstration qu'un instrument réputé valide dans une organisation pour un emploi l'est également dans une autre organisation pour un emploi similaire (Hunter et Hunter, 1984; Schmidt et Hunter, 1998). Ainsi, par la validité de généralisation et les méta-analyses, il s'agit d'appliquer les évidences reliées à la validité de certains prédicteurs obtenus par le biais des données recueillies dans diverses situations à d'autres situations similaires (Catano, 2010). Les stratégies de validation locales à petite échelle ont donc été remplacées par des séries de données agrégées, à l'intérieur desquelles les diverses sources d'erreur reliées aux petits échantillonnages ont été prises en considération (Bartram, 2004).

Malgré leur caractère parfois controversé, les méta-analyses sont largement utilisées et appliquées en recherche (Schmidt, Ones et Hunter, 1992). D'ailleurs, elles ont eu un impact majeur sur la pratique de la sélection de personnel et certains auteurs vont même jusqu'à parler d'une nouvelle ère en gestion des ressources humaines en stipulant qu'elles ont littéralement révolutionné les pratiques de validation (Pettersen, 2000). Cependant, une des limitations associées à cette portion de la littérature est que les données datent généralement de plusieurs années et on s'entend généralement pour dire qu'elles ne reflètent peut-être pas nécessairement l'étendue du rendement en emploi tel qu'il est aujourd'hui (Bartram, 2004).

Enfin, tenir compte et prévoir l'utilisation de plusieurs prédicteurs simultanément peut parfois donner des prévisions du rendement plus appréciables (Pettersen, 2000). Or, dans ce cas, il ne s'agit pas seulement de la corrélation entre le prédicteur et le rendement qui est étudiée (validité critériée), mais également si le prédicteur ajoute quelque chose à la validité des prédicteurs déjà présents (Pettersen, 2000). On parle alors de validité incrémentielle. En général, cette notion de validité permet de déterminer si la combinaison de certains instruments permet d'augmenter la valeur scientifique d'un processus de sélection en mettant l'accent sur la notion de complémentarité entre les divers éléments d'un ensemble.

#### 2.4.4 LA VALIDITE CRITERIEE DU TEST DE JUGEMENT SITUATIONNEL

---

Utilisé très souvent en sélection de personnel, plusieurs études ont tenté d'établir la validité critériée du test de jugement situationnel en lien avec le rendement futur en emploi. D'abord, il est intéressant de noter qu'une étude de Lievens et Patterson (2011) a démontré qu'un exercice de simulation avec un faible niveau représentativité (*low-fidelity simulation*), comme le TJS, n'obtenait pas de coefficients de validité critériée inférieurs à ceux obtenus par le biais d'exercices très représentatifs de l'emploi (*high fidelity simulation*). En fait, ces auteurs ont démontré que chacun de ces types de simulations se complète en ciblant chacun de l'information différente reliée positivement au rendement en emploi (Lievens et Patterson, 2011).

Cependant, malgré qu'il soit généralement admis que les résultats obtenus à un tel exercice soient corrélés positivement avec le rendement, la force de cette relation varie considérablement d'une étude à l'autre. D'aussi élevés que 0,56 (Stevens et Campion, 1999), d'autres coefficients de validité critériée de plus faible intensité ont également été publiés. Par exemple, alors que Weekley et Jones (1999) ont trouvé des coefficients de 0,23 pour leur étude 1 et de 0,16 pour l'étude 2, la recherche de Motowidlo et de ses collègues (1990) démontrait une relation variant de 0,28 à 0,37 entre l'évaluation du rendement par le superviseur et les résultats obtenus au TJS. Dans le même sens, Christian et ses collègues (2010) ont trouvé des coefficients de corrélation moyens de 0,25 lorsque le TJS avait été développé spécifiquement pour évaluer un construit relié à la qualité des relations interpersonnelles alors qu'une validité critériée de 0,19 a été trouvée lorsque l'instrument visait spécifiquement à mesurer les connaissances reliées au poste à doter (Christian et al., 2010).

Les résultats les plus régulièrement cités concernent une méta-analyse de 102 coefficients de validité réalisée par McDaniel et ses collègues (2001) qui reportait un coefficient moyen corrigé de 0,34 pour la capacité des TJS à prédire le rendement futur en emploi. Puis, plus récemment, une autre méta-analyse réalisée en 2007 avançait un coefficient de validité critériée de 0,26 lorsque corrigé (McDaniel et al., 2007). Selon ces auteurs, le fait que l'on observe un lien entre les habiletés cognitives et les résultats obtenus au TJS ( $p = 0,46$ ) pourrait, en partie, expliquer les raisons qui en font un bon prédicteur (McDaniel et al., 2001). En effet, comme de multiples études l'ont rapporté au cours des dernières décennies, la capacité de prédiction sur le rendement en emploi des mesures des habiletés cognitives générales est supérieure à n'importe quelle autre méthode (Schmidt et Hunter, 1998).

Par ailleurs, dans leur étude, McDaniel et ses collègues (2007) ont démontré que les TJS avaient une validité incrémentielle qui dépassait les prédictions généralement établies par le biais des tests de personnalité et des tests d'intelligence. Dans le même sens, d'autres auteurs ont également démontré que les TJS apportaient une contribution unique au-delà des données concernant la personnalité, l'habileté cognitive, la connaissance technique du travail et l'expérience de travail (Chan et Schmitt, 2002; Clevenger et al., 2001; T. V. Mumford, Van Iddekinge, Morgeson et al., 2008; O'Connell et al., 2007; Weekley et Jones, 1999). Ces auteurs ont également démontré qu'il ne s'agissait pas seulement d'un bon prédicteur du rendement global en emploi ou du rendement relié aux tâches, mais également d'une bonne façon de prédire le rendement contextuel (c'est-à-dire « *job dedication* » et « *interpersonal facilitation* ») (Chan et Schmitt, 2002). Dans tous les cas, ces résultats sont intéressants, mais nettement inférieurs aux coefficients de validité obtenus au moyen d'autres outils de mesure, tels que l'entrevue structurée ou les échantillons de travail pour lesquels les coefficients s'élèvent à 0,51 et 0,50 respectivement (Schmidt et Hunter, 1998).

Puisque l'amélioration de la valeur prédictive d'un test présente des avantages importants pour l'organisation tant sur le plan financier que pour l'efficacité de ses opérations, l'utilisation d'approches visant à accroître la validité prédictive des TJS représente donc une avenue à explorer. Comme le résume bien Schmidt et Hunter (1998) : « *the use of hiring methods with increased predictive validity leads to substantial increases in employee performance as measured in percentage increases in output, increased monetary value of output, and increased learning of job-related skills* » (Schmidt et Hunter, 1998, p. 262). Or, quels sont les facteurs qui influencent la variabilité des coefficients de validité critériée ? D'abord, les sources ayant permis de développer le contenu du test peuvent influencer sa

capacité à prédire le rendement en emploi. En effet, des auteurs ont démontré que les tests basés sur une analyse exhaustive de l'emploi ont démontré une validité plus élevée ( $p = 0,38$  vs.  $p = 0,29$ ) (McDaniel et al., 2001). D'autres ont avancé que la validité critériée moyenne des TJS variait selon que les directives fournies au candidat leur demandent d'indiquer « ce qu'il faut faire » ou « ce qu'il ferait » (McDaniel et al., 2007; Ployhart et Ehrhart, 2003). En effet, alors que les questions comprenant des directives de nature comportementale (would do) sont davantage corrélées avec les dimensions relatives à la personnalité, les questions avec des directives reliées aux connaissances (should do), sont davantage corrélées avec les habiletés cognitives. Par conséquent, selon McDaniel et ses collègues (2007), ces dernières démontrent une meilleure validité prédictive puisqu'elles représentent une mesure des connaissances de l'emploi, c'est-à-dire un très bon prédicteur du rendement en comparaison avec les caractéristiques de la personnalité. Finalement, des différences sur le plan méthodologique, tant pour la conceptualisation du critère que pour le développement du prédicteur, peuvent également apporter d'importantes variations sur le plan de la validité critériée des TJS (Lievens, Buyse et Sackett, 2005).

#### 2.4.5 IMPACT DE LA GRILLE DE CORRECTION SUR LA VALIDITE CRITERIEE

---

Dans cette perspective, il est clair que la stratégie adoptée par le concepteur afin d'identifier la bonne réponse et la manière de le corriger risque de conduire à des résultats forts différents concernant la capacité d'un TJS à prédire le rendement futur en emploi.

À ce sujet, les études qui se sont penchées sur l'efficacité des méthodes permettant d'établir les clés de correction ont d'abord démontré que chacune des trois approches permettait de prédire le rendement futur des candidats (Bergman et al., 2006; MacLane, Barton, Holloway-Lundy et al., 2001; Paullin et Hanson, 2001; Weekley et Jones, 1999).

Par la suite, d'autres études ont comparé entre elles différentes procédures de correction. Par exemple, MacLane et ses collègues (2001) ont comparé une clé de correction établie empiriquement par le biais d'une méthode corrélationnelle avec une autre établie par le biais des experts de contenus. Leurs résultats indiquent que les deux clés de correction apportent des niveaux de validité assez similaires (MacLane et al., 2001). Leurs conclusions rejoignent alors celles de Weekley et Jones (1999) à l'effet qu'il n'existe pas réellement de valeur ajoutée sur le plan de la validité critériée d'ajouter une méthode empirique aux approches rationnelles.

Par surcroît, Paullin et Hanson (2001) ont comparé un TJS corrigé à partir d'une approche rationnelle avec une série de clés de correction établies de manière empirique (relation verticale, horizontale, méthode corrélacionnelle et celle de la moyenne). Dans l'ensemble, ces auteurs n'ont trouvé aucune différence entre les diverses clés élaborées selon la méthode empirique et entre le fait de les avoir établies de manière rationnelle ou empirique (Paullin et Hanson, 2001).

Krokos et ses collègues (2004) ont également comparé cinq procédures de correction empiriques avec la méthode rationnelle. Leurs résultats ont démontré que seule la clé de correction établie de manière empirique par le biais d'une approche corrélacionnelle était significativement reliée au rendement en emploi, et ce, lorsque l'ensemble des clés empiriques ont été soumises à une contre-validation.

Enfin, Bergman et ses collègues (2006) ont comparé 11 différentes clés de correction pour un TJS mesurant le leadership. Leurs résultats ont démontré des variations considérables entre les différentes méthodes. Bien qu'il ait été impossible de conclure à la supériorité d'une approche sur une autre, ils ont trouvé des coefficients de validité plus élevés par le biais de la méthode hybride, c'est-à-dire en identifiant la bonne réponse de manière rationnelle et empiriquement.

Cependant, on a également observé que les corrélacions entre les clés de correction produites à partir de ces différentes méthodes sont habituellement faibles (Weekley et al., 2006). Ces études portent à croire que chacune des trois approches mesure un construit distinct ou différentes facettes du rendement. Il est également possible qu'il existe plusieurs façons de répondre adéquatement à des situations et que différentes combinaisons de réponses permettent de prédire le rendement en emploi.

---

## 2.5 LE CONCEPT DE FIDELITE

---

Dans une perspective très large, la validité d'un processus de sélection et l'interprétation que l'on fait des résultats n'ont rien à offrir si les informations qui ont servi à prendre une décision ne sont pas fiables ou fidèles, tant sur le plan du prédicteur que pour le critère (Gatewood et al., 2008). En effet, pour être utile, une mesure doit d'abord être fiable, exacte et digne de confiance (Pettersen, 2000).

Dans le pire des scénarios, les procédures d'évaluation et de sélection sont tellement peu fidèles, qu'elles n'apportent peu ou pas d'information sur la prédiction du rendement futur en emploi (Cook, 2009).

De façon plus spécifique, le concept de fidélité désigne le degré de constance ou la stabilité des résultats entre deux séries de résultats de la même mesure (Guion, 1965; Pettersen, 2000). À cet égard, depuis des décennies, la définition opérationnelle de la notion de fidélité a peu évolué dans la recherche en sélection de personnel et elle est demeurée orientée sur l'évaluation de l'erreur de mesure pour estimer l'exactitude ou la précision des résultats (Putka et Sackett, 2010). En somme, selon l'approche classique généralement adoptée en psychométrie, une mesure fidèle est une mesure sans erreur aléatoire et l'évaluation de la fidélité d'un instrument devient alors la détermination du degré de variation dû aux sources d'erreurs aléatoires (Guion, 1965). Pour évaluer la fidélité des résultats obtenus à un instrument de mesure et s'assurer qu'une mesure est exacte, on les compare habituellement à d'autres mesures indépendantes du même objet. Si les résultats sont semblables, on conclut que la mesure est fiable et s'il y a des différences, alors il y a de l'erreur (Pettersen, 2000). En gestion de ressources humaines, ces erreurs de mesure qui affectent un instrument de mesure proviennent de multiples sources, soit (Pettersen, 2000) :

- Du candidat lui-même qui n'est pas toujours constant dans sa façon de répondre à un instrument de mesure ;
- De l'examineur dont le comportement est aussi influencé par plusieurs facteurs ;
- De la situation ou des conditions dans lesquelles est utilisé l'instrument.

Dans cette perspective, la littérature en psychométrie a traditionnellement organisé les discussions sur l'estimation « directe » de la fidélité d'un instrument par catégorie ou par type de fidélité, et ce, par la détermination de l'erreur de mesure (Putka et Sackett, 2010). Malgré tout, d'autres techniques modernes de calcul de la fidélité ont été proposées par la communauté scientifique pour tenter de comprendre les composantes de la fidélité et des sources d'erreur. C'est le cas notamment de la théorie de la généralisabilité – *G-Theory* (Putka et Sackett, 2010). En résumé, cette approche renouvelée de la fidélité repose sur une méthode statistique d'analyse de la variance et représente plutôt la fidélité comme un continuum général. Cependant, ce nouveau paradigme n'a pas encore trouvé sa place dans le domaine de la recherche en sélection de personnel (Putka et Sackett, 2010).

Ainsi, les techniques d'estimation de la fidélité présentés aux pages suivantes permettent chacune d'étudier un aspect particulier de l'instrument et reposent directement ou indirectement sur l'application de méthodes corrélationnelles (Cook, 2009).

D'une part, un instrument de mesure doit permettre à un observateur d'établir des mesures similaires d'un même sujet à des moments différents, toujours avec le même instrument. Il s'agit de la fidélité test-retest (Cook, 2009). Cette méthode pour évaluer la fidélité d'un instrument de mesure consiste à faire passer aux mêmes personnes deux fois le même instrument à des moments différents et d'évaluer dans quelle mesure les résultats de la deuxième passation sont reliés à ceux de la première. Le degré de fidélité d'un test peut donc être mesuré en vérifiant la stabilité d'un test à reproduire les mêmes résultats sur une période de temps donnée (Pettersen, 2000).

D'autre part, la méthode des formes parallèles ou équivalentes est similaire à celle de la méthode du test-retest sauf que c'est une autre version de l'instrument de mesure qui est administrée lors de la deuxième passation. De façon plus précise, il s'agit d'administrer deux tests aux mêmes individus, la différence étant que le second test n'est pas identique au premier : il s'agit d'une forme alternative élaborée pour mesurer la même caractéristique, mais selon une formulation différente et de déterminer dans quelle proportion les résultats des personnes candidates sur les mêmes caractéristiques sont convergents entre eux (Gatewood et al., 2008).

Par ailleurs, de nombreux outils utilisés en gestion de ressources humaines peuvent nécessiter la participation de plusieurs examinateurs ou évaluateurs pour observer et formuler une appréciation des personnes. Comme le jugement de l'évaluateur joue un rôle important dans ce contexte, un outil visant la mesure de certaines caractéristiques de la personne doit permettre à des observateurs différents de faire des mesures concordantes d'un même sujet avec le même instrument. Pour ce faire, l'influence de l'examineur peut être évaluée en comparant deux ou plusieurs examinateurs lorsqu'ils évaluent la même personne et s'il y a concordance entre les examinateurs, les résultats sont fidèles. Pour calculer la fidélité par ce coefficient de corrélation, il s'agit donc de demander à des évaluateurs indépendants de corriger ou de porter une appréciation sur certaines caractéristiques spécifiques à partir du même bassin de personnes candidates et de déterminer dans quelle proportion les résultats convergent entre eux.

Finalement, les diverses parties d'un instrument de mesure doivent permettre de mesurer la même chose. Il s'agit de la fidélité par consistance interne (Guion, 1965). À partir de cette méthode, il s'agit en fait d'examiner les résultats obtenus par les mêmes personnes entre des parties équivalentes de l'examen à travers une seule passation. Plus les réponses des candidats à une partie de l'instrument sont reliées à celles des autres parties, plus les résultats sont fidèles (Catano, 2010). On utilise généralement le coefficient d'alpha qui examine la contribution du test en entier au score total afin d'évaluer l'homogénéité d'un instrument de mesure composé d'un ensemble de questions qui, ensemble, devraient contribuer à appréhender une même dimension (Cook, 2009). Cependant, comme le rappellent Putka et Sackett (2010), selon la composition du construit que l'on cherche à mesurer, un résultat « vrai », du point de vue de la fidélité pourrait également être considéré comme une erreur du point de vue de la validité. En effet, ce que l'on appelle généralement le « dilemme étendue-précision » survient lorsque l'on sacrifie certaines dimensions pertinentes à l'emploi lors de l'évaluation (impact sur la validité) au profit d'une évaluation plus approfondie de certaines dimensions circonscrites ou en retirant les items qui ne corrélerent pas avec les autres au profit d'autres items plus homogènes (impact sur la fidélité) (Pettersen, 2000).

En conclusion, une meilleure connaissance des informations relatives à la fidélité d'un instrument de mesure contribue à identifier les sources possibles de biais introduit dans le processus de sélection de personnel. À cet égard, Gatewood (2008, p. 157) conclut que : « *unreliability and instability of test scores are the most fundamental forms of test bias* ». Toutefois, à la question concernant le seuil minimal de fidélité acceptable pour un instrument de mesure utilisé en sélection de personnel, les recommandations peuvent varier selon la méthode d'estimation utilisée (Gatewood et al., 2008).

### 2.5.1 LA FIDELITE DU TEST DE JUGEMENT SITUATIONNEL

---

Comme les TJS sont corrigés de manière mécanique, ils bénéficient du mérite de la standardisation sur la fidélité des résultats obtenus chez les candidats en y éliminant les erreurs aléatoires dues à l'interprétation et au jugement des correcteurs. En effet, plus une approche de correction est standardisée, plus elle contribue à uniformiser la correction entre les correcteurs (accord interexamineurs) et d'un candidat à l'autre (fidélité intrajuge) (Pettersen, 2000).



Par contre, sur un autre volet, tel que mentionné précédemment, les TJS sont la plupart du temps conçus de manière multidimensionnelle (McDaniel et al., 2007), c'est-à-dire pour évaluer une variété de construits hétérogènes, ce qui mine les estimations de leur consistance interne. À cet effet, dans leur méta analyse, McDaniel et ses collègues (2001) ont publié des coefficients de consistance interne variant entre 0,43 et 0,94. Incidemment, plusieurs facteurs influencent la consistance interne d'un TJS et agissent de manière différente d'une étude à l'autre. Par exemple, le nombre de questions qui composent le test agit sur la fidélité du TJS ; plus le test est long plus les résultats seront fidèles en réduisant la probabilité d'un succès aléatoire. Dans le même sens, Ployhart et Ehrhart (2003) ont démontré que le format des directives soumis aux candidats influençait aussi la consistance interne du test. À cet égard, lorsque les candidats doivent évaluer l'efficacité de chacune des options de réponses, généralement la consistance interne au test est plus élevée (0,73). Par ailleurs, lorsque les candidats doivent identifier deux options de réponses, soit la meilleure et la pire des options, la consistance interne étaient relativement inférieure, soit de l'ordre de 0,60. En terminant, selon eux, lorsque les candidats doivent identifier une seule bonne réponse, ce type de directive mène à l'observation de la consistance interne la plus faible, c'est-à-dire 0,24 (Ployhart et Ehrhart, 2003).

Cette faible consistance interne et le fait que les TJS soient de nature multidimensionnelle sont aussi mis en évidence par le biais d'analyses factorielles qui révèlent des liens avec une multitude de facteurs difficilement interprétables (Schmitt et Chan, 2006). De cette manière, cela met en doute le fait qu'une mesure de consistance interne est adéquate pour évaluer la fidélité d'un TJS (Whetzel et McDaniel, 2009). En fait, utiliser cette mesure de la fidélité en l'appliquant aux TJS conduirait à en sous-estimer la valeur réelle (Catano, Brochu et Lamerson, 2012). Dans l'ensemble, afin d'évaluer la fidélité du test, on recommande alors habituellement la méthode du test-retest ou des formes parallèles (Catano et al., 2012; Lievens et al., 2008; O'Connell et al., 2007; Whetzel et McDaniel, 2009).

---

## 2.6 SES PRINCIPALES LIMITES

---

Malgré les avantages indéniables discutés précédemment concernant l'utilisation des TJS en sélection du personnel, certaines limites doivent également être considérées. En premier lieu, avec ce type de simulation, l'objet de la mesure est le comportement projeté plutôt que la description du comportement réellement adopté en regard aux situations de travail présentées. Dans cette perspective,

alors que certains auteurs se sont interrogés sur le risque de tricherie de la part des candidats (« *faking* ») lors de l'utilisation des TJS, la littérature sur les inventaires de personnalité renseigne également sur les biais de réponse systématiques qui entrent en ligne de compte lors de la passation d'un test à mesure projeté (Nhung T. Nguyen et al., 2005; Peeters et Lievens, 2005).

En fait, il est fort probable que les TJS soient beaucoup plus sujets à la falsification que les centres d'évaluation, par exemple, pour lesquels aucune liste de scénarios d'actions possibles n'est proposée aux candidats. Dans le même sens, les candidats peuvent avoir tendance à adopter des comportements qu'ils croient représentatifs des attentes de l'organisation sans pour autant avoir l'intention de les adopter en cours d'emploi. Or, la complexité des situations et leurs procédures de correction rendent la falsification des TJS beaucoup plus difficile si l'on compare, notamment, avec les inventaires de personnalité (Hooper, Cullen et Sackett, 2006). En effet, les énoncés des inventaires de personnalité sont généralement transparents, au point où il pourrait être facile pour les candidats de biaiser leurs réponses en fonction des enjeux en cause. Cependant, comme chaque question d'un TJS ne vise pas la mesure d'une échelle spécifique et qu'elles comprennent souvent plusieurs éléments de réponses, le traitement de cette grande quantité d'informations est un processus cognitif trop complexe pour que les candidats puissent y répondre de façon constante dans un sens qu'ils croient souhaité par l'organisation (Hooper et al., 2006). D'ailleurs, il n'est pas certain que cette falsification soit toujours à l'avantage du candidat. Prenons l'exemple d'un candidat qui postule un emploi routinier et qui se montre très ouvert aux changements afin de paraître ce qu'il croit être préférable aux yeux de l'employeur. C'est pourquoi on recommande généralement de mettre en œuvre des stratégies préventives et des mises en garde aux candidats au cours de la passation quant à l'importance de répondre de façon honnête (Boudrias et Morin, 2011).

Dans le même sens, il est essentiel de faire une distinction entre l'action consciente et délibérée de falsifier un examen et le concept de « *self-presentation* » inhérent aux tests de personnalité et aux TJS comprenant des directives de type comportemental (*would do*) (R. Hogan, Hogan et Roberts, 1996). Ce phénomène, que l'on désigne habituellement par la notion de « désirabilité sociale » constitue une tendance de la part des candidats à donner des réponses qui sont jugées plus acceptables par une grande partie de la société afin de faire bonne impression ou de se mettre en valeur (Ones, Viswesvaran et Reiss, 1996). À cet effet, selon Paulhus (1988) il est possible de conceptualiser le phénomène de désirabilité sociale à partir de deux facteurs, soit les phénomènes de « *self-deception* » et de

« *impression management* ». Alors que le premier se réfère à la tendance d'un individu à penser à lui-même dans des termes positifs, le deuxième fait référence au fait d'agir de façon à générer une impression favorable chez les autres (Paulhus, 1988). Plus spécifiquement, comme le rappellent Schmitt et Oswald (2006) : « *A self-deception factor was considered a "normal" and nondeliberate tendency to present oneself positively, whereas impression management was a deliberate attempt to present oneself in a particular manner to achieve some desirable outcome, such as a job offer in a personnel selection context* » (Schmitt et Oswald, 2006, p. 613).

Or, la recherche à ce sujet démontre généralement qu'elle n'atténue pas la validité des mesures de la personnalité (Barrick et Mount, 1996; Leietta M. Hough, Eaton, Dunnette et al., 1990; Ones, Viswesvaran et Schmidt, 1993). En effet, certains auteurs ont démontré que la désirabilité sociale est fortement reliée à des traits de personnalité comme l'adaptation, l'aspect consciencieux, la stabilité émotionnelle et d'autres critères considérés « désirables » de la part d'une organisation (Ones et al., 1996). Par ailleurs, malgré le fait que certaines études ont démontré que les candidats peuvent augmenter leurs résultats obtenus à un inventaire de personnalité lorsqu'ils ont reçu la consigne de tricher ou que la liste des candidats choisis suite à un processus de sélection peut être affectée par la tricherie, il n'est pas certain qu'un candidat participant à un processus de sélection aurait été réellement porté à le faire.

En ce qui concerne les TJS, leur falsification pourrait dépendre du type d'instruction privilégié dans le test (N. T. Nguyen, McDaniel et Biderman, 2002). À cet effet, Nguyen et ses collègues (2002) ont démontré que les directives de type « quelle est la meilleure réponse » seraient beaucoup plus résistantes à la tricherie en comparaison avec les directives de type « que feriez-vous ». Selon eux, la tricherie serait donc affectée par le construit mesuré par le TJS ; certains construits sont plus faciles à « simuler » que d'autres (N. T. Nguyen et al., 2002). Ainsi, toujours Nguyen et ses collègues (2002), un TJS devient plus résistant à la tricherie lorsque les corrélations de l'outil sont très élevées avec le facteur de l'intelligence générale en comparaison avec un TJS qui est fortement corrélé avec les traits de personnalité deviendrait plus transparent. Par contre, même si les TJS avec des directives de nature comportementale sont davantage sujets à la désirabilité sociale en comparaison avec les questions qui demandent aux candidats de démontrer leurs connaissances, l'effet de cette « distorsion » sur les TJS n'est pas clair à ce jour (Nhong T. Nguyen et al., 2005). Par ailleurs, dans le cadre d'un processus de sélection dont l'issue comporte un enjeu réel pour le candidat (*high-stake testing*), il est peu probable que ce dernier choisisse

une option de réponse de nature comportementale qu'il ne considère pas comme étant la « meilleure » réponse. Il s'agit de la conclusion de l'étude de Lievens, Sackett et Buyse (2009) qui n'ont trouvé aucune différence dans les résultats en fonction du type de directive (comportementale vs. connaissance) lorsque les TJS étaient administrés exclusivement dans un contexte comportant un enjeu réel pour les participants.

Dans cette même perspective, il est possible de s'inspirer des conclusions de Hough et Oswald (2008) relativement aux inventaires de personnalité à l'effet que le phénomène de la désirabilité sociale dépend généralement de plusieurs facteurs, notamment : « (a) *item transparency (i.e., subtle vs. obvious items)*, (b) *research setting (i.e., experimental vs. real-life applicant selection setting)*, and (c) *research design (i.e., concurrent vs. predictive – longitudinal – design)* »<sup>10</sup>. Dès lors, il est possible de contrôler en partie cet enjeu de falsification en lui accordant une attention soutenue lors de l'élaboration du test et par le biais d'analyses statistiques (par exemple, éliminer les items dont les moyennes sont très élevées, établir un paramétrage pour les contextes avec enjeux, rédiger des options de réponse pour lesquelles il n'y a pas de bonne ou de mauvaise réponse évidente, etc.) (Lievens et al., 2009).

D'un autre regard, lorsque l'on s'intéresse à l'évaluation des différences individuelles, il est implicitement reconnu que l'on peut généraliser les caractéristiques d'un individu à partir de ses réponses sur sa propension à adopter ce type de comportement ou à se sentir d'une telle manière dans une situation particulière à n'importe quel type de situation (R. Hogan, Johnson et Briggs, 1997). Cependant, comme le suggèrent certains auteurs, cela n'est peut-être pas le cas pour toutes les personnes candidates alors que certaines peuvent adopter un cadre de référence spécifique pour répondre aux questions d'un test (Schmit, Ryan, Stierwalt et al., 1995). Par exemple, lors de l'administration d'un test de personnalité, aux questions « J'aime régulièrement participer à des activités sociales » ou « Je deviens très émotive lorsqu'une personne de mon entourage vit des difficultés », le fait de répondre en fonction de son cadre de vie personnelle ou professionnelle pourrait amener une personne candidate à des résultats plus ou moins contradictoires. Ainsi, une autre limite de l'utilisation des TJS en contexte de sélection fait référence au problème du cadre de référence des répondants (« *frame-of-reference problem* ») (Mount, Barrick et Strauss, 1994). À ce sujet, la recherche sur les tests de personnalité a démontré qu'un inventaire psychométrique contextualisé par rapport à des situations

---

<sup>10</sup> Tiré de Hough, L. et Dilchert, S. (2010). Personality: Its measurement and validity for employee selection. Dans J. L. F. N. T. Tippins (Éd.), *Handbook of employee selection* (pp. 299-319). New York, NY, US: Routledge/Taylor & Francis Group.

reliées au monde du travail et présentant un cadre de référence spécifique permet habituellement d'améliorer la validité des mesures de la personnalité (Schmit et al., 1995). Selon cette perspective, comme les comportements des individus pourraient être généralement conditionnés par la situation, la prédiction du comportement futur en emploi pourrait être améliorée lorsque les personnes candidates reçoivent un cadre de référence étroitement lié à la réalité de l'emploi.

Par ailleurs, de manière opérationnelle, il existe également un enjeu de sécurité lors de l'adoption d'un TJS dans un processus de sélection. En effet, il est facile d'imaginer les montants qui doivent être déboursés par une organisation lorsque son instrument de sélection est copié, volé ou redistribué par les candidats. Dans cette perspective, puisque les TJS sont généralement accessibles via des plateformes technologiques permettant le testing à distance, les organisations vont généralement produire différentes versions parallèles du même TJS afin de contrôler dans une certaine mesure les risques associées à la confidentialité et à la sécurité du contenu du test (Ployhart et MacKenzie Jr., 2011). Cependant, compte tenu du caractère multidimensionnel des TJS, il s'avère ardu de reproduire différentes versions présentant des caractéristiques psychométriques similaires (Lievens et Sackett, 2007). À ce sujet, Lievens et Sackett (2007, p. 1050) ont noté dans une de leurs études que « *the poor alternate-form consistency results of the random assignment approach revealed that the domain sampled by an SJT should not be considered to be the only radical (i.e., a major determining characteristic) of SJTs* ». Pour ce faire, ces auteurs ont proposé un cadre conceptuel pour la création de différentes versions d'un test qui s'oriente davantage sur les caractéristiques des questions qui les composent (c'est-à-dire leur contexte, l'incident critique utilisé pour leur développement, etc.) (Lievens et Sackett, 2007).

En outre, malgré qu'il soit généralement reconnu que les TJS présentent des différences de moyennes modérées dans les scores obtenus par certains groupes visés par les programme d'égalité en emploi, les raisons qui sous-tendent ce phénomène et celles qui font varier les différences observées entre certaines populations demeurent floues à ce jour (Lievens, 2006; Ployhart et MacKenzie Jr., 2011). En effet, est-ce que cela peut être attribuable au fait que cet instrument présente une plus faible consistance interne ou à un autre facteur qui n'a pas encore été identifié (Ployhart et MacKenzie Jr., 2011) ? Selon ces auteurs, il est essentiel d'approfondir l'ensemble de la recherche portant sur le facteur culturel lors de l'utilisation des TJS étant donné que plusieurs questions demeurent sans réponse : dans quelle mesure est-il possible de généraliser l'utilisation d'un TJS à travers différentes cultures ? quelle est

la validité interculturelle d'un TJS ? quel est l'effet de la culture sur la définition des critères de rendement attendus ? (Lievens, 2006) Dans un contexte d'internationalisation des affaires, ces enjeux deviennent incontournables pour les organisations qui doivent s'assurer que leurs procédures de sélection puissent être utilisées dans plusieurs pays et complétées par des participants provenant d'origines culturelles diverses.

### 3. LE RENDEMENT EN EMPLOI

---

Comme il l'a été exposé précédemment, lorsqu'une organisation s'attarde à aligner les différentes composantes de son système de GRH de façon stratégique, elle peut en retour influencer la composition de son capital humain et contribuer à solliciter les comportements reliés à un bon rendement en emploi (Wright et al., 2001). Dans le même sens, étant donné l'ampleur des différences individuelles sur le plan du rendement, il importe donc pour une organisation d'acquérir un bassin de capital humain dont le rendement satisfera minimalement aux exigences normales des emplois à combler (Pettersen, 2000). De cette manière, le rendement en emploi, que l'on désigne également par la notion de « performance » individuelle, se trouve au cœur des problématiques de gestion des ressources humaines et de management des organisations. Il s'agit sans doute de la variable dépendante la plus étudiée dans ces disciplines (Kahya, 2007).

Dans cette perspective, comme le rendement d'une personne demeure inconnu au moment de la mise en œuvre d'un processus de sélection, la solution la plus pratique pour les responsables des ressources humaines est de tenter de le prédire par le biais d'instruments, tels que le TJS. Ainsi, la validité et l'utilité d'un système de sélection sont basées sur la façon dont ce dernier a réussi à prédire le rendement futur en emploi. Cependant, tel qu'il l'a été discuté plus tôt, la valeur scientifique d'une étude portant sur l'efficacité d'une procédure de sélection dépend en grande partie de la qualité de la mesure associée au rendement.

---

### 3.1 LA DEFINITION DU RENDEMENT

---

Paradoxalement à sa popularité dans les études touchant au monde du travail, le rendement comme objet de recherche demeure peu conceptualisé et l'on s'arrête généralement au caractère opérationnel de la mesure plutôt qu'à la définition du concept (St-Onge et Haines, 2007). D'ailleurs, bien des auteurs vont utiliser les expressions « performance » et « rendement » de façon interchangeable dans la vie de tous les jours alors que certains vont privilégier l'expression « performance » pour son caractère plus englobant qui prend également en compte les comportements (St-Onge et Haines, 2007). Cependant, sur le plan de la langue française, le terme performance fait référence davantage à la notion d'exploits sportifs qui demeure peu pertinente au monde du travail. À cet égard, selon Foucher (2007), la pratique qui semble la plus courante en gestion est celle d'utiliser le terme rendement pour désigner les résultats au niveau individuel et celui de performance pour nommer ceux qui se situent au niveau organisationnel. Ainsi, la définition de Foucher présentée dans l'ouvrage de St-Onge et Haines (2007) à l'effet que le rendement « désigne un résultat obtenu et en corollaire, le degré de réussite que représente ce résultat (...) évalué en fonction de divers critères (...) en prenant en compte les moyens mis en œuvre » (p. 67) est utile. Cependant, les deux modèles de rendement individuel les plus cités dans la documentation scientifique (Campbell, McCloy, Oppler et al., 1993; Motowildo et al., 1997) définissent le rendement en termes de comportements plutôt qu'en fonction des résultats puisque les résultats sont souvent affectés par des facteurs non contrôlés par l'individu (Campbell, Gasser et Oswald, 1996; Murphy et Cleveland, 1995). Selon ces modèles, le rendement ne doit donc pas être confondu avec la productivité (ratio d'extrants divisé par le coût de réalisation, soit les intrants) (Murphy et Cleveland, 1995). La définition de Campbell et ses collègues (1993) qui définissent le rendement comme « les comportements qui sont pertinents pour les objectifs de l'organisation et qui peuvent être mesurés en terme de niveau de contribution individuelle » (p.40) sera donc retenue pour la suite de la recherche.

Le rendement en emploi est effectivement un concept complexe comportant plusieurs dimensions. En fait, il est généralement abordé dans le contexte des tâches qu'une personne doit accomplir pour contribuer aux objectifs de l'organisation. C'est ce que la littérature dans la pratique de l'ergonomie appelle le « travail prescrit », la référence théorique formalisée à laquelle l'entreprise rapporte la performance des travailleurs (Daniellou, Grall, Martin et al., 2000). Comme Daniellou et ses collègues (2000) le présentent, le travail prescrit appartient à tout ce qui est défini objectivement par

avance par l'entreprise et donné à l'individu pour organiser et réaliser son travail : objectifs à atteindre, consignes à respecter, processus et procédures à mettre en œuvre, etc.

Or, il est impossible de faire abstraction des comportements qu'une personne va adopter au travail et qui influencent directement son rendement. C'est pourquoi ce même volet de littérature relève la différence entre le travail prescrit et le travail réel. Le travail réel, c'est le travail tel qui se réalise effectivement au quotidien dans chaque situation particulière. Travailler c'est bien plus que respecter des consignes : c'est confronter plusieurs sources de prescriptions (en provenance des collègues, des clients ou des superviseurs, par exemple), établir des priorités, trier entre elles et parfois ne pas pouvoir les satisfaire toutes tout le temps (Daniellou et al., 2000). Ainsi, pour que l'individu travaille et offre un bon rendement, il doit nécessairement y avoir une certaine forme de mobilisation d'engagement ou de motivation.

En ce sens, on s'entend généralement pour dire que les comportements d'une personne sont distincts de ses caractéristiques individuelles (KSAOs) qui représentent le capital humain d'une organisation (Wright et al., 2001). En fait, le travailleur possède les KSAOs et non l'organisation, ce qui fait que ce dernier peut toujours utiliser de son « pouvoir » discrétionnaire d'agir d'une telle ou telle façon et d'avoir des répercussions positives ou négatives sur l'organisation. À cet effet, il est clair que la motivation peut en retour diriger le comportement. Plusieurs chercheurs dans différents domaines ont d'ailleurs élaboré des théories afin de prédire et d'expliquer les comportements humains à partir du concept générique de motivation. Sans décrire chacune de ces théories qui ne sont pas centrales à la présente recherche, mentionnons seulement que plusieurs éléments variant en intensité jouent un rôle important et entrent souvent en compétition lors de la prise de décision de l'individu d'adopter les comportements que l'on attend de lui en emploi (Vallerand et Thill, 1993).

Ceci étant dit, selon St-Onge et Haines (2007), les auteurs s'entendent pour dire qu'il existe deux grandes catégories de comportements démontrés par les employés et liés au rendement individuel qui vont en retour contribuer à l'efficacité organisationnelle. Il semble que cette conceptualisation du rendement soit devenue la plus courante dans la documentation scientifique (St-Onge et Haines, 2007). À cet égard, Borman et Motowildo (1993) ont fait la distinction entre le rendement centré sur la tâche et le rendement contextuel. D'une part, le premier type de rendement fait référence à des comportements qui sont impliqués directement dans la production de biens, de services et d'activités tout en fournissant



un soutien indirect aux processus techniques de l'organisation. En général, ce type de rendement s'observe par l'atteinte des objectifs fixés et porte généralement sur l'efficacité et la qualité de la réalisation des tâches comprises dans une description de travail.

D'autre part, bien qu'il dépasse souvent le cadre normal des fonctions professionnelles d'un individu et qu'ils ont souvent un caractère volontaire, le rendement contextuel est aussi important puisque les comportements qui s'y rattachent permettent de définir le contexte social de l'organisation et ils agissent de façon globale sur le fonctionnement de l'organisation. Certains auteurs vont même jusqu'à dire que ces comportements peuvent expliquer entre 12% et 19% de la variance dans le jugement qu'un superviseur exprime quant au rendement d'une personne, leur influence étant encore plus importante lorsque les mesures quantitatives liées à des résultats ne sont pas disponibles ou ne sont pas valorisées (St-Onge et Haines, 2007). Dans la communauté scientifique, ce rendement contextuel a été identifié par plusieurs expressions : comportements de citoyenneté organisationnelle, comportements prosociaux, participation organisationnelle, etc. Alors que les comportements prosociaux visent le bien-être d'un individu ou d'un groupe de personnes, les comportements de citoyenneté organisationnelle peuvent être identifiés par le biais des dimensions suivantes : l'altruisme, la courtoisie, l'esprit sportif, la vertu civique et l'éthique et la conscience (Organ, 1997). Il s'agit d'ailleurs d'un courant de recherche important pour quiconque s'intéresse au concept de mobilisation en ressources humaines, c'est-à-dire « une masse critique d'employés qui accomplissent des actions (faisant partie ou non de leur contrat de travail, rémunérées ou non) bénéfiques au bien-être des autres, de leur organisation et à l'accomplissement d'une œuvre collective » (Tremblay et Wills, 2005, p. 38) et qui permet d'autant plus de distinguer la mobilisation de la motivation individuelle.

En bref, le rendement contextuel produit généralement des résultats favorables pour l'organisation et ils ne sont pas imposés par le contrat de travail. C'est dans une optique d'intégration que Borman et Motowidlo (1993) ont proposée une taxonomie du rendement contextuel en cinq dimensions : (1) être volontaire pour effectuer des tâches qui ne sont pas formellement requises par l'analyse d'emploi ; (2) persister avec enthousiasme et efforts supplémentaires lorsque requis pour mener à bien des activités au travail ; (3) aider et coopérer avec les autres ; (4) respecter les règles et les procédures organisationnelles même lorsqu'elles entraînent un inconvénient pour l'individu lui-même ; et (5) endosser, soutenir et défendre les objectifs organisationnels.

Certains autres chercheurs se sont intéressés au concept du rendement comme étant relié principalement à l'adaptation au travail (*adaptive job performance*) et qui consiste spécifiquement à examiner la façon dont les individus font face à de nouvelles situations ou exigences en milieu de travail afin de qualifier leur rendement (Pulakos, Arad, Donovan et al., 2000). En effet, puisque l'environnement des travailleurs est constamment en changement, ces derniers doivent démontrer un haut niveau et d'adaptation et de tolérance face à l'ambiguïté afin de continuer à y naviguer efficacement. Conséquemment, dans le cadre d'une vaste étude, Pulakos et ses collègues (2000) ont tenté de définir ces notions d'adaptation et de flexibilité en lien avec le rendement démontré par une ressource au travail. Au final, ils en sont arrivés à proposer une taxonomie déclinée en huit (8) dimensions théorique du rendement reliées au concept du « *adaptive job performance* », c'est-à-dire : « (a) la capacité à gérer des situations d'urgence et de crise, (b) la capacité à gérer le stress (c) la résolution de problèmes avec créativité, (d) la capacité à faire face à des situations de travail imprévues et incertaines, (e) la capacité à apprendre de nouvelles tâches, technologies et procédures, (f) la capacité à s'adapter sur le plan interpersonnel (g) la capacité à s'adapter culturellement, et (h) une capacité à s'adapter physiquement » (Pulakos et al., 2000, pp. 617, traduction libre).

Par ailleurs, comme il s'agit de l'ensemble des mêmes caractéristiques d'une personne (capacités, dispositions personnelles, etc.) qui finissent par influencer de la même manière toutes les dimensions reliées au rendement, selon Viswesvaran, Schmidt et Ones (2005), il existerait un facteur général du rendement qui irait au-delà du rendement axé sur les tâches et du rendement contextuel pris individuellement. En fait, à partir d'une vaste méta-analyse, ils ont confirmé l'hypothèse qu'il existait un construit spécifique relié au rendement général emploi qui compte pour 60% du total de la variance lorsque l'on procède à l'évaluation du rendement au travail d'une personne. Dans ce contexte, il s'avère intéressant de penser qu'une étude qui s'intéresse au rendement en emploi pourrait également s'orienter vers l'évaluation de ce construit spécifique, soit le rendement global en emploi.

Finalement, comme expliquée plus tôt, une étude portant sur la validité d'un processus de sélection de personnel débute avec l'identification des dimensions reliées au rendement par l'analyse de l'emploi. Dans cette perspective, les paramètres fixés qui indiquent le succès ou l'échec en emploi seront généralement reliés aux tâches. Dans le même sens, on s'intéressera à la validation des KSAOs qui contribuent au rendement en emploi. La mesure du rendement en emploi devient alors associée à un critère, c'est-à-dire un échantillon de mesures qui permettent d'identifier les différences individuelles

entre les employés en regard des comportements requis dans l’emploi pour offrir un rendement adéquat. À cet égard, les travaux de Campbell et ses collègues (Campbell, Dunnette et Hough, 1990; Campbell, McHenry et Wise, 1990) proposent une structure des dimensions du rendement au travail composée de huit dimensions génériques qui devient très utile pour quiconque veut le mesurer (voir tableau 6).

**Tableau 6**  
**Dimensions génériques du rendement**  
**(Campbell, Dunnette, et al., 1990)<sup>11</sup>**

Effacité dans les tâches techniques spécifiques et au cœur du poste
Effacité dans les tâches générales et non spécifiques à un poste
Effacité dans les communications écrites et orales
Effort soutenu et approprié en toute circonstance
Discipline personnelle
Appui et aide au travail en équipe et à celui des pairs
Supervision et leadership auprès des subordonnés au moyen de comportements d’influence et d’interactions interpersonnelles
Effacité dans les tâches administratives distinctes de la supervision

---

### 3.2 LA MESURE DU CRITERE

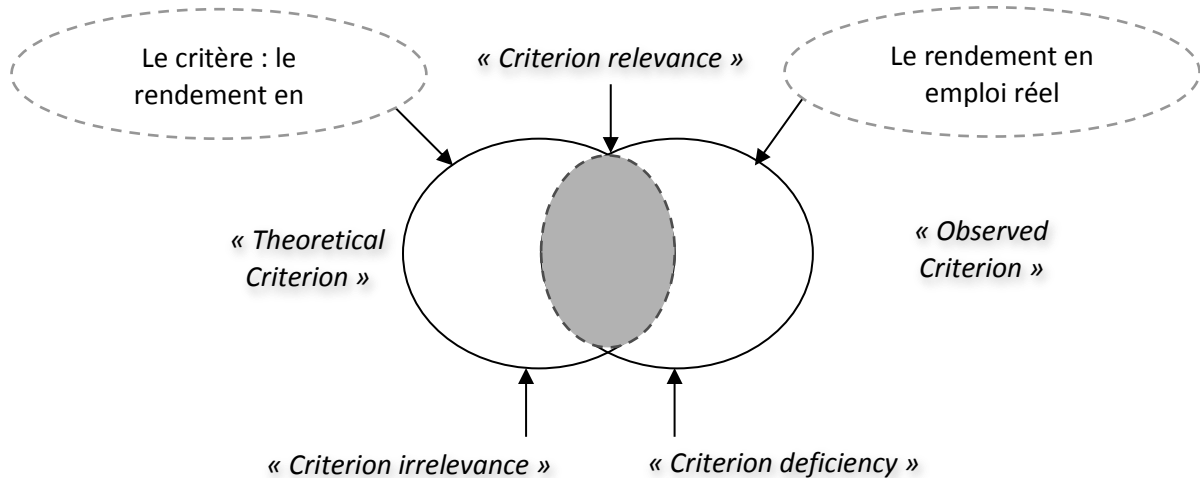
---

Encore ici, on s’inspire de la psychométrie qui vise la valorisation et l’amélioration de la précision en stipulant qu’une mesure du rendement reflète une réalité objective ou réelle et qu’il est possible d’adopter des mesures valides et pertinentes des comportements et des résultats au travail (St-Onge et Haines, 2007). Cependant, la mesure du rendement est beaucoup plus facilement dite que faite. En effet, il n’est pas aisé de choisir des mesures qui vont réussir à capter l’essence totale et la complexité du concept de rendement. En effet, bien que l’on s’intéresse généralement aux KSAOs en sélection de personnel, la relation risque de demeurer toujours assez faible si on n’inclut pas d’autres mesures qui représentent mieux l’étendue du critère (par exemple, les comportements associés au rendement contextuel). Sur le plan de la validité de la mesure du critère, les travaux de Campbell ont permis de schématiser les complexités reliées à la mesure du critère.

---

<sup>11</sup> Tiré de l’ouvrage de Pettersen, N. (2000). Évaluation du potentiel humain dans les organisations élaboration et validation d’instruments de mesure. Sainte-Foy : Presses de l’Université du Québec, page 240.

**Figure 7**  
**Caractéristiques du critère**  
**(Campbell, Dunnette, et al., 1990)<sup>12</sup>**



Dans cette perspective, dans une étude de validation, la définition des critères devrait inclure des indicateurs tangibles, qu'ils soient objectifs ou subjectifs, plutôt que de s'en tenir à des définitions conceptuelles (Guion, 2010). Dans le même sens, alors que certains auteurs avancent que les critères de rendement doivent être des comportements ou les résultats de ces comportements, d'autres préfèrent restreindre le domaine de la performance à ce qui est réellement sous le contrôle de l'employé, c'est-à-dire ses comportements. On évite alors le risque de contamination par des facteurs externes à l'individu, par exemple le contexte économique, la qualité du produit, etc. (Pettersen, 2000). Par ailleurs, on pourrait penser que puisque le rendement est multidimensionnel, le critère devrait absolument être conceptualisé de manière multidimensionnelle. Par contre, dans la pratique, il pourrait en être autrement. Selon Cascio (1987), le choix d'un critère simple ou multiple dépend du contexte à l'étude. Si la validation est réalisée dans un contexte pratique pour appuyer les décisions de sélection, un critère global pourrait suffire.

En général, on s'entend d'abord pour dire que le critère doit être valide, c'est-à-dire représentatif de l'emploi pour lequel il a été choisi comme mesure de succès. Dans le même sens, le critère doit également être relié logiquement au prédicteur à l'étude afin de pouvoir mettre les deux en

<sup>12</sup> Tiré de l'ouvrage de Landy, F. J. et Conte, J. M. (2007). *Work in the 21st century : An introduction to industrial and organizational psychology, 2nd edition*. Malden, MA: Blackwell Publishing. p.179.

relations lors d'une stratégie de validation. En effet, comme le prédicteur est en fait une mesure individuelle, le critère doit également se référer au rendement individuel et sous le contrôle de l'employé. Par ailleurs, la mesure du critère doit également permettre d'identifier des différences entre les individus. En effet, on doit pouvoir constater de la variation autour du critère pour établir une discrimination entre les personnes afin que le critère soit considéré comme étant utile. Finalement, le critère doit être pratique, c'est-à-dire pouvoir être mesuré sans interférer avec le bon fonctionnement de l'organisation (Pettersen, 2000).

D'un autre regard, la mesure du critère doit également être exacte et exempte de contamination par rapport à des effets étrangers au rendement. En fait, puisqu'il s'agit généralement de l'évaluation du rendement effectué par le superviseur qui est utilisé pour compiler cette mesure du critère, il est nécessaire de limiter les sources de biais des évaluateurs : l'effet de halo, les stéréotypes, le facteur de similarité entre l'évaluateur et la personne évaluée, la sévérité naturelle de l'évaluation ou sa clémence, l'effet de tendance centrale, l'humeur générale de l'évaluateur ou tout simplement le phénomène de désirabilité sociale (Landy, 1983; Podsakoff, MacKenzie, Lee et al., 2003). On note, par exemple, que les facteurs liés au contexte de l'évaluation du rendement (promotion, développement des compétences, diagnostic organisationnel, etc.) ont une influence sur les buts poursuivis par le superviseur ce qui viendrait, à son tour, influencer son indulgence et les autres formes d'erreurs de mesure (Tziner, Joanis et Murphy, 2000).

Par ailleurs, l'attention des chercheurs dans ce domaine a longtemps été portée à l'élaboration d'échelles d'évaluations du rendement qui permettraient de contrôler ou de réduire les biais déjà identifiés (Austin et Villanova, 1992). Par exemple, la création des échelles BARS (*Behaviorally anchored scales*) a eu un impact majeur autant pour la recherche que pour la pratique, en mettant non plus l'accent sur les traits de l'individu lors de l'évaluation faite par le superviseur, mais plutôt sur l'observation de ses comportements au travail (Austin et Villanova, 1992).

Dans le même sens, on note peu de fidélité dans la mesure de l'évaluation du rendement par les superviseurs (Viswesvaran, Ones et Schmidt, 1996). On assiste souvent à un accord inter-juges très faible, différents superviseurs accordant une évaluation et une pondération différente de l'importance de comportements liés aux tâches, au contexte organisationnel dans leur évaluation globale. De cette manière, les études démontrent que les évaluations du rendement tendent à être plus précises lorsque

les conditions suivantes sont respectées : (1) possibilité d'observer régulièrement la performance des subordonnés ; (2) accès à des informations pertinentes sur la performance des employés ; (3) familiarité du superviseur à l'égard d'une méthode d'évaluation valide et fiable et (4) compréhension des normes de performance (St-Onge et Haines, 2007). Cependant, la fidélité du critère ne doit jamais primer sur son manque de pertinence. En effet, le manque de pertinence d'un critère a des conséquences beaucoup plus irrémédiables que le manque de fidélité (Pettersen, 2000). À l'évidence, il n'est pas possible de corriger statistiquement après coup des biais qui seraient survenus dans la mesure du rendement des employés ne sachant pas exactement la nature et l'ampleur de ce biais (par exemple, mesure du rendement des employés à partir de quelques indices objectifs, mais incomplets).

## PARTIE 2 : CADRE D'ANALYSE

---

À la lumière de la documentation scientifique recensée précédemment, il appert que la sélection d'un personnel compétent en centre d'appels comporte des enjeux importants pour la satisfaction de la clientèle et la compétitivité d'une société. Or, l'efficacité des activités de dotation est tributaire des outils qui sont mis à la disposition des gestionnaires et des spécialistes en ressources humaines. En fait, des instruments d'évaluation valides permettent de choisir les ressources qui représentent la meilleure adéquation avec les postes à doter tout en permettant d'établir des inférences quant à leurs capacités et qualités personnelles. En ce sens, la recension des écrits présentée aux sections précédentes indique que les tests de jugement situationnel représentent des outils de choix permettant de répondre aux impératifs d'évaluation de candidats à des postes d'agents au service à la clientèle (McDaniel et al., 2001). Tout en étant plus efficaces et économiques pour les organisations, les TJS permettent également de prédire une portion unique du rendement en emploi qui va au-delà des mesures de l'intelligence et de la personnalité (Clevenger et al., 2001). Dans le même sens, les TJS comportent d'autres avantages indéniables, notamment sur le plan des réactions positives qu'ils génèrent chez les candidats ainsi que leur faible impact négatif envers les minorités visibles et ethniques visés par les programmes d'accès à l'égalité à l'emploi au Canada et aux États-Unis (Bauer, Truxillo, Sanchez et al., 2001; Motowidlo et Tippins, 1993).

Néanmoins, il a été aussi possible de constater que l'agenda de recherche des TJS demeure important. Que l'on pense aux raisons pour lesquelles ils permettent de prédire le rendement en emploi, à la grande variabilité observée sur le plan de leur validité critériée, au construit mesuré ou aux incidences culturelles liées à leur utilisation en contexte de sélection, plusieurs questions demeurent encore à éclaircir. Dans l'ensemble, on peut donc conclure que les TJS permettent de prédire le rendement futur, mais que (1) la force de cette relation varie selon les études (de 0,16 à 0,56) et que (2) les coefficients observés sont inférieurs à ceux obtenus par le biais d'autres exercices d'évaluation, tels que les tests d'habileté cognitive (Schmidt et Hunter, 1998). À l'évidence, il est nécessaire de se questionner sur les facteurs qui viennent influencer cette variabilité. Dans le même sens, par le biais de la recension des enjeux liés à leur développement, il s'avère que plusieurs travaux restent à compléter afin d'évaluer quelle approche permet d'en arriver à l'élaboration d'une grille de correction menant à la meilleure efficacité relative dans le cadre d'un processus de sélection. En effet, il n'existe toujours pas de consensus dans la littérature concernant la meilleure stratégie pour élaborer cette grille de cotation

(Bergman et al., 2006; Krokos et al., 2004; Weekley et al., 2006). Dans cette perspective, les prochaines lignes présenteront un cadre conceptuel portant sur l'évaluation de l'efficacité relative de différentes méthodes d'élaboration des grilles de correction du TJS. L'objectif de cette recherche, les questions de recherche qui y sont reliées ainsi que les hypothèses s'y rattachant composeront donc la partie suivante.

#### 4. CONCEPTUALISATION DE LA RECHERCHE

---

En premier lieu, il est approprié de résumer sommairement l'objectif de la recherche avant de proposer le modèle conceptuel qui sera soumis à la vérification empirique. En bref, l'objectif de ce projet de recherche est d'évaluer si l'une ou l'autre de ces trois approches permettant de déterminer la grille de correction d'un TJS conduirait à une meilleure source de discrimination et de prédiction du rendement en emploi dans le contexte de la sélection du personnel d'un centre d'appels, c'est-à-dire (1) la méthode empirique, (2) la méthode théorique, et (3) la méthode basée sur les experts de contenu (Weekley et al., 2006). La possibilité qu'une combinaison de ces trois approches permette d'obtenir une meilleure efficacité relative en sélection de personnel à l'utilisation isolée de chacune de ces approches sera également examinée : (4) la méthode hybride (Bergman et al., 2006).

---

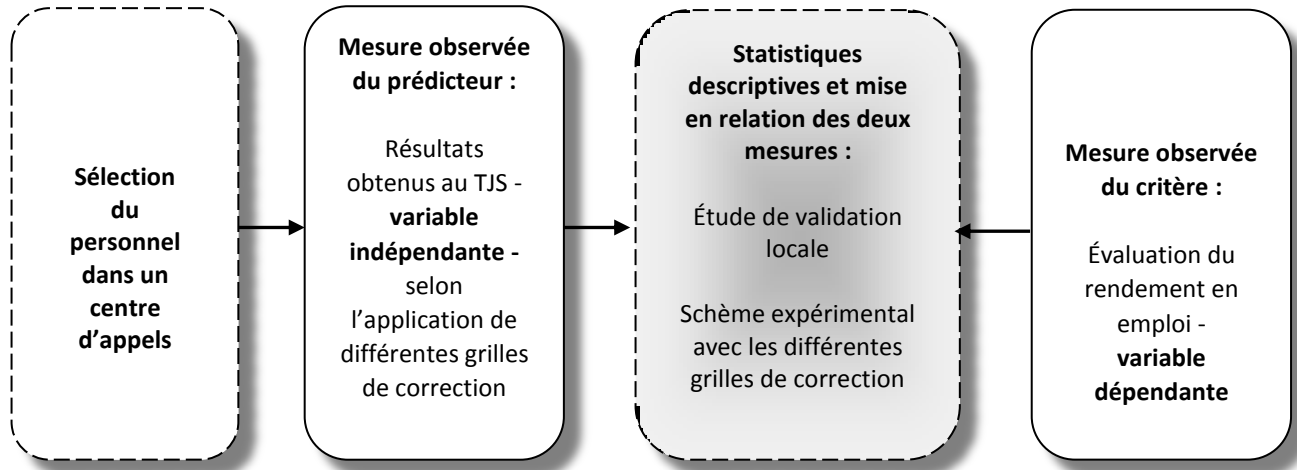
##### 4.1 QUESTIONS DE RECHERCHE

---

Le modèle conceptuel de la recherche repose en premier lieu sur les mesures observées par le prédicteur, c'est-à-dire les résultats obtenus au TJS, pour le pourvoi d'un poste dans un centre d'appels. Ces mesures représentent donc les variables indépendantes de l'expérimentation. Incidemment, afin de répondre aux questions proposées plus loin, la conceptualisation de cette recherche devra donc permettre d'isoler l'impact de chacune des méthodes d'élaboration d'une grille de correction sur la prédiction du rendement en emploi.



**Figure 8**  
**Modèle conceptuel de la recherche**



Afin de vérifier l'efficacité relative de chacune des grilles de correction, en plus d'observer les statistiques descriptives reliées au prédicteur, il sera essentiel de mettre en relation chacune de ces mesures du prédicteur – résultats obtenus au TJS lors de l'application des différentes grilles de correction – avec la mesure observée du critère – c'est-à-dire l'évaluation du rendement en emploi (variable dépendante).

En premier lieu, comme l'objectif d'un processus de sélection est de discriminer les candidats entre eux, se questionner sur l'efficacité relative des méthodes d'élaboration des grilles de correction d'un instrument élaboré dans un tel contexte justifie le fait de se questionner sur laquelle de ces méthodes permettra d'en arriver à une meilleure distribution des scores des participants. En d'autres mots :

***Q1 – Quelle méthode visant l'élaboration d'une grille de correction d'un TJS destiné à la sélection du personnel dans un centre d'appels permet d'en arriver à une meilleure distribution des scores obtenus par les participants ?***

Par ailleurs, dans la mesure où la sélection d'un personnel offrant le meilleur rendement en emploi est une exigence incontournable pour le bon fonctionnement d'un centre d'appels et étant donné que le TJS est un instrument d'évaluation fréquemment utilisé dans ce contexte, atteindre un niveau de validité critériée maximal est indispensable. Alors que la littérature demeure ambiguë sur la meilleure méthode à adopter afin d'y parvenir, l'objectif de ce projet de recherche sera également

d'évaluer l'efficacité relative des grilles de correction d'un test de jugement situationnel en se posant la question la suivante :

***Q2 – Quelle méthode visant l'élaboration d'une grille de correction d'un TJS destiné à la sélection du personnel dans un centre d'appels permet d'en arriver à un niveau de validité critériée supérieur aux autres ?***

---

## 4.2 HYPOTHESES DE RECHERCHE

---

Cette section regroupe les hypothèses se rattachant à cette recherche, lesquelles s'appuient notamment sur le cadre théorique présenté dans les sections précédentes.

En premier lieu, à l'instar de l'étude de St-Sauveur et ses collègues (2014) qui visait à évaluer l'efficacité des différentes méthodes d'attribution de pointage, l'efficacité relative des méthodes d'élaboration de grilles de correction peut également être évaluée à partir de la capacité de ces dernières à identifier les candidats possédant un certain potentiel par rapport à la population générale pour laquelle le test a été validé. En fait, il est plausible de croire qu'une bonne approche permettra de distribuer les scores de l'ensemble des personnes ayant participé au test selon une courbe normale signifiant ainsi que le test n'est pas trop facile, ni trop difficile et assurant une bonne répartition de toutes les personnes ayant posé leur candidature à un poste pour en arriver à de bonnes décisions d'embauche ou de rejet (T. P. Hogan, 2012).

Malgré que la littérature porte à croire qu'il ne semble pas y avoir réellement une valeur ajoutée à combiner une méthode empirique à une méthode rationnelle (c'est-à-dire basée sur des sources théoriques ou des experts de contenu), chacune des deux sources prises séparément apportent une perspective qui lui est propre et qui demeure pertinente lors de l'élaboration d'une grille de correction hybride (Cucina et al., 2012). En fait, par l'adoption de la méthode hybride lors du développement d'une clé de correction, il est fort probable que l'on puisse limiter les faiblesses méthodologiques liées à la méthode empirique tout en minimisant les divergences souvent observées avec les méthodes de type rationnel, telles que les bonnes réponses identifiées par différents experts de contenu. Dans cette perspective, la méthode d'élaboration hybride retenue dans le cadre de ce projet de recherche vise

d'abord à maximiser les sources d'information à la disposition d'un concepteur en combinant des sources théoriques, celles provenant des experts de contenu et les données de nature empirique. Puis, d'un point de vue théorique et pratique, il est justifié d'accorder le privilège aux experts de contenu afin qu'ils procèdent à la validation finale de la grille de correction hybride pour en arriver à une compréhension globale du rôle et des facteurs de rendement associés au test.

En ce qui a trait à la moyenne attendue, St-Sauveur et ses collègues (2014) suggèrent qu'une moyenne de 60% représente un niveau de difficulté optimal dans le cadre d'un processus de sélection. Ce score est relativement central (près de 50%) et correspond à une attente culturelle issue du monde de l'éducation où ce résultat représente souvent le seuil minimum de passage. De plus, une moyenne de 60% permet un bon niveau de variabilité vers le haut (entre 60% et 100%) ou vers le bas (entre 0% et 60%).

Par conséquent, comme une distribution normale symétrique et la moyenne contribuent à l'évaluation de la qualité d'un test utilisé en sélection de personnel, la première hypothèse de recherche postulée est la suivante :

***H1a – La méthode hybride présente la meilleure distribution des scores des candidats avec une distribution normale symétrique et une moyenne la plus près de 60,0%.***

En ce qui concerne la deuxième question de ce projet de recherche, sont postulées les hypothèses suivantes :

***H2a – La méthode hybride présente un niveau de validité critériée supérieur à la méthode théorique.***

***H2b – La méthode hybride présente un niveau de validité critériée supérieur à la méthode basée sur les experts de contenu.***

***H2c – La méthode hybride présente un niveau de validité critériée supérieur à la méthode empirique.***

Cette deuxième série d'hypothèses s'appuie d'abord sur le fait que l'approche théorique nécessite une compréhension sophistiquée du domaine d'activité et des antécédents liés à un bon rendement dans un tel poste, ce qui, la plupart du temps, rend les critères attendus très difficiles à soutenir (Bergman et al., 2006). Dans le même sens, les meilleures pratiques théoriques ne peuvent pas toujours être appliquées telles quelles dans les organisations et il arrive souvent qu'elles trouvent peu d'échos sur le terrain. Dans cette perspective, il semble plausible de croire qu'une approche hybride visant à compléter cette information provenant de sources théoriques offrira un meilleur niveau de prédiction tout en s'assurant de leur généralisation dans l'organisation visée.

Par ailleurs, lorsque l'on adopte la méthode axée sur les experts de contenu, il est clair qu'il peut parfois survenir des divergences importantes dans l'a priori de leurs appréciations des éléments importants du rendement en emploi, et ce, à l'intérieur d'une même organisation (Motowidlo, et al., 1990). En adoptant une approche hybride, on tente alors de maximiser cet effort de prédiction en validant le portrait dressé par les experts par le biais d'autres sources d'information afin de parvenir à une meilleure estimation du rendement futur en emploi.

Finalement, lors de l'adoption d'une approche strictement empirique, selon l'identification du critère et l'échantillon utilisé à titre de référence, il peut y avoir une très grande variabilité dans la prédiction du rendement d'autant plus qu'il arrive parfois que des conclusions établies puissent apparaître comme irrationnelles et très peu valides pour les « non initiés » aux méthodes statistiques (Cucina et al., 2012). En somme, il semble plausible de croire que l'approche hybride permet de s'assurer que l'identification des bonnes réponses puisse être expliquée et interprétée, tout en maximisant l'effort de prédiction apporté par les méthodes empiriques (Cucina, et al., 2012; Mumford, 1999).

## PARTIE 3 : METHODOLOGIE

---

La section qui suit vise à décrire la façon dont a été opérationnalisée la recherche ainsi que les analyses statistiques utilisées afin d'atteindre les objectifs de cette expérimentation.

### 5. OPERATIONNALISATION DE LA RECHERCHE

---

Pour rendre opérationnelles la question de recherche proposée, l'essentiel de l'expérimentation est d'utiliser un TJS (instrument de mesure pour le prédicteur) qui rencontre les normes de testing décrites précédemment en l'accompagnant de grilles de correction élaborées selon différentes approches proposées par la littérature afin de les corrélérer individuellement avec le rendement en emploi d'une population cible (via un instrument de mesure pour le critère). Cette mise en relation provenant de deux sources d'information différentes permet de prendre en compte les erreurs de mesure traditionnellement associées aux méthodes communes en vérifiant quelle grille de correction conduit à l'établissement d'une meilleure distribution des scores et d'un niveau de prédiction du rendement en emploi supérieur aux autres et, par le fait même, à une meilleure efficacité relative en sélection de personnel (Podsakoff et al., 2003).

Pour y parvenir, des contacts avec une firme de consultation en ressources humaines ont été réalisés afin de permettre l'accès à des données secondaires provenant d'une organisation œuvrant dans le domaine des centres d'appels et travaillant à mettre en place des processus de sélection de personnel visant la mesure et l'évaluation des différences individuelles de ses candidats.

---

#### 5.1 INSTRUMENTS DE MESURE POUR LE PREDICTEUR

---

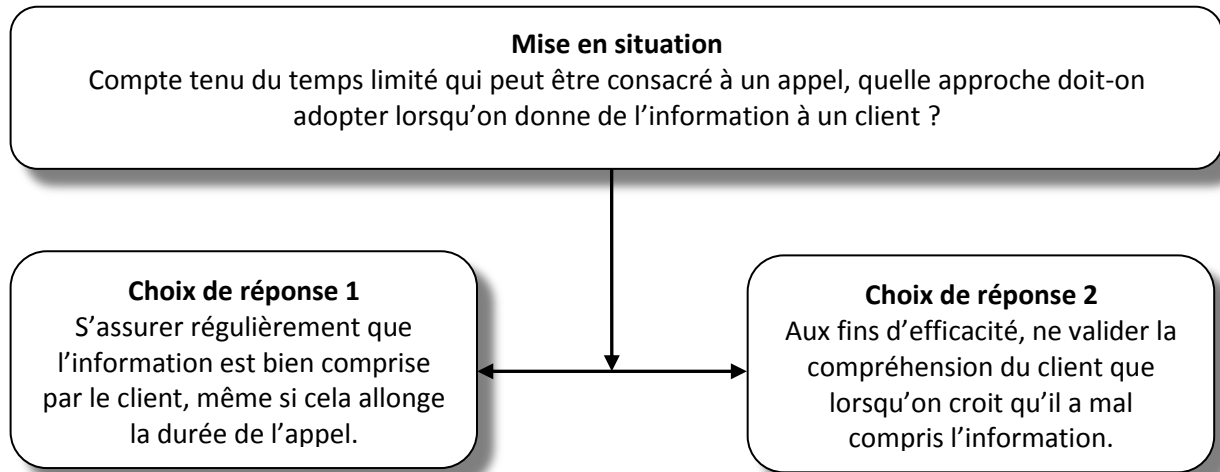
Pour le prédicteur, c'est-à-dire l'outil d'évaluation utilisé afin de prédire le rendement en emploi d'une personne, l'instrument de mesure retenu est le « Test pour Centre Contact Client (T3C) © » dont la propriété intellectuelle appartient à la firme Compmetrica (EPSI, 2015). En fait, ce TJS est largement utilisé dans plusieurs organisations au Canada et en France pour l'évaluation de compétences

essentielles à l'accomplissement des tâches effectuées par des employés œuvrant dans un centre contact client et la sélection de candidats à des postes de cette nature. Plus spécifiquement, tel que présenté dans la documentation de référence à propos du test, ce dernier comporte un tronc commun et des modules additionnels qui permettent d'évaluer des compétences directement associées à des activités de service à la clientèle, de vente et de recouvrement. En effet, l'examen des étapes qui ont été suivies lors du développement et de la validation du test, tel que précisé dans le manuel technique du test, a mené à la conclusion qu'il s'agit d'un instrument de mesure des différences individuelles des téléopérateurs rencontrant les exigences de nature scientifique pour être associées à cette recherche (EPSI, 2015).

D'emblée, cette décision d'utiliser les données secondaires fournies par la firme Compmetrica dans le cadre de cette recherche permet ainsi au chercheur de reproduire une situation similaire à celle observée réellement en contexte organisationnel, tout en élargissant les possibilités d'avoir accès à un terrain propice et des données pertinentes pour un tel projet de recherche. En effet, lors du développement de cet instrument, tel que présenté dans le manuel technique, les concepteurs du T3C ont bénéficié de l'appui de plusieurs sociétés évoluant dans le monde des centres d'appels, tant ici qu'à l'étranger (EPSI, 2015). Ces organisations ont non seulement participé à l'élaboration des mises en situation et à leur validation en y dédiant des experts de contenu (directeurs et superviseurs en centre d'appels), mais ils ont également fourni, à l'étape de l'expérimentation et de la calibration du test, plusieurs de leurs employés titulaires de postes en centre d'appels depuis un minimum de 6 mois.

Bien que la version intégrale du test T3C comporte deux parties, seule la première partie du test est utilisée dans le cadre de cette recherche. En effet, la seconde section du test vise à mesurer des traits de personnalité par le biais d'échelles psychométriques plutôt que le rendement futur en emploi par le biais questions à choix multiples, tel que celles typiquement présentées dans un TJS. Ainsi, dans la première partie du T3C, chaque question offre deux options de réponses aux candidats. Chacune de ces options de réponses représente une façon appropriée de faire face à une situation qui se présente habituellement dans le cadre des fonctions d'un téléopérateur. Cependant, le candidat doit sélectionner celle qui lui semble la plus appropriée. Notons que cet exercice est constitué de questions comprenant des directives reliées aux connaissances (*should do*). La figure 9 est une représentation schématique de la structure de chaque situation et options de réponses qui s'y rattachent.

**Figure 9**  
**Exemple d'une question du T3C**



Pour chaque question, la personne candidate doit d'abord choisir l'énoncé qui lui paraît être le plus approprié. Ensuite, une échelle de réponse en 4 points lui est présentée. La personne candidate doit alors déterminer jusqu'à quel point elle est en accord avec cet énoncé en indiquant si elle est « un peu en accord » ou « fortement en accord » avec cette option de réponse. Ainsi, pour chacune des questions, la personne candidate doit donc faire un seul choix parmi les quatre options suivantes :

- a. Le choix **a** indique que vous êtes fortement en accord avec l'énoncé 1.
- b. Le choix **b** indique que vous êtes un peu en accord avec l'énoncé 1.
- c. Le choix **c** indique que vous êtes un peu en accord avec l'énoncé 2.
- d. Le choix **d** indique que vous êtes fortement en accord avec l'énoncé 2.

Si on reprend l'exemple précédent, on note que les directives présentées dans le cahier de test indiquent que :

« En répondant à chaque question, vous devez d'abord déterminer, selon vous, quelle approche doit-on adopter lorsqu'on donne de l'information à un client lors d'un appel.

Si vous croyez qu'il est préférable de s'assurer régulièrement que l'information est bien comprise par le client, même si cela allonge la durée de l'appel, vous optez pour le choix a si vous êtes fortement en accord avec cet énoncé ou le choix b si vous êtes un peu en accord avec cet énoncé.

Si, au contraire, vous croyez qu'aux fins d'efficacité on ne devrait valider la compréhension du client que lorsqu'on croit qu'il a mal compris l'information, vous optez pour le choix c si vous êtes un peu en accord avec cet énoncé ou pour le choix d si vous êtes fortement en accord avec cet énoncé. »

En ce qui concerne le mode d'attribution du pointage pour ce TJS, les candidats obtiennent 5 points lorsqu'ils choisissent l'énoncé ayant été identifié comme la bonne réponse dans la grille de correction, alors qu'ils accumulent 2 points lorsqu'ils sont un peu en accord avec cet énoncé. Toutefois, dans le cadre du présent projet de recherche, il s'agit de la méthode du choix forcé qui a été retenu, c'est-à-dire que le participant se voit attribuer 1 point dès que sa réponse indique qu'il favorise l'énoncé identifié comme étant la bonne réponse dans la grille de correction et 0 point si ce n'est pas le cas.

Le nombre de questions et la durée du test administré au candidat varient en fonction des modules complémentaires qui sont utilisés par l'organisation et des caractéristiques qui sont requises par la nature du poste à doter. Dans le cadre de cette recherche, les questions relatives au tronc commun ainsi que celles reliées aux modules complémentaires service à la clientèle et vente sont utilisées. Il s'agit alors d'un test comprenant un total de 80 questions qui doivent être répondues à l'intérieur d'une période de 60 minutes.

Le T3C a été retenu pour l'opérationnalisation de cette recherche, car il s'agit d'un TJS possédant de bonnes caractéristiques psychométriques, qui a déjà été validé avec une population francophone nord-américaine et élaboré selon les paramètres décrits dans la littérature sur le sujet (EPSI, 2015). En effet, sur le plan de la validité basée sur le contenu, il s'agit d'un instrument qui s'inspire à la fois des deux méthodes reconnues dans la littérature pour le développement d'un TJS, soit les approches déductive et inductive (Weekley et al., 2006). Par ailleurs, étant donné qu'il est constitué de questions comprenant des directives reliées aux connaissances (*should do*), il permet en partie de contrôler l'enjeu de falsification et de désirabilité sociale inhérent aux questions de nature comportementale (*would do*). Cependant, ce type de directives pourrait en retour générer de plus grandes différences dans les scores des populations provenant des groupes cibles (minorités ethniques et minorités visibles), et ce, tout dépendant de la magnitude avec laquelle le test est fortement corrélé avec le facteur *g* (Whetzel, McDaniel et Nguyen, 2008).

D'abord, les concepteurs du T3C ont mis à profit une quantité considérable de renseignements portant sur des postes de services à la clientèle, la vente et le recouvrement y compris des dizaines



d'incidents critiques lesquels s'appuyaient sur des entrevues approfondies auprès de spécialistes en la matière provenant tant du secteur public que privé, de documents de référence (par exemple, environ 50 descriptions de poste) et publications scientifiques portant sur les meilleures pratiques en centre d'appels permettant de contribuer à la validité de contenu de cet instrument. Ils ont également procédé à plusieurs journées d'observation dans des centres d'appels différents afin de s'assurer de la représentativité des scénarios présentés.

De plus, le processus d'élaboration suivi par les concepteurs de Compmetrica s'est initialement inspiré d'un modèle de compétences afin de s'assurer que les items soient associés à des construits reliés positivement au rendement au travail d'un téléopérateur. Pour ce faire, ils ont appuyé leur instrument sur le Modèle de compétences Compmetrica©. Cet inventaire de compétences a été élaboré par un groupe de professionnels œuvrant dans le domaine de l'évaluation et de la gestion des compétences à la suite d'un examen approfondi des exigences requises par plusieurs centaines d'emplois des secteurs privés, publics et parapublics et à la suite d'une étude de plusieurs modèles de compétences existants, tant au niveau commercial qu'organisationnel. En tout, 60 compétences couramment associées aux emplois répertoriés ont été retenues et classées selon différents domaines de performance. L'analyse de l'ensemble de la documentation portant sur les centres d'appels a permis aux concepteurs du T3C d'identifier, à partir des 60 compétences de base, une liste de 17 dimensions jugées importantes lors de l'exécution des tâches dans un centre d'appels. Ces compétences et leur définition sont présentées à l'annexe A.

Puis, suite à l'élaboration préliminaire de l'instrument de mesure, quatre psychologues organisationnels expérimentés ont entrepris un examen approfondi du contenu du TJS afin de valider les liens conceptuels entre chaque énoncé de réponse et l'élément du modèle de compétences évalué. Les quatre experts ont d'abord procédé à un examen individuel du contenu des questions et se sont rencontrés par la suite en groupe afin de résoudre toute divergence de point de vue. Des modifications au contenu et à la grille de cotation ont été apportées en conséquence. Cette stratégie fait donc en sorte que chaque option de réponse est orientée vers la mesure d'une ou de deux des dimensions évaluées. En effet, malgré le fait que cet exercice mesure la « capacité à remplir efficacement le rôle et les responsabilités attribuées à une personne évoluant en centre d'appels » (résultat de la moyenne globale à l'exercice qui est habituellement utilisé en sélection de personnel), les concepteurs ont procédé à la

création de sous-échelles en fonction de leur modèle de compétences génériques afin de faciliter la rétroaction aux candidats (voir annexe A).

Finalement, cet instrument de mesure présente de bonnes propriétés psychométriques. Comme il est possible de le constater aux annexes B et C les statistiques descriptives et la structure factorielle du test en font un TJS pouvant être mis à contribution dans le cadre d'une recherche expérimentale. En effet, un examen de l'annexe B indique que les moyennes se situent entre 60% et 70% (du score maximum possible). Par ailleurs, la centralité des moyennes pour chacune des versions garantit une variance maximum (entre 1,5 et 2 écarts-type de chaque côté de la moyenne) des résultats tant vers le haut que vers le bas de l'échelle. Qui plus est, comme il l'est souligné dans le manuel technique du test T3C, la distribution des résultats autour de la moyenne suit bien une loi normale, toutes les échelles présentant des indices de symétrie et les coefficients d'aplatissement (*kurtosis*) inférieurs à 1,0 et supérieurs à -1,5 (Carricano, Poujol et Bertrandias, 2010). Dans le même sens, une analyse factorielle exploratoire a été effectuée par les concepteurs afin de déterminer la structure sous-jacente aux dimensions évaluées par le T3C. L'analyse exploratoire est présentée en composantes principales et les facteurs possédant une valeur propre égale ou supérieure à 1,0 ont été retenus. Une rotation oblique a été effectuée afin de tenir compte de l'interdépendance des composantes (EPSI, 2015). En effet, ce processus permet de faciliter l'interprétation des facteurs en maximisant les saturations les plus fortes et en minimisant les plus faibles de sorte que chaque facteur apparaisse déterminé par un ensemble restreint et unique de variables. La rotation cherche donc à simplifier les composantes en minimisant le produit croisé des coefficients de régression (les composantes sont corrélées) (Tabachnick et Fidell, 2007). L'annexe C présente les résultats de cette analyse factorielle exploratoire.

---

## 5.2 ÉLABORATION DES GRILLES DE CORRECTION

---

Comme il l'a été indiqué précédemment, trois méthodes d'élaboration de grilles de correction sont décrites dans la littérature portant sur les tests de jugement situationnel : (1) la méthode théorique, (2) la méthode basée sur les experts de contenu et (3) la méthode empirique (Weekley et al., 2006). À cette liste, on y ajoute également (4) la méthode hybride qui combine différentes clés de correction entre elles. Afin de déterminer quelle est l'efficacité relative de chacune d'entre elles par rapport à leur

efficacité relative en sélection de personnel, la présente recherche s'appuie sur l'élaboration de grilles de correction suivant chacune des méthodes identifiées plus tôt.

### 5.2.1 GRILLE DE CORRECTION THEORIQUE

---

Dans le cadre de cette recherche, la grille de correction établie de manière théorique est celle proposée par les concepteurs du T3C. En effet, le manuel technique de cet instrument de mesure décrit la manière dont les bonnes réponses ont été identifiées théoriquement lors du développement du test (grille de correction no 1). En effet, pour y parvenir, les concepteurs ont d'abord élaboré un cadre de référence suite à la revue de la documentation scientifique et professionnelle concernant les différents thèmes couverts par le test, c'est-à-dire le service à la clientèle et la vente. Cette recension des écrits leur a permis de déterminer les grandes catégories de comportements recherchées pour l'efficacité d'un centre d'appels. Par la suite, les critères pertinents pour la fonction de téléopérateur mesurés par chacun des choix de réponses présentés dans le test ont été liés conceptuellement à un cadre théorique relié au rendement en emploi, c'est-à-dire le Modèle de compétences Compmetrica.

Par la suite, un comité de travail composé de cinq (5) experts-conseils, psychologues du travail et spécialistes en ressources humaines, a été invité à participer à la validation de cette grille de cotation. Les membres de ce comité ont d'abord eu comme responsabilité de prendre connaissance de l'ensemble de l'exercice, c'est-à-dire les mises en situations et options de réponses proposées dans le T3C. Par la suite, ils devaient analyser chaque option de réponse proposée et choisir celle qui reflète davantage le bon comportement à adopter en faisant le lien conceptuellement avec chacune des dimensions évaluées par le test (généralement deux dimensions par option de réponses). De cette manière, afin d'en arriver à identifier les options de réponses à privilégier en fonction de la théorie, chaque membre du comité de travail devait procéder à la mise en relation rationnelle des choix de réponse avec les caractéristiques souhaitées et leurs différences sur le rendement. Pour y parvenir, ces derniers devaient s'appuyer sur la documentation scientifique et professionnelle fournie par les concepteurs, c'est-à-dire :

- Les descriptions des emplois visés par l'exercice précisant la nature des tâches et des responsabilités ;
- Le profil des compétences pour ce type de poste incluant les capacités et les qualités requises pour œuvrer efficacement dans un emploi de cette nature ainsi que les attentes comportementales ;
- Trois articles scientifiques qui ont été utilisés par les concepteurs lors de l'élaboration du cadre de référence du T3C<sup>13</sup>.

À la suite de l'identification des bonnes réponses de manière individuelle, les membres du comité de travail se sont rencontrés et ont passé en revue chacune des réponses recommandées. Lorsqu'il y avait divergence dans les recommandations, une discussion était générée afin d'identifier les références théoriques sous-jacentes à chacune des affirmations et d'en venir à un consensus sur l'approche à privilégier en fonction de la théorie. Dès qu'un consensus réunissant au moins 80% des membres du comité était atteint, l'option de réponse choisie par la majorité était retenue.

En regard à l'attribution des pointages suite à l'identification de la réponse la plus appropriée par le candidat, l'approche dichotomique (c'est-à-dire 1 point pour une bonne réponse et 0 pour une mauvaise réponse) a été conservée pour l'ensemble des grilles de correction élaborée afin de permettre plus facilement les comparaisons statistiques.

### 5.2.2 GRILLE DE CORRECTION PAR LES EXPERTS DE CONTENU

---

Pour l'opérationnalisation de cette recherche, plusieurs ateliers de travail avec des experts de contenu ont permis d'élaborer un autre type de grille de correction afin de vérifier l'efficacité relative de cette approche. Étant donné que plusieurs organisations distinctes ont participé à cette étape de la recherche, ces différents ateliers de validation ont également permis de vérifier s'il existe une certaine variabilité dans l'identification des bonnes réponses lors de l'élaboration de différents types de grilles de

---

<sup>13</sup> Références bibliographiques soumises à leur attention : 1) Burgers, A., Ruyter, K. d., Keen, C. et Streukens, S. (2000). Customer expectation dimensions of voice-to-voice service encounters: a scale-development study. *International Journal of Service Industry Management*, 11 (2), 142-161. 2) Gwinner, K. P., Bitner, M. J., Brown, S. W. et Kumar, A. (2005). Service Customization Through Employee Adaptiveness. *Journal of Service Research*, 8 (2), 131-148. 3) Winsted, K. F. (2000). Service behaviors that lead to satisfied customers. *European Journal of Marketing*, 34 (3/4), 399-417.

correction. En somme, aux fins de cette étude, trois grilles de correction ont été élaborées en suivant la méthode basée sur le jugement des experts de contenu (SMEs).

En premier lieu, un comité de travail de quatorze (14) experts de contenu provenant d'une grande institution financière québécoise (organisation identifiée à titre de terrain de recherche) a procédé à la validation d'une grille de correction particulière en fonction de leurs croyances concernant les comportements à adopter en centre d'appels (grille de correction no 2). Ces ressources étaient pour la plupart des directeurs(trices) supervisant le travail de téléopérateurs. Au cours de cet atelier, le contenu de l'exercice ainsi que toutes les paires d'énoncés leur ont été présentés. Les experts devaient déterminer, pour chacune des paires, l'énoncé (a) ou (b) qui représente la façon la plus appropriée de faire face à la situation. Ce choix a été fait en fonction de leur expérience, de leur connaissance de l'emploi et du contexte particulier de leur organisation. Comme dans le cas de la grille de correction établie de manière théorique, dès qu'un consensus réunissant au moins 80% des membres du comité était atteint, l'option de réponses choisie par la majorité était retenue.

Par ailleurs, afin de constater si une certaine variabilité est présente lors de l'élaboration de différentes grilles de correction provenant de sources d'information similaires, deux ateliers de validation supplémentaires ont été mis à contribution dans le cadre de cette recherche. En fait, un groupe de six (6) experts de contenu et superviseurs d'employés de centre d'appels dont la tâche consiste à offrir des produits et services aux clients d'une organisation œuvrant dans le secteur de l'énergie s'est également penché sur l'identification des bonnes réponses en fonction de leur connaissance et expérience du milieu du centre d'appels (grille de correction no 3). Finalement, des données provenant de cinq (5) experts de contenu d'une vaste agence gouvernementale québécoise ont également été soumises à l'analyse et une grille de correction a été élaborée en utilisant une démarche identique à celle utilisée pour les grilles précédentes (SMEs) (grille de correction no 4).

### 5.2.3 GRILLE DE CORRECTION EMPIRIQUE

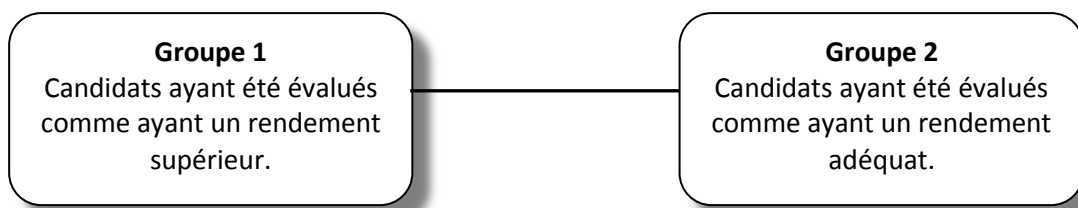
---

Pour l'établissement d'une grille de correction provenant de sources empiriques, des analyses statistiques ont été réalisées à partir de la base de données fournie par l'organisation identifiée à titre de terrain de recherche qui a participé à la validation du T3C. En effet, lors du déploiement de ce test au sein de l'organisation, 312 personnes titulaires de postes en centre d'appels ont été invitées à compléter

le test. Plus de détails concernant la composition de cet échantillon sont décrits à la section traitant spécifiquement de ce sujet (voir section 5.5).

Puis, pour l'ensemble de ces candidats, l'organisation a demandé à leur supérieur immédiat de leur attribuer une cote de rendement afin d'identifier les employés démontrant un rendement supérieur en emploi. Cette étape d'évaluation a été réalisée à l'extérieur du cadre de l'évaluation officielle du rendement de l'entreprise. La présentation détaillée de l'opérationnalisation de cette mesure du rendement est présentée spécifiquement à la section suivante de cette thèse (voir section 5.3). Notons simplement ici que les sujets de l'étude ont été catégorisés en deux groupes, soit les employés dits « adéquats » (n=160) (cote de 1 représentant un score de 1 ou 2 sur l'échelle d'évaluation du rendement utilisée), soit les employés dits « performants » (n= 152) (cote de 2 représentant un score de 3 ou 4 sur l'échelle d'évaluation du rendement utilisée). Ainsi, pour les besoins de la présente recherche, un échantillon de grilles de correction empiriques a été élaboré en mettant en relation un critère (le rendement en emploi) attribué à un groupe référence (voir figure suivante).

**Figure 10**  
**Critères de référence pour la grille de correction empirique**



Habituellement, sur le plan pratique, les organisations qui utilisent l'approche empirique identifient un nombre relativement restreint d'employés (souvent entre 30 et 60 individus), la moitié d'entre eux étant considérés comme « performants » et l'autre comme étant « adéquats ». À l'évidence, les organisations sont la plupart du temps réticentes à administrer un nouveau test destiné à la sélection au personnel en place et à mobiliser les énergies de plusieurs intervenants de l'organisation afin de fournir des données de rendement aux concepteurs immédiatement avant le déploiement du test.

Lorsque certains employés ont été choisis afin de participer à cet exercice, ces derniers complètent alors le test et leurs réponses soumises à l'examen sont ensuite corrélées à leur niveau de

rendement en emploi. Si un plus grand nombre d'employés « performants » choisissent l'option « A » et qu'un plus grand nombre d'employés « adéquats » choisissent l'option « B », il est donc déterminé empiriquement que la bonne réponse devrait être « A », peu importe qu'elle corresponde ou non à un schème théorique ou à une orientation privilégiée par l'entreprise. Une autre façon de faire consiste, pour chacun des items, à corrélérer les choix de réponses au niveau de performance associée à chaque répondant. Prenons un exemple qui correspond au modèle de TJS utilisé dans le cadre de cette étude. Lorsqu'un participant choisit un énoncé, on lui accorde alors une cote de 1 (soit choix A ou soit choix B). Lorsqu'il ne choisit pas cette option, il obtient un score de 0 (voir tableau 7 pour l'illustration de cette démarche). Parallèlement à cet exercice, les employés performants sont identifiés avec un score de « 2 » et les employés adéquats avec un score de « 1 ». Finalement, une corrélation entre les réponses au choix A (0 ou 1) et le niveau de performance des employés (1 ou 2) est alors établit et la même démarche est réalisée pour le choix B. Dans ce cas, la bonne réponse est celle qui donne une corrélation positive. Dans le cadre de cet exemple, comme le TJS présente un choix dichotomique, la corrélation obtenue pour le choix A est la même que celle obtenue pour le choix B, seul le signe variant entre les deux mesures (ex : Choix A : +0,25 Choix B : -0.25). La même procédure est utilisée pour chacun des items, jusqu'à ce qu'une grille de correction complète soit générée.

**Tableau 7**  
**Exemple de résultats à un des items du test**

	<b>Choix A</b>	<b>Choix B</b>	<b>Performance 1= Adéquat 2= Performant</b>
Employé 1	0	1	1
Employé 2	1	0	2
Employé 3	1	0	1
Employé 4	0	1	1
Etc.			

Opérationnellement, cette approche n'est habituellement effectuée qu'une seule fois, la grille de correction finale étant tributaire de la qualité des employés choisis comme étant performants ou adéquats. Cette approche est donc en soi sujette à un biais possible d'échantillonnage. Afin de minimiser l'impact de ce biais, un échantillon de 30 grilles de correction empiriques ont été conçues à partir de l'ensemble des sujets ayant participé à l'étude. Ce faisant, l'approche qui est habituellement utilisée par

les concepteurs de TJS a été reproduite aux fins de la présente recherche. Dans l'ensemble, la génération de chacune de ces grilles a suivi la démarche suivante :

1. 30 employés performants et 30 employés adéquats ont été choisis au hasard parmi les 312 sujets ayant participé à l'étude ;
2. Les réponses de ces 60 employés à chacun des items ont été mises en corrélation avec le niveau de performance identifié ;
3. Pour chacun des items, le choix (A ou B) démontrant une corrélation positive avec le niveau de performance a été retenu comme étant la bonne réponse pour cette grille ;
4. La grille de correction ainsi générée était ensuite appliquée aux 252 autres participants faisant partie de la base de données. De cette façon, les 60 personnes ayant contribué à générer la grille de correction étaient éliminées des analyses subséquentes. Cette approche évitait d'inclure des sujets qui par définition auraient augmenté la corrélation finale (puisque les choix de réponses ont été basés sur les corrélations obtenues par ces participants).

Cette étape a été répétée 30 fois. Dans chacun des cas, les 60 employés étaient choisis au hasard, de sorte que chaque échantillon était différent.

Aux fins de l'analyse et de la présentation des résultats, trois grilles de correction (no 6 no 19 et no 28) ont été retenues, soit celle démontrant un coefficient de détermination le plus près de la moyenne lors de la comparaison entre les trente grilles, la grille ayant démontré le meilleur coefficient de détermination et celle ayant démontré le moins bon coefficient de détermination (voir annexe D). Au final, la sélection de ces trois grilles de correction permet de couvrir le large spectre de possibilités en termes de prédiction du rendement par le biais de l'approche empirique, et ce, tout en respectant le scénario typiquement rencontré lors du développement d'un nouveau TJS. Néanmoins, afin de tirer des conclusions sur l'efficacité de cette méthode, cette recherche s'appuiera davantage de la grille 19 lors de la comparaison entre les différentes approches étant donné qu'elle représente la moyenne de tous les scénarios produits par l'approche empirique. Ce choix est également motivé par le fait que c'est cette grille qui a été utilisée par les experts lors de l'élaboration de la grille de correction « hybride ».



La clé de correction s'inspirant d'une approche hybride combine les informations recueillies par le biais des trois méthodes décrites précédemment. Elle est donc appuyée sur des données provenant de sources empiriques et d'autres de sources rationnelles. En effet, pour y parvenir, un comité de travail réunissant 5 experts de contenu (experts du domaine financier appartenant à l'organisation ayant participé au terrain de recherche et différents de ceux qui ont participé aux autres comités) a été rencontré afin de juxtaposer l'ensemble des informations provenant des grilles de correction élaborées précédemment avec pour objectif de déterminer une grille de correction « finale ». Les experts ont d'abord pris connaissance de l'instrument de mesure, des questions et des options de réponses proposées. Par la suite, pour chacune des questions, ils ont eu accès aux trois grilles de correction élaborées précédemment (théorique, basé sur les experts du domaine financier et grille empirique 19). En fonction de l'ensemble de ces informations, les experts devaient alors déterminer quelle réponse devrait être identifiée comme étant la « bonne réponse ».

Plus spécifiquement, pour chacune des questions de l'exercice, un tableau leur présentait quelle option de réponses a été identifiée comme la meilleure façon de résoudre la situation selon la théorie, selon l'avis d'experts de contenu (experts financiers) et selon une source rationnelle (grille empirique 19 possédant un coefficient de détermination moyen). Par le biais de la discussion, les experts devaient en venir à un consensus sur la bonne réponse à retenir.

Notons que dans le cadre de cet exercice, nous avons pu observer que les réponses privilégiées tenaient compte de l'ensemble des informations fournies. Dans certains cas, la puissance statistique de la méthode empirique prévalait dans le choix de la « bonne réponse ». Par contre, dans d'autres cas, la difficulté à interpréter rationnellement cette option pouvait les mener à choisir une réponse correspondant au choix des experts de contenu ou même de l'approche théorique. Dans tous les cas, la réponse retenue devait faire l'objet d'un consensus.

---

### 5.3 INSTRUMENT DE MESURE POUR LE CRITERE

---

Pour l'évaluation du critère, c'est-à-dire le rendement global dans un poste de centre d'appels, une évaluation a été effectuée, pour le supérieur hiérarchique de chaque participant, et ce, à partir d'un questionnaire d'évaluation élaboré spécifiquement à cette fin (voir annexe E). Celui-ci était complété en ligne, sur une base volontaire, à l'intérieur de la même période de temps qui a été allouée aux participants afin qu'ils complètent le test.

De façon générale, la question principale s'inspirait du construit de rendement général proposé par Viswesvaran et ses collègues (2005) et discuté précédemment. Lorsque les superviseurs possédaient des données objectives concernant le rendement de l'individu (par exemple, la quantité de ventes, le nombre d'appels moyens, etc.), ils devaient identifier la liste des personnes qui ont obtenu les meilleurs résultats à partir des critères jugés essentiels. Par la suite, ils devaient examiner la liste et déterminer, pour chacune des personnes identifiées, par rapport à des employé(e)s observés par le passé et qui occupent des emplois similaires, si cette personne se situe dans le 3e ou le 4e groupe (voir l'échelle d'évaluation ci-dessous). Si c'est le cas, il était alors possible de déduire que cette personne pouvait être considérée comme ayant un rendement supérieur. En effet, non seulement elle obtient des résultats supérieurs, mais, en plus, elle se démarque par rapport aux compétences jugées essentielles pour le poste. Dans le cas des employés pour lesquels aucune donnée objective n'était disponible, l'échelle d'évaluation en quatre points d'ancrage devait leur permettre d'identifier les individus avec un rendement supérieur en emploi.

Ainsi, en ce qui concerne l'échelle de réponses possibles, elle comportait quatre niveaux, telle que suggérée par Durivage, Pettersen et Thibault (2006). Selon ces auteurs l'évaluation par rang est avantageuse par rapport à l'utilisation d'une échelle de type Likert parce que cette dernière serait sujette à une restriction de la variance significative en contexte d'évaluation du rendement (Durivage et al., 2006; Goffin et Olson, 2011). En effet, il semblerait que les premiers échelons (1-2) d'une échelle de type Likert ne soient pas utilisés dans ce contexte bien précis puisque les employés auxquels pourrait correspondre cette évaluation ne sont pas embauchés ou ont déjà été congédiés (Durivage et al., 2006).

De plus, la signification pratique et concrète de la plupart des échelles de type jugement absolu demeurerait floue pour bon nombre de superviseurs, ce qui pourrait poser un biais à l'évaluation. En

effet, prenons la situation la plus courante en milieu organisationnel qui exige qu'un superviseur évalue ses employés en fonction d'une échelle allant de « Très faible » à « Excellent ». De manière générale, comme l'indiquent Goffin et Olson (2011) en lien avec l'utilisation avec ce type d'échelles : « *verbal anchors and descriptions of performance dimensions have been criticized as poorly defined and responsible for interrater differences in interpretation* » (Goffin et Olson, 2011, p. 50). Ainsi, deux sujets au rendement véritablement équivalent pourraient se voir accorder au final une cote de rendement différent (Adéquat vs. Bon) seulement parce que leur rendement a été évalué par des superviseurs différents ayant chacun leur conception et des attentes différentes de ce que l'on entend par un rendement adéquat vs. un bon rendement.

D'autre part, même avec le développement de nouvelles échelles à ancrages comportementaux (Behaviorally Anchor Rating Scales, BARS) afin d'adresser ces critiques, ces dernières posent également leur lot de problèmes. En effet, Landy et Farr (1980) rapportent qu'il est généralement très difficile d'identifier les ancres centrales de ce type d'échelle et qu'elles demandent habituellement beaucoup de temps, d'efforts et de spécialistes pour en assurer un développement adéquat. Dans le même sens, il semblerait qu'une échelle à ancrages comportementaux n'est valide que pour le contexte d'évaluation précis pour lequel elle a été développée (Landy et Farr, 1980). Qui plus est, les BARS demeuraient également sujets à des erreurs de mesure - erreurs psychométriques de distribution, de halo, de distorsion systématique, de première impression, etc.; (W.C. Borman, 1991). En bref, toujours selon ces auteurs, l'investissement nécessaire en temps et en ressources ne serait que très rarement justifiable: « *Innovations in rating, although plentiful, are likely to result in robbing Peter to pay Paul* » (Landy et Farr, 1980, p. 84).

En résumé, l'évaluation comparative proposée par cette échelle a été retenue pour les avantages qu'elle présente non seulement sur le plan de la pratique, mais également d'un point de vue conceptuel. D'ailleurs, comme le rapporte Longpré (2013), les résultats d'au moins trois méta-analyses, comparant les évaluations du rendement comparatives et absolues, tendent à démontrer que les premières permettent d'obtenir des coefficients de validité critériée corrigés supérieurs aux secondes ( $\rho = 0,66$  vs  $\rho = 0,21$  ;  $\rho = 0,45$  vs  $\rho = 0,35$  et  $\rho = 0,51$  vs  $\rho = 0,34$ )<sup>14</sup>.

---

<sup>14</sup>Basés sur les références suivantes citées dans Longpré (2013) : a) Nathan, B. R., & Alexander, R. A. (1988). A comparison of criteria for test validation: A meta-analytic investigation. *Personnel Psychology*, 41(3), 517-535. b) Heneman, R. L. (1986). The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis. *Personnel Psychology*, 39(4), 811-826. c) Bommer, W. H., Johnson, J., Rich, G. A.,

En effet, malgré que ce mode d'évaluation puisse parfois également conduire à effectuer des comparaisons « boîteuses » basées sur l'utilisation de stéréotypes ou de paramètres différents lors de cet exercice de jugement, il est intéressant de se rapporter aux propos de Goffin et Olson (2011) à ce sujet :

*“Simply put, the biases also occur reliably on absolute measures. (...) It is important to recognize that the phenomena occur because the referent group for the judgment is unspecified, thus allowing respondents to choose their own standards or reference groups, which can introduce systematic bias. In contrast to these errors caused by uncontrolled choices of reference groups, the types of relative measures we advocate rest on the selection and specification of a single, clear comparison group for all respondents.” (Goffin et Olson, 2011, p. 55)*

En tenant compte de ces recommandations, une échelle de rangs a été utilisée en spécifiant le critère de comparaison, c'est-à-dire les autres employés œuvrant dans ce poste (voir tableau 8). Qui plus est, les superviseurs ont également été avisés que les personnes évaluées offrent habituellement un rendement qui est minimalement adéquat. Par conséquent, un score de 1 (25 % inférieur) ne signifie pas nécessairement que la personne ne répond pas aux attentes de l'emploi, mais plutôt qu'elle démontre un moins bon rendement en regard au comportement ou à la dimension évaluée.

**Tableau 8**  
**Échelle d'évaluation du rendement**  
**(Durivage et al., 2006, pp. 3-4)**

<b>1</b>	La personne éprouve régulièrement des difficultés par rapport à ce critère. Si on la compare aux autres employés œuvrant dans ce poste, elle se situe, pour ce critère, dans le quart inférieur (25% inférieur).
<b>2</b>	La personne démontre une performance qui est généralement dans la moyenne par rapport à ce critère. Cependant, elle éprouve parfois des difficultés à cet égard. Par rapport à ce critère, elle se situe dans le 25% immédiatement sous la moyenne des personnes œuvrant dans ce poste.
<b>3</b>	La personne démontre une bonne performance par rapport à ce critère. Elle n'éprouve que très rarement des difficultés à cet égard. Par rapport à ce critère, elle se situe dans le 25% immédiatement au-dessus de la moyenne des personnes œuvrant dans ce poste.
<b>4</b>	La personne démontre une excellente performance par rapport à ce critère. Si on la compare aux autres employés œuvrant dans ce poste, elle se situe, pour ce critère, dans le quart supérieur (25% supérieur).

---

Podsakoff, P. M., & MacKenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology*, 48(3), 587-605.

Finalement, notons que les superviseurs ont été informés que cette démarche d'évaluation du rendement n'entraîne en aucun temps en conflit avec l'évaluation de la performance annuelle exigée par l'organisation, qu'elle demeurera en tout temps confidentielle, qu'en aucun temps les candidats ou l'organisation pourront avoir accès à cette information et que les données transmises ne seront utilisées qu'à des fins statistiques pour l'amélioration de notre connaissance sur les méthodes d'évaluation et de sélection de personnel. Les alphas de Cronbach de ces scores variaient entre 0,86 et 0,90. Notons que tous les participants à l'étude ont été supervisés pendant au moins 6 mois par le répondant qui a rempli le questionnaire d'évaluation.

---

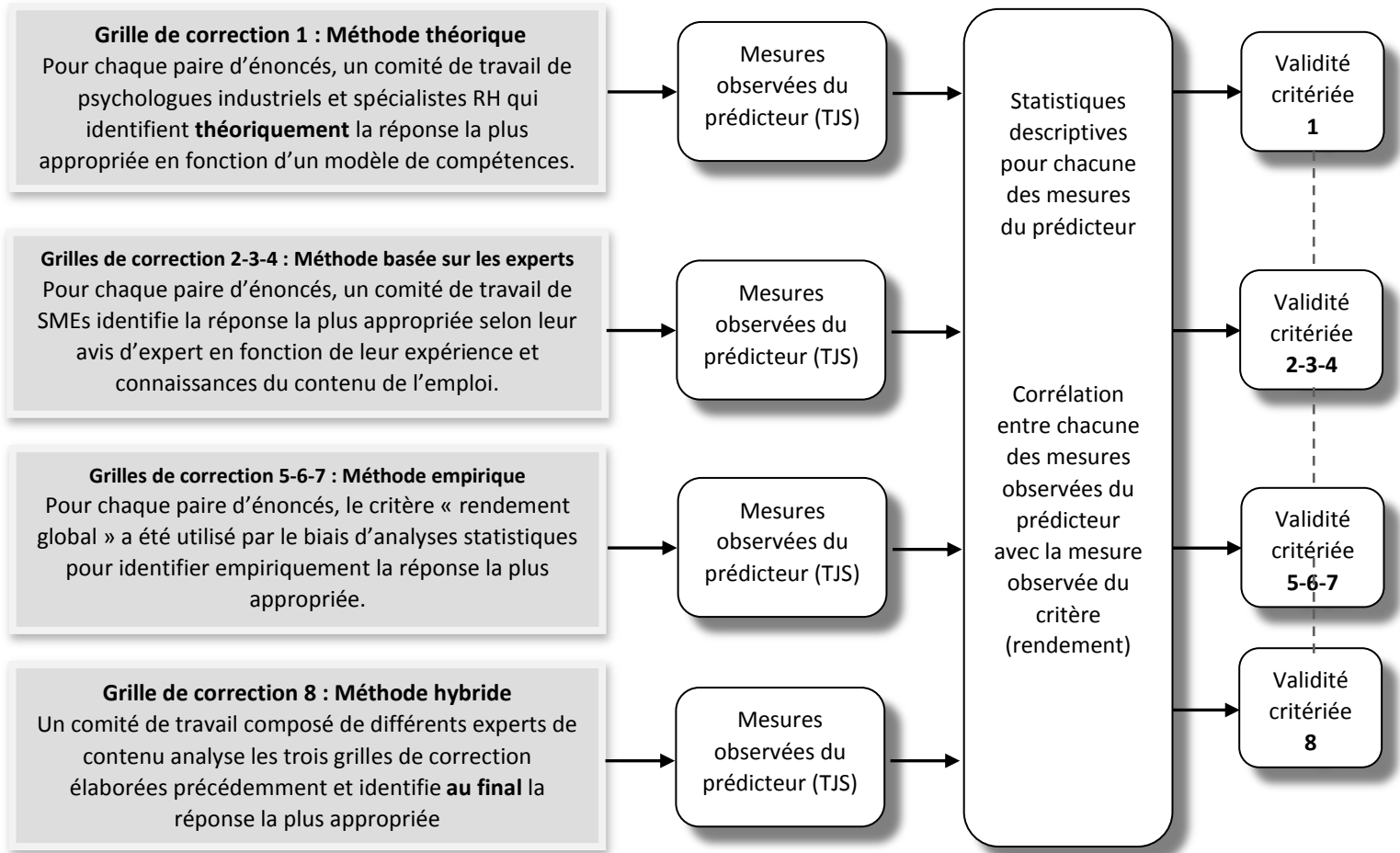
#### 5.4 DEROULEMENT DE L'EXPERIMENTATION

---

Comme mentionné ci-dessus, la collecte de données s'est déroulée selon un devis de recherche concomitant appuyé à partir des données secondaires fournies par la firme Compmetrica. En fait, tel que mentionné précédemment, une organisation ayant participé à la validation du test a identifié plusieurs centaines d'employés occupant un poste dans son centre d'appels afin qu'ils complètent le T3C (voir section 5.1 pour une description détaillée de ce test de jugement situationnel). Ce test a été administré entre janvier 2013 et février 2013 via une plateforme de testing en ligne sécurisée. Comme les conditions d'administration sont prises en charge par la plateforme, des procédures standardisées et des consignes strictes ont été mises en place par l'organisation-participante pour encadrer les passations. Pour leur part, les mesures de rendement en emploi ont été recueillies au cours de la même période, soit entre janvier 2013 et février 2013 dans le cadre d'un projet de validation locale de cet outil d'évaluation par le biais d'un questionnaire développé spécifiquement à cette fin tel que présenté à la section précédente (section 5.3).

La figure 11 présente le schème expérimental utilisé afin de répondre à la question de recherche principale et les hypothèses qui y sont associées.

**Figure 11**  
**Schème expérimental**



## 5.5 ÉCHANTILLON

Les données ont été recueillies dans le contexte d'une étude de validation locale concomitante par le biais de données secondaires entièrement anonymisées. Les participants incluaient 312 personnes provenant d'une grande institution financière québécoise pour lesquels leur superviseur immédiat leur a attribué une cote de rendement global. Ces individus occupaient tous des fonctions de téléopérateurs, et ce, au sein de différents secteurs de l'organisation (services d'assurances, épargnes et placements, cartes de crédit, etc.). En somme, la nature de leur travail consiste donc à la fois à répondre aux besoins de la clientèle tout en leur offrant des produits et services financiers. À titre indicatif, voici quelques-uns des postes occupés par les personnes ayant complété le T3C et pour lesquels des données sur leur

rendement ont été compilées : agent accueil téléphonique, agent service à la clientèle, percepteur, agent assistance routière et voyage, agent télécommercialisation, incluant certains conseillers d'affaires et experts en sinistre.

Les participants n'étaient pas tenus de fournir leurs informations biographiques. C'est pourquoi moins de la moitié des participants n'ont pas répondu à la question concernant leur origine ethnique. Par ailleurs, parmi ceux qui ont mentionné appartenir à un groupe cible (minorité ethnique ou minorité visible) cela représente moins de 13% de l'échantillon total, dont près de la moitié des participants se sont identifiés comme faisant partir de la population « noire ». Cependant, tous les sujets ont indiqué leur genre. Ainsi, il a été comptabilisé que 59% de l'échantillon est de sexe féminin alors que 41% est de sexe masculin. Par ailleurs, 75% de l'échantillon ont indiqué leur tranche d'âge ce qui a permis de constater que 49% de l'échantillon est âgé de 30 ans ou moins, 28% est âgé entre 31 ans et 40 ans alors que seulement 10% de l'échantillon est âgé de plus de 41 ans. Ces proportions, tant sur le plan du genre que pour la catégorie d'âge, sont similaires et représentatives de ce que l'on observe dans l'industrie des centres d'appels canadiens. À preuve, une vaste étude canadienne réalisée en 2007 visant à mieux connaître les pratiques RH dans le domaine des centres d'appels (basés sur l'examen des pratiques dans plus de 400 centres d'appels à travers le pays) a publié des données aussi élevées que 69% concernant la représentativité féminine dans ces unités d'affaires ainsi qu'une main-d'œuvre composée à 74% d'étudiants de niveau collégial ou universitaire (Van Jaarsveld et al., 2007).

---

## 5.6 METHODES D'ANALYSE

---

Lorsque la base de données anonymisée a été transférée aux fins d'analyse, elle contenait les réponses brutes à l'ensemble des questions du test, la cote de rendement qui a été attribuée par le superviseur immédiat ainsi que certains renseignements biographiques (sexe et âge). D'abord, les réponses brutes du candidat ont été recodées, les réponses A ou B indiquant une préférence plus ou moins prononcée pour le premier énoncé recevant une cote de 1 et les réponses C ou D indiquant une préférence plus ou moins prononcée pour le deuxième énoncé recevant une cote de 2 (voir section 5.1 pour de plus amples informations sur le format des réponses du candidat aux questions du TJS). Par la suite, en ce qui concerne la mesure du critère et aux fins des analyses subséquentes, les scores associés à l'échelle de mesure du rendement attribués par le supérieur immédiat ont également été recodés. En

fait, lorsque le candidat avait obtenu un score de 1 ou de 2, il pouvait alors être considéré comme ayant un rendement adéquat (cote de 1) alors que lorsque son supérieur immédiat lui avait attribué un score de 3 ou 4, il était considéré comme ayant un rendement supérieur (cote de 2).

Dans le même sens, des analyses devaient permettre de comparer les différences observées entre les « bonnes réponses » identifiées par les différentes grilles de correction. Pour ce faire, une matrice de corrélation a été produite afin de déterminer le degré de convergence entre chacune des grilles de correction. Dans la même optique, le pourcentage des items du test pour lesquels des réponses différentes ont été identifiées a été calculé en comparant chacune des grilles de correction entre elles. Les résultats de ces analyses sont présentés à la section 6.1.

Des analyses descriptives (fréquence, moyenne, médiane, mode, écart-type et indices d'asymétrie et d'aplatissement) ont été effectuées pour chacune des mesures du prédicteur en fonction de l'application des différentes grilles de correction afin de vérifier les mesures de tendance centrale, de dispersion des variables étudiées et de permettre de répondre à la première question de recherche, c'est-à-dire quelle méthode visant l'élaboration d'une grille de correction d'un TJS permet d'en arriver à une meilleure distribution des scores obtenus par les participants. Dans cette même perspective, des histogrammes de distribution des fréquences et des diagrammes quantile-quantile ont été utilisés pour comparer les distributions observées les unes avec les autres.

Finalement, dans le but de tester la deuxième série d'hypothèses, c'est-à-dire la relation entre la performance obtenue au prédicteur selon chacune des grilles de correction et le rendement, tel qu'évalué par le supérieur immédiat, des analyses de corrélations linéaires ont été menées. Ces analyses ont permis de déterminer s'il existe des liens significatifs entre ces variables (un seuil de  $p < .05$  a été retenu). Pour ce faire, une corrélation point bisériale (*r.bis*) a été utilisée étant donné que la relation impliquait une variable dichotomique (rendement adéquat ou supérieur) d'une part et une variable quantitative continue d'autre part (le résultat obtenu au TJS dont les valeurs sont considérées comme des points séparés). De façon plus spécifique, tel que recommandé par Cohen (1983) et cité dans le manuel de Bobko (2001), les corrélations bisériales sont généralement recommandées afin de prendre en considération une certaine perte d'information qui affecte les données lorsqu'une variable continue est artificiellement dichotomisée : « *The biserial correlation attempts to estimate the true underlying population correlation between two continuous variables when the sample data for one variable have*



*been artificially dichotomized* » (Bobko, 2001, p. 38). Mathématiquement équivalente au coefficient de corrélation de Pearson, cet indice varie de -1 à +1 et permet de vérifier la linéarité d'une relation entre deux variables. Les intervalles de confiance ont également été calculés et ont été établis à 95% à l'intérieur desquels la corrélation réelle se situera 19 fois sur 20.

Pour tenir compte d'éventuelles erreurs associées à l'atténuation de la fidélité pour le critère, une mesure de correction a également été effectuée (Hunter, Schmidt et Le, 2006). Malgré qu'il n'existe pas de consensus sur la meilleure façon de corriger l'erreur de mesure reliée à la fidélité des évaluations effectuées par un superviseur, plusieurs auteurs considèrent que la fidélité interexamineurs est le moyen le plus approprié d'évaluer l'impact des erreurs de mesure de l'évaluation du rendement faite par un superviseur (Vinchur, Schippmann, Switzer et al., 1998). En effet, dans le contexte de l'évaluation de la performance globale d'un employé, cet indice de concordance entre deux (ou plusieurs) examineurs donne des renseignements sur les erreurs de mesures qui pourraient être occasionnés par les caractéristiques personnelles d'un évaluateur, par exemple. Dans la présente recherche, tel que proposé par Viswesvaran et ses collègues, un coefficient de 0,52 sera utilisé pour effectuer cette deuxième correction pour compenser l'erreur de mesure due à la fidélité du critère (Viswesvaran et al., 1996).

Finalement, des vérifications supplémentaires ont été réalisées pour compléter les analyses corrélationnelles par le biais d'une analyse de variance et par l'examen des différences de moyennes standardisées.

## PARTIE 4 : RESULTATS

---

La section qui suit décrit la façon dont la recherche a été opérationnalisée ainsi que les analyses statistiques qui ont été réalisées sur les données recueillies afin de répondre aux questions de recherche énoncées précédemment.

---

### 6.1 ANALYSES PRELIMINAIRES

---

Dans un premier temps, chacun des scores totaux a été vérifié par le biais des fréquences par rapport au nombre de questions répondues par les participants. Aucune donnée manquante n'a été identifiée ce qui s'explique par le fait que le test était administré via une plateforme informatique qui a contrôlé l'administration de l'examen et qui enregistrait chacune des réponses en temps réel en plus de s'assurer que le candidat avait répondu à l'ensemble des questions avant de mettre fin au test.

#### 6.1.1 CONVERGENCE ENTRE LES GRILLES DE CORRECTION

---

En premier lieu, une matrice de corrélations a été produite afin de vérifier le degré de convergence entre les grilles de correction. Le tableau 9 présenté à la page suivante indique jusqu'à quel point les résultats obtenus en fonction de chacune des grilles de correction sont reliés les unes aux autres.

**Tableau 9**  
**Matrice de corrélations entre les grilles de correction**

	Grille 1 Théorique	Grilles experts			Grilles empiriques			Grille 8 Hybride
		Grille 2 Experts financiers	Grille 3 Experts Énergie	Grille 4 Experts secteur public	Grille 5 Empirique (6)	Grille 6 Empirique (19)	Grille 7 Empirique (28)	
Grille 1 Théorique	1	,759**	,830**	,185**	-,200**	,754**	,324**	,862**
Grille 2 Experts Financiers		1	,826**	,035	-,029	,752**	,480**	,830**
Grille 3 Experts Énergie			1	,141*	-,166**	,764**	,451**	,932**
Grille 4 Experts Public				1	-,08	,255**	,014	,047
Grille 5 Empirique 6					1	-,14*	,44**	-,156**
Grille 6 Empirique 19						1	,23**	,788**
Grille 7 Empirique 28							1	,472**
Grille 8 Hybride								1

\*p ≤ 0,05 \*\* p ≤ 0,01

Dans l'ensemble, on remarque que la majorité des grilles de correction sont corrélées positivement les uns aux autres, ce qui n'est pas surprenant en soi. En effet, il est logique de croire que des groupes d'experts en charge d'évaluer la pertinence de certains comportements reliés au service à la clientèle pourraient s'inspirer en partie des meilleures pratiques reliées à ce domaine d'activité professionnelle.

De façon plus particulière, on note que les corrélations entre la grille théorique et les grilles élaborées à l'aide d'experts de contenu sont relativement élevées (0.759\*\* pour les experts financiers et 0.830\*\* pour les experts du domaine de l'énergie). Par contre, on note que la corrélation obtenue avec la grille créée par les experts du secteur public est positive, mais relativement faible (0,185\*\*). Finalement, on observe que les corrélations entre la grille théorique et les grilles empiriques varient considérablement selon la grille avec laquelle elle est mise en relation (-0,200 ; 0,324\*\* ; 0,754).

De leur côté, les grilles créées par le biais de l'approche basée sur l'implication d'experts de contenu présentent soit une corrélation élevée (0,826\*\*) entre celles établies par les experts du domaine financier et ceux de l'énergie ou soit des corrélations plutôt faibles (0,035 entre les experts du domaine public et ceux du secteur financier et 0,141 entre les experts du domaine public et ceux du domaine de l'énergie). Tout comme pour la grille théorique, on observe également que les corrélations entre les grilles créées par des experts et par le biais de l'approche empirique varient considérablement selon la grille avec laquelle elle est mise en relation (entre -0,166 et 0,764).

Finalement, on observe que la corrélation entre les résultats obtenus par le biais de la grille hybride et ceux établis théoriquement est élevée (0,862\*\*), tout comme les corrélations avec les résultats associés aux experts financiers (0,830\*\*) ou à ceux de l'énergie (0,932\*\*). Dans le cas de la grille établie par les experts du domaine public, on observe une corrélation relativement faible (0,047). Finalement, les corrélations observées avec les grilles empiriques varient fortement selon la grille utilisée (de -0,156\*\* à 0,788\*\*).

#### 6.1.2 VARIABILITES ENTRE LES GRILLES DE CORRECTION

---

Le tableau 10 présenté à la page suivante compare les différences observées entre les « bonnes réponses » identifiées par les différentes grilles de correction. De façon plus spécifique, le tableau 10 présente le pourcentage des items du test pour lesquels des réponses différentes ont été identifiées. Par exemple, on observe que dans 30 % des cas (24 questions sur 80), les experts financiers (grille 1) ont identifié une « bonne réponse » différente de celle choisie par les experts théoriques (grille 2). De façon générale, on constate qu'en moyenne 32 % des « bonnes réponses » identifiées variaient en fonction du groupe d'experts impliqué.

En ce qui a trait aux grilles empiriques, on observe qu'en moyenne 48 % des « bonnes réponses » variaient, lorsque comparées à celles identifiées par les groupes d'experts. Lorsque comparées les unes aux autres, les grilles empiriques (grille 6, grille 19 et grille 28) variaient en moyenne de 52 %. Finalement, on constate que les experts financiers (grille 1) ont modifié plus du tiers de leurs réponses originales (35 %) suite à l'examen des informations fournies lors de la création de la grille hybride (grille 8). La moyenne des différences entre la grille hybride et les grilles empiriques s'est établie à 46 %.

Ces analyses préliminaires nous permettent donc de confirmer que les variations sont suffisamment importantes pour soutenir les objectifs et questions de la présente recherche.

**Tableau 10**  
**Pourcentages de différences entre les « bonnes réponses »**

	Grilles experts				Grilles empiriques			Grille 8 Hybride
	Grille 1 Théorique	Grille 2 Experts financiers	Grille 3 Experts Énergie	Grille 4 Experts secteur public	Grille 5 Empirique (6)	Grille 6 Empirique (19)	Grille 7 Empirique (28)	
Grille 1 Théorique	-	30 %	28 %	31 %	43 %	44 %	56 %	25 %
Grille 2 Financiers		-	33 %	36 %	45 %	49 %	54 %	35 %
Grille 3 Énergie			-	36 %	40 %	44 %	59 %	15 %
Grille 4 Public				-	50 %	46 %	41 %	36 %
Grille 5 Empirique 6					-	44 %	46 %	41 %
Grille 6 Empirique 19						-	58 %	43 %
Grille 7 Empirique 28							-	55 %
Grille Hybride								-

### 6.1.3 ANALYSES DESCRIPTIVES LIEES AU CRITERE

En ce qui concerne le critère, c'est-à-dire le rendement global des individus (appréciation globale de la personne dans son milieu de travail), il a été mesuré par une seule question (voir annexe E). Tel que mentionné à la section précédente, les réponses ont été recodées afin de créer deux groupes, soit les individus ayant été identifiés comme ayant un rendement adéquat (1) et ceux ayant été identifiés comme ayant un rendement supérieur en emploi (2).

Les analyses descriptives reliées à l'étendue et à la moyenne du critère sont présentées au tableau 11. Dans l'ensemble, on remarque que la moyenne s'établit à 1,49 avec un écart type de 0,50.

**Tableau 11**  
**Statistiques descriptives pour le critère**

<u>Critère</u>	<u>N</u>	<u>Min</u>	<u>Max</u>	<u>M</u>	<u>É-T</u>
Rendement	312	1	2	1,49	0,50

De façon plus spécifique, afin de comparer plus facilement les deux catégories d'individus en fonction de la variable du critère, le tableau 12 présente la répartition observée. Alors que 160 personnes font partie du premier groupe (rendement adéquat), 152 individus ont été identifiés comme ayant un rendement supérieur en emploi, pour une répartition quasi équitable entre les deux catégories de variables (51,3% pour la première et 48,7% pour la deuxième).

**Tableau 12**  
**Répartition de l'échantillon en fonction de la mesure du critère**

<u>Variables</u>	<u>N</u>	<u>% de l'échantillon</u>
Rendement adéquat	160	51,3%
Rendement supérieur	152	48,7%

---

## 6.2 VERIFICATIONS DES HYPOTHESES

---

Dans cette section seront traités plus spécifiquement les résultats obtenus afin de répondre aux deux questions de la présente recherche énoncées précédemment.

### 6.2.1 ANALYSES DESCRIPTIVES LIEES AU PREDICTEUR

---

Des analyses descriptives liées au prédicteur en fonction de chacune des variables indépendantes du schème expérimental permettent d'abord de répondre à la première question de recherche. Pour ce faire, ces analyses ont été effectuées principalement pour vérifier les mesures de tendances centrales et de dispersion des variables étudiées et ainsi répondre à la première question de recherche, c'est-à-dire les scores minimum et maximum, la moyenne (en fonction du nombre d'items et converti en %), l'écart type, l'indice d'asymétrie et de kurtose, et ce, pour chacune des grilles de

correction. De plus, les coefficients de cohérence interne pour chacune des grilles de correction ont été calculés. Le tableau 13 présente l'ensemble de ces résultats.

**Tableau 13**  
**Statistiques descriptives pour chacune des grilles de correction**

<b>Méthode théorique</b>	<b>N</b>	<b>#items</b>	<b>Min</b>	<b>Max</b>	<b>M</b>	<b>M%</b>	<b>É-T</b>	<b>Asy.</b>	<b>Kurt.</b>	<b>α</b>
Grille 1	312	80	29,00	63,00	47,43	59,3%	6,80	-0,18	-0,39	0,51
<b>Méthode Experts</b>	<b>N</b>	<b>#items</b>	<b>Min</b>	<b>Max</b>	<b>M</b>	<b>M%</b>	<b>É-T</b>	<b>Asy.</b>	<b>Kurt.</b>	<b>α</b>
Grille 2 « financier »	312	80	28,00	60,00	44,53	55,7%	6,48	0,07	-0,41	0,52
Grille 3 « énergie »	312	80	25,00	61,00	46,80	58,5%	7,32	-0,22	-0,55	0,63
Grille 4 « public »	312	80	28,00	63,00	45,49	56,9%	6,13	-0,00	-0,08	0,51
<b>Méthode Empirique</b>	<b>N</b>	<b>#items</b>	<b>Min</b>	<b>Max</b>	<b>M</b>	<b>M%</b>	<b>É-T</b>	<b>Asy.</b>	<b>Kurt.</b>	<b>α</b>
Grille 6	312	80	22	57,00	40,00	50,0%	6,20	-0,12	0,12	0,53
Grille 19	312	80	23	65,00	45,23	56,5%	8,12	-0,04	-0,30	0,73
Grille 28	312	80	20	59,00	42,93	53,7%	6,91	-0,22	0,12	0,62
<b>Méthode Hybride</b>	<b>N</b>	<b>#items</b>	<b>Min</b>	<b>Max</b>	<b>M</b>	<b>M%</b>	<b>É-T</b>	<b>Asy.</b>	<b>Kurt.</b>	<b>α</b>
Grille 8	312	80	22,00	70,00	48,03	60,0%	9,78	-0,29	-0,58	0,68

Dans l'ensemble, on remarque que les moyennes générées par les différentes grilles de correction varient entre 50% (grille empirique 6) et 60% (grille hybride). Par ailleurs, les indices d'asymétrie et de kurtose démontrent la normalité de la distribution de ces résultats (score compris entre -1 et 1). À cet effet, l'annexe F présente graphiquement la distribution des résultats pour la population étudiée, et ce, pour l'ensemble des grilles de correction.

Finalement, les coefficients de consistance interne (alpha) sont présentés pour chacune des grilles. Dans l'ensemble, on observe qu'ils se situent entre 0,51 et 0,73, un niveau supérieur à ce que l'on a observé dans la littérature portant sur les TJS. En effet, tel que mentionné précédemment, l'étude de Poyart et Enhart (2003) indiquant que lorsque le TJS demande aux participants de choisir une seule

bonne réponse parmi celles présentées, les coefficients de consistance interne étaient généralement plus faibles si situant autour de 0,24.

En ce qui a trait à l'**hypothèse 1a**, cela permet d'abord d'attester que la distribution des résultats pour les huit grilles de correction suit une distribution normale. En effet, à l'analyse des histogrammes présentés à l'annexe F, on note qu'environ 67% à 73% des participants se retrouvent à -1 et +1 écart-type de la moyenne (alors que la normalité correspond à 68,26 % de l'échantillon qui se retrouve à cette même distance de la moyenne). Notons cependant que la grille de correction hybride produit une courbe légèrement asymétrique (plus à droite) que les autres méthodes d'élaboration des grilles de correction. La grille de correction élaborée par les experts publics se rapproche davantage d'une courbe normale et présente un indice d'asymétrie plus près du 0. Ces résultats sont également corroborés par les diagrammes quantile-quantile qui permettent de comparer visuellement les distributions entre elles et de vérifier la corrélation entre leur distribution et la normalité (voir annexe G). En fait, à l'observation de ces diagrammes, on note que, mise à part la grille produite par les experts du secteur public, l'approche empirique produit une distribution des résultats qui est normale, et ce, peu importe la grille de correction appliquée.

Par ailleurs, toutes les moyennes obtenues au prédicteur se situent entre 50% et 60%. Dans la mesure où les personnes faisant partie de l'expérimentation occupaient déjà un emploi de ce type, il est normal que les résultats soient un peu plus élevés que le point milieu (50%). Par ailleurs, on note que les grilles de correction établies de manière empirique se rapprochent beaucoup plus d'une moyenne à 50%, alors que la méthode hybride mène à la moyenne la plus élevée. Cette différence est relativement importante (jusqu'à 10%) et indique que la méthode privilégiée a un impact majeur sur les résultats obtenus par les participants. Le fait que la méthode hybride s'appuie à la fois sur des éléments théoriques et opérationnels (c'est-à-dire qu'ils reflètent les façons de faire privilégiées par l'organisation) peut expliquer le fait que les participants, qui œuvrent déjà au sein de l'organisation aient obtenu des résultats plus élevés, ces derniers étant déjà au fait des attentes associées à leur poste. À l'inverse, les grilles de correction basées sur l'approche empirique, qui ne tient pas nécessairement compte de ces aspects, produisent des résultats qui sont généralement plus faibles (50,0%, 53,7% et 56,5%).

En somme, les résultats de ces analyses descriptives supportent en partie l'**hypothèse 1a** qui stipulait que la méthode hybride utilisée pour élaborer la grille de correction d'un TJS conduirait à une



meilleure distribution des scores des participants en présentant une distribution symétrique des résultats ainsi qu'une moyenne obtenue à l'exercice de 60%. Si la méthode hybride n'a pas démontré de supériorité par rapport aux autres grilles en ce qui a trait au niveau de normalité, elle a tout au moins produit une moyenne de 60%, qui correspond exactement aux attentes stipulées par St-Sauveur et ses collègues (2014).

## 6.2.2 ANALYSES CORRELATIONNELLES ENTRE PREDICTEUR ET CRITERE

L'analyse bivariée permet de mesurer les relations entre les résultats obtenus au TJS selon chacune des grilles de correction élaborée avec le rendement observé en emploi. Cette analyse permet alors de vérifier la série d'hypothèses associées à la deuxième question de ce projet de recherche. Le tableau 14 présente les résultats de ces analyses. Les diagrammes de dispersion permettant de visualiser le lien entre les variables et d'attester de la présence d'une relation linéaire sont présentés à l'annexe H.

**Tableau 14**  
**Corrélation entre les scores obtenus à chacune des grilles et le rendement**

	<b>Grille 1 Théorique</b>	<b>Grille 2 Experts financiers</b>	<b>Grille 3 Experts Énergie</b>	<b>Grille 4 Experts Public</b>	<b>Grille 5 Empirique (6)</b>	<b>Grille 6 Empirique (19)</b>	<b>Grille 7 Empirique (28)</b>	<b>Grille 8 Hybride</b>
<b>N</b>	312	312	312	312	252	252	252	252
<b>Corrélation brute</b>	,162**	,164**	,222**	,132*	0,022	0,154*	0,278**	,247**
<b>Corrélation corrigée</b>	<b>0,31</b>	<b>0,32</b>	<b>0,38</b>	<b>0,26</b>	<b>0,04</b>	<b>0,25</b>	<b>0,49</b>	<b>0,42</b>

\* $p \leq 0,05$  \*\*  $p \leq 0,01$

Il est à noter que pour mesurer les corrélations entre les grilles de correction empiriques et le rendement en emploi, les données des participants qui ont été utilisées afin de déterminer les bonnes réponses (soit les 30 employés performants et les 30 employés adéquats) ont été exclues. Dans le même sens, comme le comité de travail a pris en considération la grille de correction empirique (19) lors de l'élaboration de la grille de correction hybride, les données qui ont été utilisées pour déterminer les

bonnes réponses de cette grille ont été exclues des analyses corrélationnelles. Tel qu'expliqué précédemment, ces corrélations ont été corrigées par la suite afin de tenir compte de l'atténuation dans le critère.

Dans l'ensemble, on observe une certaine variabilité dans les coefficients de validité critériée. En fait, la corrélation brute entre les résultats obtenus au TJS et le rendement en emploi passe de 0,132\* (experts public) à 0,247\*\* (approche hybride) selon l'approche adoptée pour élaborer la grille de correction (on ne retient ici que la grille 19 pour la méthode empirique). Toujours selon cette même prémisse, on note que l'approche permettant d'obtenir le coefficient de corrélation le plus élevé est celle qui permet aux experts de prendre leurs décisions en fonction d'un maximum d'informations (méthode hybride) confirmant ainsi les **hypothèses 2a, 2b et 2c**.

En effet, ces analyses corrélationnelles indiquent qu'une approche multiple d'élaboration d'une grille de correction d'un TJS (méthode hybride) permet de mieux prédire le rendement des employés ( $r=0,247^{**}$ ) comparativement à l'utilisation de la théorie pour définir les attentes ( $r=0,162^{**}$ ), le recours à des experts de contenu œuvrant dans le domaine de pratique ( $r=0,164^{**}$ ) ou l'utilisation d'une approche empirique (résultats variant entre  $r=0,022$  et  $r=0,278^{**}$ ). Par ailleurs, il est intéressant de noter que l'ensemble de ces corrélations se sont avérées significatives. D'ailleurs, lorsqu'on corrige pour tenir compte de l'erreur de mesure associée à la fidélité du critère, les coefficients de corrélation corrigés vont également dans le même sens ( $r=0,42$  pour la méthode hybride vs.  $r=0,31$  pour l'approche théorique,  $r=0,32$  pour les experts du domaine et des résultats de  $r=0,25$  pour l'utilisation d'une approche empirique).

Finalement, les intervalles de confiance ont été calculés pour l'ensemble de ces corrélations (voir tableau 15). Ainsi, non seulement l'augmentation du score obtenu au TJS est en relation directe (et positive) avec le rendement en emploi observé par les supérieurs immédiats, mais les intervalles de confiance inférieurs et supérieurs viennent également supporter le fait que la méthode hybride présente des corrélations supérieures aux autres méthodes d'élaboration en situant la corrélation réelle 19 fois sur 20 entre 0,131 et 0,363.

**Tableau 15**  
**Intervalles de confiance des corrélations**

	<b>Grille 1 Théorique</b>	<b>Grille 2 Experts financiers</b>	<b>Grille 3 Experts Énergie</b>	<b>Grille 4 Experts Public</b>	<b>Grille 5 Empirique (6)</b>	<b>Grille 6 Empirique (19)</b>	<b>Grille 7 Empirique (28)</b>	<b>Grille 8 Hybride</b>
<b>N</b>	312	312	312	312	252	252	252	252
<b>Intervalle de confiance supérieur</b>	0,054	0,056	0,116	0,023	-0,102	0,029	0,166	0,131
<b>Intervalle de confiance inférieur</b>	0,270	0,272	0,328	0,241	0,146	0,271	0,394	0,363

---

## 6.3 ANALYSES STATISTIQUES SUPPLEMENTAIRES

---

Afin de valider les tendances observées lors du traitement des résultats présentés précédemment, des analyses statistiques supplémentaires ont été réalisées.

### 6.3.1 COMPARAISON INTER-GROUPE

---

Une analyse de variance a été effectuée afin de vérifier si les moyennes de résultats obtenus par le groupe d'employés performants étaient significativement supérieures à celles obtenues par les employés considérés comme adéquats. L'objectif de cette analyse supplémentaire était de vérifier si la prédiction du rendement futur en emploi pouvait être attribuable aux résultats obtenus au TJS ou s'il s'agit plutôt d'un effet du hasard. Le tableau 16 présente les résultats de cette analyse de variance (voir page suivante). Il est à noter que les participants qui ont été utilisés pour la création des trois grilles de correction empiriques (grille no 6, grille no 19 et grille 28) ont été éliminés de l'échantillon pour effectuer ces analyses supplémentaires.

**Tableau 16**  
**Analyse de variance en fonction du rendement et des différentes grilles de correction**

	Sommes des carrés		dl	Carré moyen	F	Sign
<b>Grille 1 Théorique</b>	Entre groupes	517,475	1	517,475	12,050	0,001
	Dans le groupe	6957,086	162	42,945		
	Total	7474,561	163			
<b>Grille 2 Experts financiers</b>	Entre groupes	325,399	1	325,399	8,491	0,004
	Dans le groupe	6208,650	162	38,325		
	Total	6534,049	163			
<b>Grille 3 Experts Énergie</b>	Entre groupes	915,944	1	915,944	19,647	0,000
	Dans le groupe	7552,367	162	46,620		
	Total	8468,311	163			
<b>Grille 4 Experts Publics</b>	Entre groupes	86,138	1	86,138	2,411	0,122
	Dans le groupe	5788,423	162	35,731		
	Total	5874,561	163			
<b>Grille 5 Empirique (6)</b>	Entre groupes	1,168	1	1,162	0,030	0,862
	Dans le groupe	6211,094	162	38,640		
	Total	6212,262	163			
<b>Grille 6 Empirique (19)</b>	Entre groupes	729,359	1	729,359	12,023	0,001
	Dans le groupe	9827,635	162	60,664		
	Total	10556,994	163			
<b>Grille 7 Empirique (28)</b>	Entre groupes	631,393	1	631,393	11,996	0,001
	Dans le groupe	8526,851	162	52,635		
	Total	9158,244	163			
<b>Grille 8 Hybride</b>	Entre groupes	1972,716	1	1972,716	22,690	0,000
	Dans le groupe	14084,424	162	86,941		
	Total	16057,140	163			

Dans tous les cas on constate que chaque grille de correction produit un résultat moyen supérieur pour les employés démontrant un rendement supérieur en emploi, lorsque comparé au score moyen de ceux qui ont obtenu une cote de rendement adéquat. Appuyant les analyses corrélationnelles, l'analyse de variance indique que, mises à part la grille élaborée par les experts du secteur public et la grille empirique no 6 qui était faiblement corrélée avec le rendement en emploi, les moyennes observées entre les deux groupes d'employés sont significativement différentes, peu importe les autres méthodes adoptées pour l'élaboration de la grille de correction. Ici, plus la différenciation entre les groupes est grande, meilleure est la validité du test. Cette information supplémentaire vient donc en appui à la vérification de l'hypothèse selon laquelle la méthode de l'approche hybride est la meilleure approche permettant l'élaboration de la grille de correction d'un TJS (le F observé étant le plus élevé).

Lorsque l'on s'attarde à l'analyse des scores moyens bruts obtenus aux différents exercices, on remarque au tableau 16 que les différences de moyennes produites par les différentes grilles de correction supportent la supériorité de la méthode hybride dans la mesure où cette dernière produit une différence de moyenne plus élevée (9,7% d'écart en comparaison avec 4,5% d'écart pour l'approche théorique et les experts du domaine et de 5,1% pour l'approche empirique ayant une relation moyenne avec le rendement - grille 19).

Finalement, l'indice  $d$  de Cohen permet d'apprécier les différences de moyennes standardisées entre les groupes et de caractériser la magnitude d'un effet associé dans une population donnée par rapport à une hypothèse nulle. Le tableau 17 présente les résultats de l'indice  $d$  obtenu pour chacune des méthodes d'élaboration des grilles de correction. Traditionnellement, un indice  $d$  entre 0,10 et 0,30 est décrit comme ayant un effet « faible », alors qu'un indice  $d$  entre 0,31 et 0,50 est qualifié comme un ayant un effet « moyen » et un indice  $d$  supérieur à 0,51 est qualifié de « fort » (Cohen, 1992). Il est intéressant de constater que ces analyses supplémentaires indiquent que seule la méthode hybride produit un effet considéré comme étant « fort » lorsque l'on exclut la grille 28 produite par la méthode empirique.

**Tableau 17**  
**Différences de moyennes standardisées**

	N	Rendement adéquat		Rendement supérieur		Indice $d$
		<u>M</u>	<u>É-T</u>	<u>M</u>	<u>É-T</u>	
<b>Grille 1 Théorique</b>	312	46,4	7,22	48,6	6,16	0,33
<b>Grille 2 Experts financiers</b>	312	43,5	6,40	45,6	6,42	0,33
<b>Grille 4 Experts Énergie</b>	312	45,2	7,79	48,5	6,40	0,46
<b>Grille 4 Experts Publics</b>	312	44,7	6,42	46,3	5,72	0,26
<b>Grille 5 Empirique (6)</b>	252	40,13	6,17	40,38	5,62	0,04
<b>Grille 6 Empirique (19)</b>	252	43,66	7,83	46,02	7,42	0,31
<b>Grille 7 Empirique (28)</b>	252	40,77	7,09	44,81	6,90	0,58
<b>Grille 8 Hybride</b>	252	45,5	10,35	50,4	8,44	0,52

\* $p \leq 0,05$  \*\*  $p \leq 0,01$

En somme, à la lecture des résultats présentés à cette section, on constate d'abord qu'il peut exister des variations importantes dans les grilles de correction associées à un TJS, alors que peu plus du tiers des réponses identifiées peuvent varier en fonction de la composition du comité de validation. On observe également que la validité critériée varie grandement en fonction de la stratégie privilégiée lors de l'élaboration de la grille de correction. Il existe également une très grande variabilité dans la qualité de l'effort de prédiction par le biais de la méthode empirique qui produit à la fois le coefficient de corrélation le plus fort, mais également celui le plus bas. Finalement, on note que l'approche permettant d'obtenir le coefficient de corrélation le plus intéressant est celle qui permet aux experts de prendre leurs décisions en fonction d'un maximum d'informations (méthode hybride). Les implications de ces résultats seront discutées dans le cadre de la section suivante.

## PARTIE 5 : DISCUSSION

---

Cette recherche s'est penchée sur l'influence des stratégies associées à la détermination des grilles de correction de TJS dans un processus de sélection réalisé dans un centre d'appels. En effet, par le biais des instruments de sélection à la disponibilité des responsables de l'embauche dans ce secteur, les tests de jugement situationnel sont de plus en plus considérés comme un outil présentant de nombreux avantages. Or, mettre en œuvre un processus de sélection implique de prendre des décisions valides, justes et équitables. Dans le même sens, utiliser un instrument d'évaluation des aptitudes pour discriminer les candidats entre eux doit incidemment donner une mesure utile, c'est-à-dire valide et fidèle du rendement futur en emploi.

Dans cette perspective, plusieurs auteurs ont souligné la nécessité d'approfondir la recherche conceptuelle et empirique sur les TJS (Bergman et al., 2006; Sackett et Lievens, 2008). L'objectif de cette expérimentation était donc de renseigner sur l'efficacité relative des différentes approches disponibles lors du développement de la grille de correction d'un TJS concernant 1) la distribution des scores des participants et 2) l'effort de prédiction du rendement en emploi de chacune des méthodes. À l'évidence, il peut être très coûteux et difficile de faire appel à des experts de contenu lorsque l'on développe un nouveau test et la valeur ajoutée de combiner différentes méthodes lors de l'élaboration de la grille de correction n'est peut-être pas suffisante. Dans le même sens, il peut être facile pour les candidats de déceler les bonnes réponses découlant d'un ouvrage de référence spécifique au domaine d'activités n'attribuant pas en retour de la valeur à cette méthode d'élaboration d'une grille de correction aux processus RH d'une organisation. Par ailleurs, il n'est pas toujours aisé de recueillir suffisamment de données en lien avec un critère de référence représentatif et suffisamment stable avant le déploiement d'un nouveau test dans le but d'utiliser une méthode empirique pour élaborer la grille de correction d'un TJS. En somme, les implications théoriques de ce projet de recherche concernent d'abord l'aspect conceptuel et prédictif des TJS, mais également ses considérations opérationnelles.

À l'aide d'un TJS visant le contexte des centres d'appels et validé par une grande institution financière, cette recherche a donc mis à contribution des données de sources secondaires en y appliquant un schème expérimental comprenant plusieurs grilles de correction. À ce compte, huit différentes grilles de correction ont été développées afin d'y mettre en relation les résultats obtenus par des téléopérateurs pour lesquels des données réelles de rendement ont également été recueillies. Ce

chapitre propose donc de revenir sur une analyse critique des principaux résultats obtenus. Par ailleurs, une discussion sur les limites méthodologiques et des pistes de recherches futures seront également présentées.

## 6. ANALYSE CRITIQUE

---

Malgré qu'il soit plutôt difficile de conclure, hors de tout doute, du caractère supérieur d'une méthode sur les autres par le biais d'un seul TJS et d'un seul échantillon, certaines pistes de réflexion se dégagent de cette recherche.

Premièrement, on note que la grille de correction empirique moyenne (grille 19) présente une bonne corrélation avec la grille créée à partir des aspects théoriques ( $r=0,754^{**}$ ). Il en va de même pour la grille élaborée par les experts financiers ( $r=0,759^{**}$ ). Ces résultats sont intéressants dans la mesure où ces liens entre les différentes approches sont beaucoup plus élevés que ceux observés par Muros (2008) dans une recherche similaire portant sur le TJS. En fait, ce dernier avait trouvé que le lien produit entre la grille validée par le biais d'une approche théorique était faiblement corrélé à celle déterminée par les experts de contenu ( $r=0,31$ ) alors que la méthode empirique avait produit une grille de correction très peu corrélée aux bonnes pratiques ( $r=0,09$ ).

Les résultats de la présente recherche apportent un éclairage additionnel sur cette question. En effet, il a été observé que la corrélation entre les grilles de correction empiriques et théorique variaient considérablement d'une grille empirique à l'autre. À preuve, on observe que la corrélation entre la grille empirique 6 (la moins prédictive) et la grille basée sur la théorique est négative ( $r=-0,200^{**}$ ) alors que la grille de correction empirique 28 (la plus prédictive) présente un niveau de corrélation modéré ( $r=0,324^{**}$ ) avec cette dernière. Il semble donc que la sélection des sujets qui font partie de l'échantillon à partir duquel sont déterminées les bonnes réponses influencera en retour la qualité de la relation avec le rendement futur. Ce biais d'échantillonnage peut possiblement expliquer la différence observée entre les résultats de la présente étude (avec la grille empirique ayant une prédiction modérée) et ceux rapportés par Muros (2008).

Compte tenu des variations observées, il semble que l'utilisation isolée d'une approche empirique lors de l'élaboration de la grille de correction d'un TJS pourrait être à l'origine de décisions



malheureuses dans les organisations. En effet, si l'institution financière ayant participé à la présente étude avait déterminé les bonnes réponses au TJS avec la grille de correction empirique 6 (la moins prédictive), elle aurait alors pris des décisions d'embauche ou de rejet à partir d'un protocole où la relation avec le rendement futur était presque nulle ( $r=0,022$ ), et ce, sans qu'il y ait de lien direct avec les construits théoriques sous-jacents aux comportements attendus ( $r= -0,200^{**}$ ). Bien sur, d'autres grilles générées à partir de l'approche empirique produisent des résultats plus intéressants. Cependant, les risques associés à l'utilisation unique de cette approche demeurent présents. Une stratégie combinant à la fois l'approche empirique et une ou plusieurs autres approches paraît donc plus prudente.

Lorsque l'on observe les résultats obtenus avec les grilles de correction générées à partir d'un groupe d'experts, on observe à la fois des convergences et des divergences. Sur le plan des similarités, on note que la corrélation entre les grilles déterminées par les experts financiers et les experts du secteur de l'énergie est très élevée ( $r= 0,826^{**}$ ), et ce, même si on note que 33% des réponses sont différentes entre les deux grilles. D'une part, les deux organisations œuvrent dans le secteur privé et ont des enjeux de service à la clientèle qui sont directement reliés à la rentabilité. On peut donc croire que les différences de réponses touchaient probablement des questions qui n'avaient pas un caractère fondamental commun au secteur privé. En d'autres mots, les items ayant été identifiées comme ayant les mêmes réponses (67% des questions) sont probablement celles qui jouent un rôle central dans la différenciation des employés performants et non-performants pour un poste de téléopérateur au sein du secteur privé.

Or, telle n'est pas la même situation pour la grille de correction générée à partir d'un groupe d'experts provenant du secteur public. En effet, on observe que les corrélations obtenues entre les résultats avec la grille d'experts du secteur public et celles du secteur privé étaient faibles ( $r=0,035$  avec les experts financiers et  $r=0,141^*$  avec les experts du domaine de l'énergie). Pourtant, dans les deux cas, les réponses ne variaient que de 36%. On peut donc croire que les différences observées entre les grilles de correction étaient probablement reliées à des aspects fondamentaux qui distinguent les téléopérateurs œuvrant dans le secteur public vs. privé.

Cette constatation peut être reliée au concept fondamental de la validité. En effet, Pettersen (2000) indique qu'un test n'est valide que dans la mesure où il est directement relié aux concepts et au

contenu de l'emploi pour lequel il est utilisé. Or, les environnements et les objectifs recherchés dans le secteur public et le secteur privé sont suffisamment différents pour justifier une grille de correction particulière à chacun de ces milieux. Si l'organisation financière qui a participé à la présente étude avait utilisé le TJS validé par les experts du secteur public, elle aurait pris des décisions de sélection qui ne correspondent pas aussi bien à leurs besoins. Ces résultats sont donc importants puisqu'ils mettent en lumière la nécessité d'utiliser un test qui à la fois répond aux caractéristiques de l'emploi (c'est-à-dire le contenu des questions), mais qui valorise également les comportements attendus spécifiquement dans le milieu dans lequel il est utilisé.

D'un point de vue plus général, ces résultats indiquent également qu'une des limites de l'utilisation d'experts de contenu touche au choix des SMEs qui participent à l'exercice d'élaboration de la grille de correction. Dans le cadre de notre étude, on observe par exemple que les experts de l'institution financière ont permis de générer une grille de correction qui est adéquate (la corrélation avec le rendement étant de  $r=0,32$ ) mais qui est légèrement inférieure à celle obtenue par les experts du secteur privé œuvrant dans le domaine de l'énergie ( $r=0,38$ ). Par contre, l'utilisation d'experts ayant une orientation s'apparentant au secteur public aurait produit une grille de correction dont l'efficacité aurait été moins grande ( $r=0,26$ ). Les organisations doivent donc s'assurer de choisir des experts qui représentent bien leur mode de pensée, de fonctionnement, et qui valorisent les comportements à l'origine d'un rendement supérieur.

Finalement, on observe que la grille de correction qui a été élaborée à partir d'une approche hybride est celle qui permet le mieux de tenir compte des différents aspects. Par exemple, on peut noter que la force de la relation entre l'approche théorique et l'approche basée sur les experts de contenu est la plus élevée ( $r=0,862^{**}$ ). Parallèlement, on observe que les corrélations entre la grille hybride et les grilles élaborées par les experts de contenu du secteur privé sont très élevées ( $r=0,830^{**}$  avec les experts financiers et  $r=0,932^{**}$  avec les experts de l'énergie). De plus, les différences de réponses avec les différentes grilles est relativement faible (35% et 15% respectivement), ce qui est normal puisque les décisions relatives aux bonnes réponses reposent sur des critères similaires (c'est-à-dire même type de milieu et d'objectifs recherchés ainsi qu'une information empirique additionnelle qui permet de calibrer les réponses en fonction du rendement des employés).

Les résultats démontrent également que 35% des bonnes réponses déterminées ont été modifiés par les experts de contenu suite à l'examen des informations fournies lors de la création de la grille hybride. C'est donc dire que, lorsqu'on leur soumet de l'information supplémentaire (tant rationnelle qu'empirique), un comité d'experts est généralement enclin à réviser certains de leurs jugements initiaux qui s'avéraient moins pertinents.

En somme, les résultats démontrent qu'il est essentiel que le choix des participants à une activité de validation soit effectué avec soin, les conséquences sur les qualités métrologiques de l'outil et sur les résultats des candidats pouvant être très importantes. De plus, même lorsqu'on adopte une méthode empirique souvent qualifiée de supérieure au jugement humain à cause de son caractère mécanique, on constate que cette approche produit des grilles de correction qui peuvent varier en moyenne de 52 %. Cette approche n'est donc pas à toute épreuve et la composition de l'échantillon retenu peut avoir un impact important sur la composition de la grille de correction finale.

Si on se penche maintenant sur les résultats associés à la distribution des résultats, les hypothèses se sont avérées partiellement confirmées. En fait, il est difficile de conclure à la supériorité de la méthode hybride, notamment en rapport à sa capacité de distribuer les résultats le plus normalement possible. En fait, bien que l'ensemble de grilles de correction ont produit des distributions normales, la méthode hybride n'a pas conduit à l'établissement d'une normalité supérieure aux autres grilles de correction. Il s'agit plutôt de l'approche empirique qui, à tout coup, a produit des distributions respectant le plus la courbe gaussienne. Comme il l'a été commenté précédemment, les résultats obtenus lors de la vérification de cette hypothèse ont conduit à la représentation fidèle de la puissance des relations statistiques. Cependant, comme l'objectif d'un processus de sélection de personnel est d'abord de prédire le rendement de ces individus une fois embauchés, le rejet de cette hypothèse ne milite pas nécessairement en faveur de l'adoption de l'approche empirique lors de l'élaboration d'un TJS.

Selon cette même perspective, concernant les résultats moyens obtenus aux différents exercices, on note que c'est l'approche hybride qui produit une moyenne correspondant aux attentes professionnelles (St-Sauveur, 2014). On note en effet qu'à l'exception de la grille hybride, toutes les approches produisent des moyennes inférieures à 60%.

Dans les faits, des moyennes variant entre 50% et 60% sont tout à fait adéquates sur le plan psychométrique dans la mesure où elles permettent aux candidats d'obtenir des résultats supérieurs ou inférieurs avec suffisamment de variabilité (vs. une moyenne qui serait de 85% ou de 15% par exemple). Cependant, une moyenne de 60% est habituellement perçue comme étant la norme, cette dernière étant supérieure au point milieu (50%). De plus, dans un contexte organisationnel, il peut être mal perçu qu'un groupe d'employés en poste obtienne une moyenne aussi basse que 50%. Un tel résultat est souvent interprété comme étant une démonstration que les employés en place font face à des problèmes de rendement. À l'inverse, une moyenne de 60% est souvent considérée comme étant adéquate.

Lorsque la moyenne obtenue à un instrument de sélection s'élève à 60%, cette situation peut parfois se révéler avantageuse aux yeux des professionnels RH responsables de l'embauche (St-Sauveur et al., 2014). En fait, non seulement ces derniers n'auront pas tendance à croire que l'instrument était trop facile pour les participants, mais ils pourraient préférer un instrument leur permettant d'établir une note de passage à 60%, c'est-à-dire autour de la moyenne. Comme le rappellent St-Sauveur et ses collègues (2014, p. 235), « *even if a mean near 60% is generally sought by practitioners, in a unionized context or where there is a low number of applicants compared with the number of positions available, and in the context where the cutoff point is set to the mean, for example, a higher mean could be sought* ».

Sur le plan de la dispersion des résultats, on note que cette dernière fournit aux professionnels RH une source d'information permettant soit de cibler une note de passage, soit de distinguer plus facilement les meilleurs candidats ayant participé au processus. En effet, comme le font remarquer St-Sauveur et ses collègues (2014, p. 235), « *when a SJT is used right before the final interview or when the candidate is very large and the selection ratio considered rather small, a larger standard deviation should be preferred in order to identify the very best candidates and to determine a cutoff point that few candidates approach with their results, in order to limit contestation* ».

Par contre, il est clair que la qualité d'une grille de correction élaborée par le biais des méthodes empiriques est largement compliquée par la difficulté à avoir accès à des échantillons suffisamment larges et représentatifs permettant de tirer des conclusions fermes sur la qualité de ces mesures. De cette manière, il est plausible de croire que la méthode hybride est une excellente façon d'utiliser la

force des statistiques pour disperser les résultats adéquatement tout en s'assurant de pouvoir ajuster le tir dans l'éventualité où la grille de correction mènerait à des choix de bonnes réponses irrationnelles ou très peu valides aux yeux des professionnels RH, des experts de contenu ou même des candidats recevant du feedback suite à leur participation au processus de sélection.

Or, les méthodes qui vont permettre d'obtenir une meilleure dispersion des résultats ne sont pas nécessairement celles qui permettent d'atteindre la meilleure estimation de la fidélité. Dans l'ensemble, on a pu constater que les résultats obtenus à la cohérence interne (technique d'estimation de l'homogénéité des items basée sur le patron de corrélations des items entre eux et le score total) ont été modérés (entre  $r=0,51$  et  $r=0,68$ ) pour chacune des méthodes d'élaboration des grilles de correction. On peut dès lors en déduire qu'aucun de ces TJS ne ciblait un seul et même construit homogène. Cette constatation est cohérente avec ce qu'on observe généralement pour ce type d'instrument d'évaluation en contexte de sélection.

Ceci étant dit, plus l'homogénéité est élevée, plus les items mesurent le « vrai score » de l'individu, diminuant ainsi l'erreur de mesure qui fait varier le score total d'une mesure à l'autre dans le temps. Nunnally (1978) ira même jusqu'à recommander un coefficient d'alpha de Cronbach ( $\alpha$ ) supérieur à 0,60 pour se prononcer sur la fiabilité d'une mesure. Or, comme le fait remarquer Bergman et ses collègues (2006, p. 231), « *broader STJs are likely to have a more scoring difficulties than narrower ones* ». De cette perspective, l'intercorrélacion entre chacune des questions d'un TJS rendrait beaucoup plus facile la détermination d'une grille de correction peu importe la méthode utilisée. Encore ici, la méthode empirique (grille 19) a permis d'obtenir la meilleure estimation de la fidélité en produisant un indice de cohérence interne supérieur à celle observée par le biais de toutes les autres méthodes d'élaboration des grilles de correction.

Dans l'optique de la deuxième série d'hypothèses, on constate d'abord que les résultats obtenus au TJS dans le cadre de cette recherche sont reliés positivement au rendement en emploi, et ce, peu importe la méthode d'élaboration de la grille de correction utilisée. Cependant, on observe que la méthode adoptée pour élaborer la grille de correction d'un TJS agit comme modérateur sur le plan de la validité critériée de ce test. En fait, comme Bergman et ses collègues (2006) l'avaient présenté avant cette recherche, les résultats soumis à la section précédente indiquent globalement que la validité prédictive d'un TJS peut varier grandement (passant de  $r=0,132^*$  à  $r=0,247^{**}$ ) en fonction de la stratégie

privilegiée lors de la conception de la grille de correction. Il va sans dire que cette variabilité dans la capacité d'un instrument à prédire efficacement le rendement futur en emploi peut avoir en retour des impacts non négligeables sur le pourcentage de bonnes décisions d'embauche et de rejet prises dans le cadre d'un processus réel de sélection de personnel. Ces résultats vont d'ailleurs dans le même sens que ceux obtenus dans le cadre d'autres études (Christian et al., 2010; Motowidlo et al., 1990; Stevens et Campion, 1999; Weekley et Jones, 1999).

Cependant, la présente recherche offre une explication partielle aux variations observées. D'abord, pour un même test et une même population de candidats, l'approche utilisant uniquement des experts de contenu produit des résultats variables qui ne sont pas nécessairement reliés à la connaissance que les experts ont de l'emploi spécifique pour le test est utilisé. Par exemple, si la grille élaborée par les experts de contenu du domaine financier est plus efficace ( $r=0,164^{**}$ ) que celle élaborée par les experts du secteur public ( $r= 0,132^*$ ), elle l'est cependant moins que celle qui a été déterminée par les experts d'un autre domaine d'activités (secteur de l'énergie) ( $r= 0,222^{**}$ ). Tel que mentionné précédemment, ces résultats peuvent s'expliquer en partie par le fait que les deux derniers groupes d'experts œuvrent dans le secteur privé. On pourrait donc croire que, dans le cas des téléopérateurs, les comportements qui permettent de prédire le rendement futur en emploi sont moins étroitement reliés aux tâches spécifiques effectuées dans une organisation, mais plutôt à des facteurs sous-jacents généraux, tels que la qualité du service à la clientèle ou la capacité de vendre un produit qui déterminerait la qualité du travail effectué. Ce qui compterait donc ce seraient avant tout les compétences nécessaires à l'emploi et non les aspects spécifiques particuliers à une organisation. Cette hypothèse est plausible et irait dans le sens des études portant sur l'importance des compétences génériques en emploi (Durivage, 2004; Spencer et Spencer, 1993) Dans cette perspective, le fait d'œuvrer dans le domaine des finances ou de l'énergie aurait peu d'importance dans la mesure où les compétences nécessaires seraient similaires (vs. le domaine public où les aspects d'aide au citoyen prennent une place beaucoup plus importante que les dimensions d'affaires propre au secteur privé).

Par ailleurs, ces résultats sont intéressants et portent à l'attention le fait que l'expertise des ressources assignées à un comité de travail mis sur pied lors de la validation d'un test ne doit pas être prise à la légère. Même s'il est impossible de conclure qu'il s'agit du niveau d'expertise des experts de contenu face au domaine professionnel à évaluer qui a fortement amélioré l'effort de prédiction de l'instrument utilisé dans le cadre de ce processus de sélection, il est nécessaire de réitérer ces

considérations opérationnelles qui peuvent avoir un impact important sur l'efficacité d'un outil d'évaluation validé par des experts de contenu étant moins au fait des compétences requises pour un emploi. Tel que mentionné précédemment, comme il est souvent difficile et coûteux pour une organisation de libérer des experts de contenu pour siéger sur des comités de travail responsables d'examiner chaque situation et énoncés de réponse d'un TJS, les résultats ne militent pas nécessairement en faveur de l'utilisation unique de cette approche.

Par ailleurs, il semble qu'il n'ait pas été beaucoup plus avantageux de procéder avec le recours à la théorie ( $r= 0,162^{**}$ ) pour valider la grille de correction du TJS. En fait, lorsque l'on se concentre davantage sur des sources rationnelles, comme l'approche théorique ou le recours aux experts, on procède généralement de manière intuitive, déductive et la compréhension du domaine à évaluer prend beaucoup d'importance dans l'identification des bonnes réponses à l'exercice. Or, cela peut s'avérer très difficile dans le cas d'un domaine aussi général que le jugement situationnel en emploi et le service à la clientèle. En effet, bien que les TJS semblent être plus difficiles à déjouer (« *less fakeable* ») que les mesures de personnalité, il n'en demeure pas moins que procéder à l'élaboration d'une grille de correction en se basant exclusivement sur la théorie pourrait mener à des questions plus facilement détectables de la part des candidats. Encore ici, les résultats obtenus dans le cadre de cette étude ne militent pas en faveur d'une telle approche unique.

D'un autre regard, certaines études antérieures ont établi que la manipulation empirique était optimale pour prédire un critère spécifique par rapport à l'approche rationnelle. Dans le cadre de cette recherche, lorsque soumise à une contre-validation exhaustive, il est vrai que l'approche empirique a produit un coefficient plus élevé comparé avec d'autres méthodes ( $r=0,278^{**}$  pour la grille 28). Ce résultat va d'ailleurs dans le même sens que l'étude de Weekly et Jones (1997) qui a démontré que l'approche empirique corrélait plus fortement avec le rendement en emploi que la méthode basée sur les experts de contenu ( $r= 0,35$  vs.  $r= 0,14$ ). Cependant, plusieurs critiques importantes peuvent être adressées à l'approche empirique et doivent être prises en considération lors de l'élaboration d'un TJS visant la sélection du personnel. D'une part, lorsque l'on s'est attardé à l'analyse des différences de moyennes produites par les grilles, on constate que l'approche des experts de contenu du secteur de l'énergie ainsi que la méthode hybride a produit des différences de moyennes supérieures au mode empirique (lorsque comparé avec la grille 19). D'autre part, les méthodes empiriques présentent des défis par rapport à leur dépendance envers le groupe de personnes à partir desquelles les analyses sont

effectuées, la présence ou l'absence d'un biais d'échantillonnage pouvant jouer un rôle important dans la qualité des résultats obtenus. À cet effet, la méthodologie adoptée dans le cadre de cette thèse a notamment fait ressortir qu'il est nécessaire de se questionner sur la stabilité de cette approche ainsi que sur son caractère généralisable compte tenu de son fort pourcentage de variabilité. Qui est plus, il est important de noter que l'approche empirique requiert d'abord de recueillir une grande série de données ce qui n'est pas toujours aisé et facile à opérationnaliser dans les activités RH courantes d'une organisation.

Dans notre effort de prédiction du rendement en emploi, la combinaison des différentes approches (la méthode hybride) s'est avérée positive. En effet, malgré que certains auteurs ont démontré précédemment qu'il n'y avait pas de réelle valeur ajoutée à combiner l'approche empirique à une méthode rationnelle, la méthodologie adoptée dans le cadre de cette étude par le biais d'une approche hybride d'élaboration de grilles de correction a produit des résultats différents et légèrement supérieurs. En effet, en consultant les experts de contenu du secteur financier une seule fois ou en s'appuyant exclusivement sur l'approche théorique, la prédiction du rendement en emploi pour les téléopérateurs de cette organisation aurait été de l'ordre de  $r=0,31$  et  $r=0,32$  (lorsque les coefficients sont corrigés). Cependant, consulter un groupe d'experts du milieu une deuxième fois, en leur fournissant de l'information concernant les meilleures pratiques et en leur exposant les réponses choisies par les individus de leur organisation démontrant un rendement supérieur en emploi a permis d'améliorer cet effort de prédiction à  $r=0,42$  (lorsque corrigé), ce qui représente une amélioration substantielle. Force est de constater, à la lecture de ces résultats, qu'il existe une réelle valeur ajoutée à procéder par le biais d'une approche hybride lors de l'élaboration d'un nouveau TJS. Ces résultats sont d'ailleurs cohérents avec les conclusions tirées par Bergman et ses collègues (2006) qui ont également trouvé des coefficients de validité plus élevés par le biais d'une méthode hybride, c'est-à-dire en identifiant la bonne réponse de manière rationnelle et empirique.

À cet effet, malgré qu'il n'existe aucun consensus sur l'interprétation de la force de la relation lorsque représenté par un coefficient de validité critériée en sélection de personnel, on s'entend généralement pour dire qu'un coefficient de corrélation corrigé supérieur à  $r=0,35$  est considéré comme élevé (Bourhis, 2007). De ce fait, en plus de miser sur l'avantage des approches empiriques qui contribuent généralement à identifier les candidats démontrant le meilleur potentiel, solliciter des experts de contenu une deuxième fois pour la détermination d'une grille de correction finale permet



également de s'assurer que le TJS permettra de cibler des comportements qui sont liés aux objectifs stratégiques de l'organisation, tout en s'assurant qu'ils se remettront en question concernant leur appréciation initiale de certains comportements.

Néanmoins, il demeure nécessaire de poursuivre les recherches afin de s'assurer de la supériorité de l'approche hybride dans la mesure où l'application de cette méthode demeure dépendante des circonstances dans lesquelles elle est appliquée. En effet, les options de réponse pour un TJS varient le long d'un continuum allant de la meilleure réponse à la pire réponse. Or, l'identification de la bonne réponse est d'autant plus difficile à déterminer qu'elle variera sans doute selon le type de poste et le type d'organisation en jeu. Incidemment, il appert clairement que la grande variabilité des coefficients de validité prédictive indique l'importance de choisir la bonne méthode afin de maximiser l'utilité d'un TJS dans un processus de sélection de personnel. Dans le même sens, alors que l'on en conclut que le contenu d'un TJS n'est pas valide, il s'agit peut-être seulement de sa grille de correction qui n'a pas été élaborée convenablement ou selon les bonnes conditions (Bergman et al., 2006). Quand on pense aux possibilités de contestations légales, il s'avère primordial que les organisations soient en mesure non seulement de déterminer que les bonnes réponses identifiées dans un TJS sont reliées au contenu de cet examen utilisé dans le cadre d'un processus de sélection, mais également quelle stratégie de validation de la grille de correction ils ont utilisée et pour quelles raisons.

## 7. FORCES ET LIMITES DE LA RECHERCHE

---

D'abord, en dressant l'inventaire des différentes méthodes de correction pouvant guider l'élaboration d'un TJS en contexte organisationnel, cette recherche offre un repère intéressant pour quiconque vise à en évaluer l'efficacité relative en gestion des ressources humaines. Ensuite, les données de cette recherche ont été recueillies auprès d'un échantillon de candidats québécois avec un échantillon important compte tenu des contraintes habituellement présentes dans le milieu organisationnel (N=312). Le nombre de sujets impliqués a d'ailleurs permis de comparer, avec un bon degré de confiance, les principales conclusions de cette recherche avec les études qui ont été réalisées ailleurs sur l'élaboration des grilles de correction des TJS, notamment aux États-Unis et en Europe. Par ailleurs, les données ont été recueillies par le biais d'individus occupant réellement un poste de téléopérateur et évoluant spécifiquement dans le secteur d'activité des centres d'appels pour lequel le

test a été développé. Cette caractéristique de l'échantillon ajoute à la validité des conclusions tirées, ce qui aurait été plus difficile si on avait utilisé des sujets qui se seraient portés volontaires et qui provenaient de la population générale. Or, l'utilisation de ce dernier type de sujets représente le schème expérimental qui est le plus souvent retenu par la communauté scientifique s'intéressant aux TJS et aux caractéristiques entourant leur validité (McDaniel et al., 2007). Finalement, le fait d'avoir pu comparer les résultats provenant de trois comités d'experts de contenu provenant d'organisations différentes représente une valeur ajoutée non négligeable. Ces données provenant d'autres organisations permettent d'ailleurs de poursuivre la recherche menant à une meilleure compréhension des différentes méthodes permettant de corriger les TJS.

Cependant, comme dans toute recherche, celle-ci comprend un certain nombre de limites. Premièrement, cet échantillon était composé d'individus déjà en poste dans un centre d'appels. On peut donc se questionner sur l'influence d'un schème concomitant sur le plan de la validité prédictive d'un instrument de mesure en comparaison à l'utilisation d'un bassin de candidatures participant à un processus réel de sélection. En fait, bien que le schème concomitant offre l'avantage de faire connaître assez rapidement les résultats concernant la validité de l'instrument étudié, il faut également demeurer prudent étant donné que l'utilisation de cet échantillon peut également présenter certains biais en sélection de personnel (Pettersen, 2000). En effet, les employés déjà en place représentent une population distincte de celles des candidats qui aspirent à occuper un poste en particulier et on assiste généralement à une restriction de la variance qui ne serait pas présente en utilisant l'ensemble d'un bassin de candidatures potentielles et liée au fait que les répondants sont différents en termes d'expérience, d'ancienneté et de motivation. En effet, les personnes déjà en poste peuvent avoir tendance à compléter un test avec un niveau de motivation moins élevé qu'une personne ayant posé sa candidature pour obtenir un poste en particulier. Dans le même sens, ce type d'échantillon entraîne une réduction de la variance sur le plan de la mesure du critère (le rendement en emploi) étant donné que certains employés pourraient avoir quitté leur emploi à cause d'une mauvaise performance ou selon le fait que des mesures de sélection ont été utilisées initialement au moment de l'embauche. Or, les conclusions sur l'influence de la configuration du schème expérimental sur la validité prédictive des TJS utilisés en sélection de personnel demeurent mitigées. À cet effet, Weekley et ses collègues (2004) ont utilisé deux types d'échantillon pour comparer les réponses des titulaires d'emploi avec celles de candidats à un processus de sélection par le biais d'un TJS orienté sur l'évaluation des relations interpersonnelles et du service à la clientèle. Sur le plan de la validité prédictive, les résultats ont été

relativement similaires entre les deux types de population ( $r=0,16^*$  pour les candidats et  $r=0,12^*$  pour le personnel en place) (Weekley, Ployhart, et al., 2004).

Dans une autre perspective, comme la recherche a été réalisée par le biais d'un seul échantillon, il est légitime de se questionner sur le caractère généralisable de cet échantillon. À quel point, la détermination des bonnes réponses dans les grilles de correction empiriques n'est pas seulement le fruit d'une influence culturelle des participants qui ont complété le test affectant en retour la validité prédictive de cette méthode. Dans le même sens, le fait que les grilles de correction vont varier en fonction de la provenance des experts de contenu en produisant des coefficients de validité critériée différents, il serait souhaitable de continuer l'expérimentation en organisant différents ateliers de travail en présentant des données empiriques et les notions théoriques du domaine afin de solliciter une validation de la grille de correction en mettant en œuvre à nouveau l'approche multiple.

Par exemple, comme Weekly et Jones (1997) l'ont fait précédemment, solliciter des clients plutôt que des supérieurs immédiats à titre d'experts de contenu au début ou à la fin de la détermination des bonnes réponses pourrait permettre de contre-valider les résultats présentés précédemment. En effet, comme le service à la clientèle sert souvent de « pierre angulaire » dans le monde des affaires d'aujourd'hui, s'assurer que l'opinion des clients soit prise en compte dans les processus visant à sélectionner les futurs téléopérateurs pourrait représenter une approche qui contribue également à l'atteinte des objectifs stratégiques de l'organisation. Enfin, il serait intéressant d'isoler l'impact de la nature des activités du centre d'appels (par exemple faire des appels vs. recevoir des appels), de vérifier l'impact de la nature des opérations de cette unité d'affaires (type de clients, durée des appels, etc.) ainsi que de comparer les différentes méthodes d'évaluation des opérations et des employés (qualité vs. quantité) sur la capacité des différentes méthodes d'élaboration des grilles de correction du TJS à prédire le rendement futur en emploi.

Par ailleurs, le TJS utilisé dans le cadre de cette étude a été construit selon la méthode du choix forcé où pour chaque item, présentant deux énoncés de réponses différents, existait une seule bonne réponse. Dans cette perspective, la méthode d'attribution du pointage était de 1 point ou 0 point selon que la personne avait déterminé la bonne réponse ou non. Or, il est légitime de se questionner si le fait d'avoir utilisé un TJS avec plusieurs énoncés de réponses qui aurait permis d'y attribuer différentes méthodes de pointage (« *best-and-worst-answer method* », « *rank-ordering of response options*

*method* » or « *rank-ordering with negative correction method*») aurait permis d'avoir une vision plus complète et plus large de la capacité d'une personne à résoudre des problèmes en situation de travail et plus représentative de sa performance globale au travail en pénalisant ses erreurs de jugement (St-Sauveur et al., 2014).

Certaines limites peuvent également être soulevées par rapport à la méthode adoptée lors de l'élaboration des grilles de correction par le biais d'une approche empirique. Dans le cadre de cette étude, une technique mettant en lien l'utilisation d'un seul critère (c'est-à-dire le rendement en emploi) et l'utilisation d'une seule manipulation statistique (méthode corrélacionnelle) a été adoptée. Incidemment, est-ce que le fait de retenir un critère combinant différentes informations (par exemple le nombre d'années d'expérience comme téléopérateurs, le pourcentage d'atteinte d'objectifs quantitatifs, etc.) et de réaliser un autre type d'analyses statistiques (relations entre les individus, relations entre les variables, etc.) aurait conduit à des résultats supérieurs pour la capacité de la méthode empirique à prédire le rendement futur en emploi ? Comme on s'entend généralement pour dire que l'utilisation de la méthode empirique est traditionnellement associée à l'attribution d'une pondération relative par item (particulièrement dans la littérature portant sur les données biographiques), il serait intéressant de constater si cette approche aurait conduit à des résultats différents dans le cadre de cette étude (M. D. Mumford et Owens, 1987).

Notons également qu'il serait nécessaire d'examiner de façon plus approfondie les conditions dans lesquelles la méthode hybride est utilisée. Dans le cas de notre étude, les experts de contenu ont pris leur décision en utilisant les données empiriques provenant du groupe de personnes ayant permis un niveau de prédiction modéré (grille 19 avec une corrélation de 0,154). Les conclusions auraient-elles été les mêmes, si les experts avaient eu à prendre leurs décisions à l'aide de la grille la moins performante (grille 5 avec une corrélation de 0,022) ou la plus performante (grille 28 avec une corrélation de 0,278) du point de vue prédictif ? Dans le cas de la grille la moins performante, aurait-on assisté à une baisse du niveau de prédictivité par rapport à celle obtenue par le biais de la grille initiale (experts de contenu seulement) ou les experts auraient-ils décidé de ne pas tenir compte d'une information qu'ils auraient jugée moins adéquate ? À l'inverse, l'utilisation d'une grille plus performante aurait-elle ajouté à la capacité de prédire le rendement ou existe-t-il un plafond où les experts maintiennent leurs décisions, peu importe qu'ils soient ou non confrontés à des réponses qui sont empiriquement plus prédictives ?

Finalement, lorsque l'on s'intéresse à la validité critériée d'un instrument d'évaluation utilisé en sélection de personnel, le rendement en emploi est le critère le plus communément utilisé. Or, la mesure du rendement comporte plusieurs défis. En fait, nombreuses sont les erreurs de jugement qui peuvent être attribuables à l'évaluation des supérieurs immédiats. Malgré que cela devrait être compensé en partie par la mesure de correction statistique concernant l'atténuation de la fidélité du critère, il aurait été intéressant d'avoir accès à un certain nombre d'autres variables qui auraient pu soutenir l'analyse de la validité prédictive de chacune des méthodes mises en œuvre dans le cadre de cette étude. À titre indicatif, des critères quantitatifs comme des indices de production (par exemple, nombre d'appels complétés), des indices de qualités (par exemple, nombre de plaintes) ou des informations compilées dans le dossier des employés ayant participé à la recherche (par exemple, assiduité au travail ou avancement salarial) auraient pu servir à rendre compte du caractère multidimensionnel du rendement global des employés tout en limitant les sources de biais traditionnellement reliés à l'évaluation du rendement, lorsque réalisé par les supérieurs immédiats.

## 8. PISTES DE RECHERCHES

---

Cette recherche offre une perspective nouvelle sur la question de l'élaboration des grilles de correction des TJS. En effet, par la mise en œuvre d'une expérimentation visant à tester une méthode originale combinant différentes approches pour leur correction, la méthode hybride présentée précédemment semble une voie prometteuse pour la prédiction future du rendement en emploi. Or, le sujet demeure une piste de recherche intéressante et certaines des questions discutées ci-après devront être explorées par le biais d'études complémentaires.

D'une part, le critère de rendement global qui a été recueilli par le biais de cette étude et utilisé, autant pour la construction des grilles de correction empiriques que pour l'établissement de la validité prédictive, pourrait être bonifié afin de contre valider les résultats présentés plus tôt. En effet, dans le cadre d'une recherche subséquente, il serait intéressant de recueillir des données provenant d'une évaluation de la part du supérieur immédiat portant plutôt sur chacune des dimensions génériques du rendement pour les sujets en question (voir le modèle de Campbell et ses collègues (1990) à la page 62) afin de vérifier l'efficacité relative de chacune des grilles de correction élaborées précédemment. Par le

biais d'un critère de rendement davantage multidimensionnel, il serait intéressant d'établir une distinction entre le rendement axé sur les tâches et le rendement contextuel afin d'en vérifier l'impact sur la prédiction du rendement futur en emploi par le biais d'un TJS utilisé pour la sélection du personnel dans un centre d'appels.

Dans le même sens, le critère de rendement utilisé ici contenait seulement une appréciation subjective reflétant la perception globale de chaque supérieur immédiat impliqué dans la démarche. Or, il pourrait être souhaitable de combiner à cette évaluation du rendement d'autres critères de rendement de nature quantitative (par exemple, nombre de plaintes reçues, nombre d'appels traités par quart de travail, pourcentage d'atteinte des objectifs de ventes, etc.) afin de vérifier si les résultats observés précédemment seraient également en cohérence avec cette autre forme de critère.

Finalement, d'autres variables pourraient servir de critère de référence autant pour l'établissement des grilles de correction empiriques que pour valider le pouvoir de prédiction des différentes méthodes analysées, telles que le nombre d'années dans un poste (ancienneté et/ou expérience), une cote de satisfaction à l'égard de l'employé de la part de la clientèle, etc. Par ailleurs, ces informations supplémentaires, combinées à d'autres types de variables, comme le sexe, l'origine ethnique, la provenance du candidat (unité d'affaires, région géographique, etc.) etc., pourraient également servir à analyser des modérations éventuelles dans la relation de prédiction entre l'instrument et le rendement en emploi. Par exemple, il serait intéressant de vérifier si le fait de provenir d'une petite unité d'affaires agit comme modérateur dans la relation entre les deux variables à l'étude. Dans le même sens, est-ce que l'effort de prédiction serait plus élevé pour les candidats possédant le plus grand nombre d'années en poste ? En bref, plusieurs questions intéressantes pourraient être envisagées par l'introduction de ces variables modératrices lors de la poursuite des travaux sur le sujet.

D'autre part, la construction de la grille de correction préliminaire par le biais d'une approche empirique pourrait faire appel à différentes techniques outre que celle explorée précédemment de façon à en vérifier l'impact en retour sur la détermination finale des bonnes réponses par les experts (méthode hybride). Par exemple, tel que proposé par Lievens (2000), il pourrait être intéressant d'utiliser « l'analyse factorielle des correspondances » afin d'établir, entre autres, une pondération relative des items du TJS. En effet, comme le mentionne Lievens (2000), une des difficultés inhérentes à la correction des tests de jugement situationnel est que, dans l'ensemble, les items se voient

habituellement attribuer la même pondération. Or, bien que cela ne représente pas la manière optimale d'en arriver à un score total, il devient pratiquement impossible pour des experts de contenu d'en arriver à définir la pondération optimale de chaque item d'un test ainsi que la valeur relative des situations présentées à travers le test en entier (Lievens, 2000; Ostroff, 1991). De cette manière, une des voies intéressantes pour la recherche future sur les tests de jugement situationnel serait de vérifier l'efficacité relative de cette méthode d'analyse multi variée à partir de laquelle « *the item options are scored and weighted taking account of their relationships with all the other item options (including the ones from other items)* » (Lievens, 2000, p. 120). En établissant la valeur relative des situations présentées dans un TJS et ainsi que des énoncés de réponses, les items sont alors retenus pour leur pouvoir discriminant. Bien que plus difficile à opérationnaliser en sélection de personnel, il s'agit d'une manière intéressante de déterminer a posteriori si cette approche d'identification des bonnes réponses d'un TJS conduit à des résultats différents, soit en surpassant la méthode hybride lorsqu'analysée de manière indépendante, soit en lui permettant d'atteindre un niveau de validité prédictive encore plus intéressante que celle avancée précédemment.

Une autre piste de recherche consisterait à évaluer l'impact de la qualité prédictive des grilles empiriques sur les décisions finales prises par un groupe d'experts lors de l'élaboration d'une grille hybride. Il serait par exemple possible d'exposer différents groupes d'experts à des grilles empiriques plus ou moins efficaces et d'examiner l'impact au niveau du caractère prédictif final. Compte tenu des enjeux associés à la sélection du personnel, une telle étude pourrait prendre place dans un milieu expérimental (ex : une université où les étudiants jugeraient de la pertinence des réponses à un TJS évaluant l'efficacité de différents modes d'enseignement).

Par ailleurs, au-delà de l'approche adoptée pour identifier les bonnes réponses, plusieurs travaux restent à faire sur l'attribution des pointages et les différents formats de réponses présentés à l'intérieur des TJS. Dans le cadre de la présente recherche, chaque situation était associée à une seule bonne réponse pour laquelle le candidat accumulait un point ou aucun selon qu'il avait identifié la bonne réponse ou non. Or, il pourrait être intéressant de reproduire l'application des différentes grilles de correction en y associant d'autres formes d'attribution des pointages. Par exemple, par le biais d'un instrument comprenant une série d'énoncés de réponses, une étude complémentaire pourrait s'inspirer de la méthode originale proposée par St-Sauveur et ses collègues (2014) qui attribuait des points au candidat selon la mise en rang des choix de réponses effectuée, en plus de lui en soustraire des points

lorsque l'ordre des réponses soumis était totalement inversé où lorsque la meilleure réponse était identifiée comme la pire et vice-versa. Chaque question devient alors associée à un pointage variant entre quatre (+4) points et moins quatre (-4) points. Cette perspective pourrait être intéressante afin d'en vérifier l'impact sur l'efficacité relative des différentes méthodes présentées précédemment.

Finalement, il va sans dire que cette expérimentation a laissé plusieurs autres variables non mesurées. En effet, cette étude s'est déroulée dans un secteur d'activité bien précis, les centres d'appels, par le biais de données provenant exclusivement d'une institution financière pour un niveau de postes bien particulier. Incidemment, il pourrait être intéressant de recueillir des données de rendement des personnes en poste, notamment avec l'organisation du secteur de l'énergie et celle du secteur public ayant contribué parallèlement à cette recherche. La continuation de ces travaux sur l'élaboration des grilles de correction des TJS permettrait de répondre à d'autres questions en y impliquant le type d'environnement et de culture organisationnelle selon les centres d'appels étudiés. Ces données pourraient également permettre de vérifier si on y observe des différences culturelles lors de la création de grilles de correction par des experts de contenu selon leur pays d'origine (France/Québec). En bref, afin de s'assurer du caractère généralisable de cette étude, ce schème expérimental devrait être reproduit avec d'autres types postes (différents secteurs d'activités) avec un type d'échantillon différent (différents niveaux de postes).



## CONCLUSION

---

Dans le secteur de la gestion des ressources humaines, il faut d'abord sélectionner les personnes les plus qualifiées pour occuper un emploi, donc les plus susceptibles de réussir et d'optimiser le rendement général d'une organisation. Dans ce contexte, une des contributions les plus importantes de la psychologie industrielle et organisationnelle au domaine du travail concerne sans doute l'utilisation d'instruments d'évaluation visant à permettre l'identification des individus démontrant le meilleur potentiel à réussir dans un poste en particulier. À cet égard, les tests d'aptitudes professionnelles, notamment les tests de jugement situationnel (TJS), ont fait l'objet de nombreuses recherches scientifiques et sont de plus en plus utilisés dans les opérations d'embauches quotidiennes d'une multitude d'organisations publiques et privées pour leur capacité à fournir une mesure des comportements qui seront adoptés par un candidat dans le cadre de ses fonctions ou de sa connaissance de la meilleure façon de faire face aux situations rencontrées. À l'évidence, pour combler les postes vacants à l'intérieur d'un centre de contact client ou centre d'appels, une unité d'affaires de ce type sera d'abord intéressée, par exemple, à cerner à l'avance si les personnes ayant posé leur candidature sont susceptibles d'atteindre leur objectif de ventes en connaissant les principales techniques de négociations ou d'adopter, au bon moment, des comportements qui démontrent une approche respectueuse axée sur la clientèle. C'est le cas de l'institution financière ayant contribué à cette recherche.

Pour ce faire, les TJS existent depuis plusieurs décennies, mais ils ont récemment gagné en popularité notamment due au fait qu'ils sont relativement peu coûteux et faciles à administrer, qu'ils présentent de faibles différences de moyennes entre les personnes provenant de groupes protégés (genre et origine ethnique) et qu'ils génèrent des réactions positives de la part des candidats (Gessner et Klimoski, 2006; Weekley et Ployhart, 2006). Or, leur qualité n'est pas équivalente en soi et on constate que la validité des TJS varie considérablement d'une étude à l'autre. À cet effet, les écrits scientifiques vont démontrer que plusieurs des décisions qui sont prises lors de leur développement, à l'étape de la conception, de la validation ou de leur administration préliminaire, vont en retour affecter la validité et l'utilité de ces tests. Le protocole de recherche mis en œuvre dans le cadre de cette thèse s'inscrivait donc dans ce questionnement global portant sur les meilleures approches permettant de maximiser l'utilité de ces instruments lors d'un processus de sélection en plus d'en assurer une meilleure validité prédictive. Pour ce faire, il s'agit des différentes formes d'identification des bonnes réponses qui ont été

étudiées dans le cadre de cette recherche, c'est-à-dire, plus spécifiquement, qu'elle est l'efficacité relative des différentes méthodes d'élaboration d'une grille de correction d'un TJS ?

À partir d'un échantillon de 312 téléopérateurs en situation d'emploi dans le domaine bancaire québécois qui ont complété un test de jugement situationnel, l'application de différentes grilles de correction en lien avec une mesure de rendement global de chacun de ces individus a permis de répondre à ce questionnement principal. Dans un premier temps, sur le plan de la distribution des résultats, la méthode empirique a démontré une meilleure distribution des scores des participants au processus. Par ailleurs, la méthode empirique et la méthode hybride ont démontré plus de variance que la méthode théorique ainsi que celle basée sur le jugement des experts. En deuxième lieu, on note que l'approche permettant d'obtenir la meilleure validité prédictive est celle qui permet aux experts de prendre leurs décisions en fonction d'un maximum d'informations (approche hybride). Malgré les limites de la recherche indiquées précédemment, c'est donc dire que les résultats de la présente étude militent en faveur de l'utilisation d'une telle approche lors de l'élaboration de tests de jugement situationnel lors de processus d'évaluation du potentiel.

## RÉFÉRENCES BIBLIOGRAPHIQUES

---

- Allworth, E. et Hesketh, B. (1999). Construct-oriented biodata: Capturing change-related and contextually relevant future performance. *International Journal of Selection and Assessment*, 7 (2), 97-111.
- Arrègle, J.-L. (2006). Analyse « Resource Based » et identification des actifs stratégiques. *Revue Française de Gestion* (160), 241-259.
- Arthur Jr, W. et Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, 93, 435-442.
- Arvey, R. D. et Faley, R. H. (1992). *Fairness in selecting employees* (2 ed.). Don Mills, Ont.: Addison-Wesley Publishing Company.
- Austin, J. T. et Villanova, P. (1992). The criterion problem: 1917-1992. *Journal of Applied Psychology*, 77 (6), 836-874.
- Banki, S. et Latham, G. P. (2010). The Criterion-Related Validities and Perceived Fairness of the Situational Interview and the Situational Judgment Test in an Iranian Organisation. *Applied Psychology: An International Review*, 59 (1), 124-142.
- Barrick, M. R. et Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, 81 (3), 261-272.
- Bartram, D. (2004). Assessment in Organisations. *Applied Psychology: An International Review*, 53 (2), 237-259.
- Bauer, T. N., Maertz Jr, C. P., Dolen, M. R. et Campion, M. A. (1998). Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. *Journal of Applied Psychology*, 83 (6), 892-903.
- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P. et Campion, M. A. (2001). Applicant Reactions to Selection: Development of the Selection Procedural Justice Scale (SPJS). *Personnel Psychology*, 54 (2), 387-419.
- Bechet, T. P. (2008). *Strategic staffing a comprehensive system for effective workforce planning, second edition*. New York: American Management Association.

- Becker, B. E. et Huselid, M. A. (2006). Strategic Human Resources Management: Where Do We Go From Here. *Journal of Management*, 32 (6), 898-925.
- Béraud, M., Colin, T. et Grasser, B. (2008). La qualification dans les centres d'appels : affaiblissement ou recomposition ? *Travail et emploi*, 114 (2), 45-57.
- Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B. et Juraska, S. E. (2006). Scoring Situational Judgment Tests: Once You Get the Data, Your Troubles Begin. *International Journal of Selection & Assessment*, 14 (3), 223-235.
- Binning, J. F. et Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, 74 (3), 478-494.
- Bledow, R. et Frese, M. (2009). A Situational Judgment Test of Personal Initiative and its Relationship to Performance. *Personnel Psychology*, 62 (2), 229-258.
- Bobko, P. (2001). *Correlation and Regression: Applications for Industrial Organizational Psychology and Management*. California: Sages Publications.
- Boivin, J. (1987). Les relations industrielles: une pratique et une discipline. *Relations Industrielles*, 42 (1), 179-196.
- Borman, W. C. (1991). Job behavior, performance, and effectiveness. Dans M. D. Dunnette & L. M. Hough (Éds.), *Handbook of industrial and organizational psychology* (2 éd., Vol. 2, pp. 271-326). Palo Alto, CA: Consulting Psychologists Press.
- Borman, W. C. et Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. Dans N. S. W. C. Borman (Éd.), *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.
- Boudrias, J.-S. et Morin, D. (2011). Démystifier les inventaires de personnalité. *Gestion*, 36 (3), 61-73.
- Bourhis, A. (2007). *Recrutement et sélection du personnel*. Montréal: Gaétan Morin Éditeur.
- Breaugh, J. A. (1981). Relationships between Recruiting Sources and Employee Performance, Absenteeism, and Work Attitudes. *The Academy of Management Journal*, 24 (1), 142-147.

- Burgers, A., Ruyter, K. d., Keen, C. et Streukens, S. (2000). Customer expectation dimensions of voice-to-voice service encounters: a scale-development study. *International Journal of Service Industry Management*, 11 (2), 142-161.
- Burke, R. J. et Ng, E. (2006). The changing nature of work and organizations: Implications for human resource management. *Human Resource Management Review*, 16 (2), 86-94.
- Buscatto, M. (2002). Les centres d'appels, usines modernes ? Les rationalisations paradoxales de la relation téléphonique. *Sociologie du Travail*, 44 (1), 99-117.
- Campbell, J. P., Dunnette, M. D. et Hough, L. M. (1990). Modeling the performance prediction problem in industrial and organizational psychology. Dans *Handbook of industrial and organizational psychology, Vol. 1 (2nd ed.)*. (pp. 687-732). Palo Alto, CA, US: Consulting Psychologists Press.
- Campbell, J. P., Gasser, M. B. et Oswald, F. L. (1996). The substantive nature of job performance. Dans K. R. Murphy (Éd.), *Individual differences and behaviour in organizations* (pp. 258-299). San Francisco: Jossey-Bass.
- Campbell, J. P., McCloy, R., Oppler, S. H. et Sagger, C. E. (1993). A theory of performance. Dans N. Schmitt, Borman & Associates (Éd.), *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass.
- Campbell, J. P., McHenry, J. J. et Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology*, 43 (2), 313-333.
- Cappelli, P. (2003). Will There Really Be a Labor Shortage? *Organizational Dynamics*, 32 (3), 221-233.
- Carricano, M., Poujol, F. et Bertrandias, L. (2010). *Analyse de données avec SPSS* (2e ed.). Paris: Pearson Education.
- Catano, V. M. (2010). *Recruitment and selection in Canada* (4 ed.). Toronto: Nelson Education
- Catano, V. M., Brochu, A. et Lamerson, C. D. (2012). Assessing the reliability of situational judgment tests used in high-stakes situations. *International Journal of Selection and Assessment*, 20 (3), 333-346.
- Chan, D. (2006). Interactive Effects of Situational Judgment Effectiveness and Proactive Personality on Work Perceptions and Work Outcomes. *Journal of Applied Psychology*, 91 (2), 475-481.

- Chan, D. et Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82 (1), 143-159.
- Chan, D. et Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15 (3), 233-254.
- Chapman, D. S. et Webster, J. (2003). The Use of Technologies in the Recruiting, Screening, and Selection Processes for Job Candidates. *International Journal of Selection and Assessment*, 11 (2-3), 113-120.
- Chênevert, D., Dubé, M. et Déom, E. (2008). Les nouvelles formes d'organisation du travail. *Relations Industrielles*, 63 (1), 134-159.
- Christian, M. S., Edwards, B. D. et Bradley, J. C. (2010). Situational Judgment Tests : Constructs Assessed and a Meta-Analysis of their Criterion-Related Validities. *Personnel Psychology*, 63 (1), 83-117.
- Cihuelo, J. (2010). Le cas d'un modèle hybride de centre d'appels. *Réseaux*, 6 (164), 161-196.
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N. et Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86 (3), 410-417.
- Cohen, J. (1983). The cost of dichomotization. *Applied Psychological Measurement*, 7, 247-253.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112 (1), 155-159.
- Conger, J. A. et Riggio, R. E. (2007). *The practice of leadership developing the next generation of leaders*. San Francisco, Calif.: Jossey-Bass.
- Cook, M. (2009). *Personnel selection : adding value through people*. Hoboken, N.J.: J. Wiley & Sons.
- Cronbach, L. J. et Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52 (4), 281.
- Crook, T. R., Todd, S. Y., Combs, J. G., Woehr, D. J. et Ketchen, D. J., Jr. (2011). Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *Journal of Applied Psychology*, 96 (3), 443-456.

- Cucina, J. M., Caputo, P. M., Thibodeaux, H. F. et Maclane, C. N. (2012). Unlocking the Key to Biodata Scoring: A Comparison of Empirical, Rational, and Hybrid Approaches at Different Sample Sizes. *Personnel Psychology*, 65 (2), 385-428.
- Cullen, M. J., Sackett, P. R. et Lievens, F. (2006). Threats to the Operational Use of Situational Judgment Tests in the College Admission Process. *International Journal of Selection & Assessment*, 14 (2), 142-155.
- Dalessio, A. T. (1994). Predicting Insurance Agent Turnover Using a Video-Based Situational Judgment Test. *Journal of Business & Psychology*, 9 (1), 23-32.
- Daniellou, F., Grall, J., Martin, C. et Six, F. (2000). *Prescriptions, injonctions et pressions*. Communication présentée au Actes des journées de Bordeaux sur la pratique de l'ergonomie, Université Victor Segalen Bordeaux 2.
- Dawson, K. (2004). *The Call Center Handbook: The Complete Guide to Starting, Running, and Improving Your Call Center* (5 ed.). San Francisco: CMP Books.
- Delery, J. E. et Doty, D. H. (1996). Modes of theorizing in strategic human resource management: Tests of universalistic, contingency, and configurational performance predictions. *Academy of Management Journal*, 39 (4), 802-835.
- Demers, D. L. et Messing, K. (2001). Les tests de sélection : une course à obstacles vers l'égalité économique des femmes. *Recherches féministes*, 14 (1), 15-48.
- Duder, J. C. et Rosenwein, M. B. (2001). Towards "zero abandonments" in call center performance. *European Journal of Operational Research*, 135 (1), 50-56.
- Durivage, A. (2004). La gestion des compétences et la dotation du personnel au Mouvement Desjardins. *Gestion*, 29 (1), 10-18.
- Durivage, A. et Pettersen, N. (2007). L'administration de tests en sélection du personnel. *Gestion*, 4 (32), 20-27.
- Durivage, A., Pettersen, N. et Thibault, J. (2006). *Validité prédictive d'un inventaire de personnalité au travail*. Communication présentée au 13e congrès de l'Association Internationale de Psychologie du Travail de Langue Française (AIPTLF)., Université de Bologne (Italie).
- Edenborough, R. (2007). *Assessment methods in recruitment, selection & performance : a manager's guide to psychometric testing, interviews and assessment centres*. London: Kogan Page.

- England, G. W. (1971). *Development and use of weighted application blanks* Minneapolis: University of Minnesota.
- EPSI. (2015). *Manuel technique : Test pour Centre Contact Client (T3C)*. Gatineau: Compmetrica.
- Farr, J. L. et Tippins, N. T. (2010). Handbook of employee selection: An introduction and overview. Dans J. L. F. N. T. Tippins (Éd.), *Handbook of employee selection* (pp. 1-6). New York: Routledge.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51 (4), 327-358.
- Gatewood, R. D., Feild, H. S. et Barrick, M. R. (2008). *Human resource selection* (6 ed.). Mason, Ohio: Thomson South-Western.
- Gessner, T. L. et Klimoski, R. J. (2006). Making Sense of Situations. Dans J. A. W. R. E. Ployhart (Éd.), *Situational judgment tests: Theory, measurement, and application* (pp. 13-38). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Goffin, R. D. et Olson, J. M. (2011). Is It All Relative?: Comparative Judgments and the Possible Improvement of Self-Ratings and Ratings of Others. *Perspectives on Psychological Science*, 6 (1), 48-60.
- Guérin, G., Le Louarn, J.-Y. et Wils, T. (1988). *L'intégration des ressources humaines à la planification d'entreprise : une justification et un cadre conceptuel*. Montréal: École de relations industrielles, Université de Montréal
- Guion, R. M. (1965). *Personnel testing*. New York McGraw-Hill.
- Guion, R. M. (2010). Employee selection: Musings about its past, present, and future. Dans J. L. F. N. T. Tippins (Éd.), *Handbook of employee selection* (pp. 943-957). New York, NY, US: Routledge/Taylor & Francis Group.
- Gwinner, K. P., Bitner, M. J., Brown, S. W. et Kumar, A. (2005). Service Customization Through Employee Adaptiveness. *Journal of Service Research*, 8 (2), 131-148.
- Harvey, R. J. et Lozada-Larsen, S. R. (1988). Influence of Amount of Job Descriptive Information on Job Analysis Rating Accuracy. *Journal of Applied Psychology*, 73 (3), 457-461.
- Hogan, R., Hogan, J. et Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist*, 51 (5), 469-477.



- Hogan, R., Johnson, J. A. et Briggs, S. R. (1997). *Handbook of personality psychology*. San Diego, CA, US: Academic Press.
- Hogan, T. P. (2012). *Introduction à la psychométrie*. Montréal: Chenelière-éducation.
- Hooper, A. C., Cullen, M. J. et Sackett, P. R. (2006). Operational Threats to the Use of SJTs: Faking, Coaching, and Retesting Issues. Dans J. A. W. R. E. Ployhart (Éd.), *Situational judgment tests: Theory, measurement, and application* (pp. 205-232). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Hough, L. et Dilchert, S. (2010). Personality: Its measurement and validity for employee selection. Dans J. L. F. N. T. Tippins (Éd.), *Handbook of employee selection* (pp. 299-319). New York, NY, US: Routledge/Taylor & Francis Group.
- Hough, L. et Paullin, C. (1994). Construct-oriented scale construction: The rational approach. Dans G. S. Stokes & M. D. Mumford (Éds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 109-145). Palo Alto, CA: CPP Books.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D. et McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75 (5), 581-595.
- Hough, L. M. et Oswald, F. L. (2008). Personality testing and industrial-organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1 (3), 272-290.
- Hunter, J. E. et Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96 (1), 72-98.
- Hunter, J. E., Schmidt, F. L. et Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, 91 (3), 594-612.
- Huselid, M. A. (1995). The Impact of Human Resource Management Practices on Turnover, Productivity, and Corporate Financial Performance. *Academy of Management Journal*, 38 (3), 635-672.
- Joiner, D. A. (2002). Assessment Centers: What's New? *Public Personnel Management*, 31 (2), 179.

- Kahya, E. (2007). The effects of job characteristics and working conditions on job performance. *International Journal of Industrial Ergonomics*, 37 (6), 515-523.
- Kanning, U. P., Grewe, K., Hollenberg, S. et Hadouch, M. (2006). From the subjects' point of view: Reactions to different types of situational judgment items. *European Journal of Psychological Assessment*, 22 (3), 168-176.
- Krokos, K. I., Meade, A. W., Cantwell, A. R., Pond, S. B. et Wilson, M. A. (2004). *Empirical keying of situational judgment tests: Rational and some examples*. Communication présentée au 19<sup>e</sup> conférence annuelle de la Society for Industrial and Organizational Psychology, Chicago, Illinois.
- Landy, F. J. (1983). *The measurement of work performance : methods, theory, and applications*. New York: Academic Press.
- Landy, F. J. et Conte, J. M. (2007). *Work in the 21st century : An introduction to industrial and organizational psychology, 2nd edition*. Malden, MA: Blackwell Publishing.
- Landy, F. J. et Farr, J. L. (1980). Performance Rating. *Psychological Bulletin*, 87 (1), 72-107.
- Legree, P. J., Psotka, J., Tremble, T. et Bourne, D. R. (2005). Using Consensus Based Measurement to Assess Emotional Intelligence. Dans R. S. R. D. Roberts (Éd.), *Emotional intelligence: An international handbook* (pp. 155-179). Ashland, OH, US: Hogrefe & Huber Publishers.
- Lengnick-hall, C. A. et Lengnick-hall, M. L. (1988). Strategic Human Resources Management: A Review of the Literature and a Proposed Typology. *Academy of Management Review*, 13 (3), 454-470.
- Lengnick-Hall, M. L., Lengnick-Hall, C. A., Andrade, L. S. et Drake, B. (2009). Strategic human resource management: The evolution of the field. *Human Resource Management Review*, 19 (2), 64-85.
- Lévy-Leboyer, C. (2002). *Évaluation du personnel : quels objectifs? quelles méthodes?* Paris: Éditions d'Organisation.
- Lievens, F. (2000). Development of an empirical scoring scheme for situational inventories. *European Review of Applied Psychology/Revue Européenne de Psychologie Appliquée*, 50 (1), 117-125.

- Lievens, F. (2006). International Situational Judgment Tests. Dans J. A. W. R. E. Ployhart (Éd.), *Situational judgment tests: Theory, measurement, and application* (pp. 279-300). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Lievens, F., Buyse, T. et Sackett, P. R. (2005). The Operational Validity of a Video-Based Situational Judgment Test for Medical College Admissions: Illustrating the Importance of Matching Predictor and Criterion Construct Domains. *Journal of Applied Psychology, 90* (3), 442-452.
- Lievens, F. et Chan, D. (2010). Practical intelligence, emotional intelligence, and social intelligence. Dans J. L. F. N. T. Tippins (Éd.), *Handbook of employee selection* (pp. 339-359). New York, NY, US: Routledge/Taylor & Francis Group.
- Lievens, F. et De Soete, B. (2012). Simulations. Dans *The Oxford handbook of personnel assessment and selection* (pp. 383-410). New York, NY, US: Oxford University Press.
- Lievens, F. et Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *Journal of Applied Psychology, 96* (5), 927-940.
- Lievens, F., Peeters, H. et Schollaert, E. (2008). Situational judgment tests: a review of recent research. *Personnel Review, 37* (4), 426-441.
- Lievens, F. et Sackett, P. R. (2007). Situational Judgment Tests in High-Stakes Settings: Issues and Strategies with Generating Alternate Forms. *Journal of Applied Psychology, 92* (4), 1043-1055.
- Lievens, F., Sackett, P. R. et Buyse, T. (2009). The effects of response instructions on situational judgment test performance and validity in a high-stakes context. *Journal of Applied Psychology, 94* (4), 1095-1101.
- Lucas, R. W. (2001). *How to be a great call center representative*. Saranac Lake, N.Y.: American Management Association.
- MacLane, C. N., Barton, M. G., Holloway-Lundy, A. E. et Nickels, B. J. (2001). *Keeping score: Expert weights on situational judgment responses*. . Communication présentée au 16e conférence annuelle de la Society for Industrial and Organizational Psychology, San Diego, Californie.
- Marcoux, G. (2010). L'expérience de travail en centre d'appels : aux confins de l'engagement et de la distanciation. *Relations Industrielles, 65* (4), 654-672.

- McDaniel, M. A., Hartman, N. S., Whetzel, D. L. et Grubb, W. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60 (1), 63-91.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A. et Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86 (4), 730-740.
- McDaniel, M. A. et Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, 9 (1-2), 103-113.
- McDaniel, M. A., Psozka, J., Legree, P. J., Yost, A. P. et Weekley, J. A. (2011). Toward an understanding of situational judgment item validity and group differences. *Journal of Applied Psychology*, 96 (2), 327-336.
- McDaniel, M. A. et Whetzel, D. L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence*, 33 (5), 515-525.
- McDaniel, M. A., Whetzel, D. L., Hartman, N. S., Nguyen, N. T. et Grubb Iii, W. L. (2006). Situational Judgment Tests: Validity and an Integrative Model. Dans J. A. W. R. E. Ployhart (Éd.), *Situational judgment tests: Theory, measurement, and application* (pp. 183-203). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Motowidlo, S. J., Dunnette, M. D. et Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75 (6), 640-647.
- Motowidlo, S. J., Hooper, A. C. et Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, 91 (4), 749-761.
- Motowidlo, S. J. et Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology*, 66 (4), 337-344.
- Motowidlo, S. J., Borman, W. C. et Schmit, M. J. (1997). A Theory of Individual Differences in Task and Contextual Performance. *Human Performance*, 10 (2), 71.
- Mount, M. K., Barrick, M. R. et Strauss, J. P. (1994). Validity of observer ratings of the big five personality factors. *Journal of Applied Psychology*, 79 (2), 272-280.
- Mumford, M. D. (1999). Construct Validity and Background Data: Issues, Abuses, and Future Directions. *Human Resource Management Review*, 9 (2), 117.

- Mumford, M. D. et Owens, W. A. (1987). Methodology Review: Principles, Procedures, and Findings in the Application of Background Data Measures. *Applied Psychological Measurement*, 11 (1), 1-31.
- Mumford, M. D. et Whetzel, D. L. (1997). Background data. Dans D. L. W. G. R. Wheaton (Éd.), *Applied measurement methods in industrial psychology* (pp. 207-239). Palo Alto, CA, US: Davies-Black Publishing.
- Mumford, T. V., Van Iddekinge, C. H., Morgeson, F. P. et Campion, M. A. (2008). The Team Role Test: Development and validation of a team role knowledge situational judgment test. *Journal of Applied Psychology*, 93 (2), 250-267.
- Muros, J. P. (2008). *Know the score: An exploration of keying and scoring approaches for situational judgment tests*. Dissertation, University of Minnesota.
- Murphy, K. R. et Cleveland, J. (1995). *Understanding performance appraisal : social, organizational, and goal-based perspectives*. Thousand Oaks, Calif.: Sage.
- Nguyen, N. T., Biderman, M. D. et McDaniel, M. A. (2005). Effects of Response Instructions on Faking a Situational Judgment Test. *International Journal of Selection and Assessment*, 13 (4), 250-260.
- Nguyen, N. T., McDaniel, M. A. et Biderman, M. D. (2002). *Response instructions in situational judgment tests: Effects on faking and construct validity*. Communication présentée au 17e conférence annuelle de la Society for Industrial and Organizational Psychology, Toronto, Canada.
- Nunnally, J. C. (1978). *Psychometric Theory*. New-York: McGraw-Hill.
- O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, W. L., III et Lawrence, A. (2007). Incremental Validity of Situational Judgment Tests for Task and Contextual Job Performance. *International Journal of Selection and Assessment*, 15 (1), 19-29.
- O'Sullivan, M. (2007). Trolling for trout, trawling for tuna. Dans G. Matthews, Zeidner, M., & Roberts, R. R. (Éd.), *The science of emotional intelligence—Knowns and unknowns* (pp. 258–287). New York: Oxford University Press.
- Ones, D. S., Viswesvaran, C. et Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81 (6), 660-679.

- Ones, D. S., Viswesvaran, C. et Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78 (4), 679-703.
- Organ, D. W. (1997). Organizational Citizenship Behavior: It's Construct Clean-Up Time. *Human Performance*, 10 (2), 85-97.
- Ostroff, C. (1991). Training Effectiveness Measures and Scoring Schemes : A Comparison. *Personnel Psychology*, 44 (2), 353-374.
- Paulhus, D. L. (1988). Measurement and control of response bias. Dans J. P. Robinson, P. Shaver & L. Wrigitsman (Éds.), *Measures of personality and social psychological attitudes* (Vol. 1, pp. 17-60). New York: Academic Press.
- Paullin, C. et Hanson, M. A. (2001). *Comparing the validity of rationally-derived and empirically-derived scoring keys for a situational judgement inventory*. Communication présentée au 16e conférence annuelle de la Society for Industrial and Organizational Psychology, San Diego, California.
- Pearlman, K. et Sanchez, J. I. (2010). Work analysis. Dans J. L. F. N. T. Tippins (Éd.), *Handbook of employee selection* (pp. 73-98). New York, NY, US: Routledge/Taylor & Francis Group.
- Peeters, H. et Lievens, F. (2005). Situational Judgment Tests and their Predictiveness of College Students' Success: The Influence of Faking. *Educational and Psychological Measurement*, 65 (1), 70-89.
- Pettersen, N. (2000). *Évaluation du potentiel humain dans les organisations élaboration et validation d'instruments de mesure*. Sainte-Foy, Québec: Presses de l'Université du Québec.
- Phillips, J. (2012). *Strategic staffing*. Upper Saddle River, N.J. : Pearson Prentice Hall.
- Ployhart, R. E. et Ehrhart, M. G. (2003). Be Careful What You Ask For: Effects of Response Instructions on the Construct Validity and Reliability of Situational Judgment Tests. *International Journal of Selection & Assessment*, 11 (1), 1-16.
- Ployhart, R. E. et MacKenzie Jr., W. I. (2011). Situational judgment tests : A critical review and agenda for the future Dans *APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization* (pp. 237-252). Washington, D.C.: American Psychological Association

- Ployhart, R. E. et Weekley, J. A. (2010). Strategy, selection, and sustained competitive advantage. Dans J. L. F. N. T. Tippins (Éd.), *Handbook of employee selection* (pp. 195-212). New York, NY: Routledge.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y. et Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88* (5), 879-903.
- Pulakos, E. D., Arad, S., Donovan, M. A. et Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology, 85* (4), 612-624.
- Putka, D. J. et Sackett, P. R. (2010). Reliability and validity. Dans J. L. F. N. T. Tippins (Éd.), *Handbook of employee selection* (pp. 9-49). New York, NY, US: Routledge/Taylor & Francis Group.
- Saba, T. et Dolan, S. (2013). *La gestion des ressources humaines : tendances, enjeux et pratiques actuelles*. Montréal: Pearson.
- Sackett, P. R. et Lievens, F. (2008). Personnel selection. *Annual Review of Psychology, 59*, 419-450.
- Salovey, P. et Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality, 9*, 185-211.
- Schmidt, F. L. et Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124* (2), 262-274.
- Schmidt, F. L., Ones, D. S. et Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology, 43*, 627-670.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L. et Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80* (5), 607-620.
- Schmitt, N. (2012). *The Oxford handbook of personnel assessment and selection*. New York: Oxford University Press.
- Schmitt, N. et Chan, D. (2006). Situational Judgment Tests: Method or Construct? Dans *Situational judgment tests: Theory, measurement, and application* (pp. 135-155). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

- Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., et al. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology, 94* (6), 1479-1497.
- Schmitt, N. et Oswald, F. L. (2006). The impact of corrections for faking on the validity of noncognitive measures in selection settings. *Journal of Applied Psychology, 91* (3), 613-621.
- Schulze, R., Wilhem, O. et Kyllonen, P. C. (2007). Approaches to the assessment of emotional intelligence. Dans G. Matthews, Zeidner, M., & Roberts, R. R. (Éd.), *The science of emotional intelligence—Knowns and unknowns* (pp. 199–229). New York: Oxford University Press.
- Sekiou, L. (2001). *Gestion des ressources humaines* (2e éd. ed.). Bruxelles: Bruxelles : De Boeck Université.
- Silzer, R. et Jeanneret, R. (2011). Individual psychological assessment: A practice and science in search of common ground. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 4* (3), 270-296.
- Spencer, L. M. et Spencer, S. M. (1993). *Competence at Work : Models for Superior Performance*. New York: John Wiley & Sons, Inc.
- St-Onge, S. (2013). *Relever les défis de la gestion des ressources humaines* (4e éd. ed.). Montréal: Montréal : G. Morin.
- St-Onge, S. et Haines, V. (2007). *Gestion des performances au travail : bilan des connaissances*. Bruxelles: De Boeck.
- St-Sauveur, C., Girouard, S. et Goyette, V. (2014). Use of Situational Judgment Tests in Personnel Selection: Are the different methods for scoring the response options equivalent? *International Journal of Selection & Assessment, 22* (3), 225-239.
- StatistiqueCanada. (2005). *Statistiques des télécommunications*: Le Quotidien.
- Stemler, S. E. et Sternberg, R. J. (2006). Using Situational Judgment Tests to Measure Practical Intelligence. Dans J. A. W. R. E. Ployhart (Éd.), *Situational judgment tests: Theory, measurement, and application* (pp. 107-131). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Sternberg, R. J., Wagner, R. K., Williams, W. M. et Horvath, J. A. (1995). Testing common sense. *American Psychologist, 50* (11), 912-927.



- Stevens, M. J. et Campion, M. A. (1999). Staffing Work Teams: Development and Validation of a Selection Test for Teamwork Settings. *Journal of Management*, 25 (2), 207-228.
- Tabachnick, B. G. et Fidell, L. S. (2007). *Using multivariate statistics (5th ed.)*. Boston, MA: Allyn & Bacon/Pearson Education.
- Taylor, P., Mulvey, G., Hyman, J. et Bain, P. (2002). Work Organization, Control and the Experience of Work in Call Centres. *Work, Employment & Society*, 16 (1), 133-150.
- Thorndike, E. L. (1920). Intelligence and its uses. *Harper's Magazine*, 140 (227-235).
- Tippins, N. T., Beaty, J., Drasgow, F., Gibson, W. M., Pearlman, K., Segall, D. O., et al. (2006). Unproctored internet testing in employment settings. *Personnel Psychology*, 59 (1), 189-225.
- Tippins, N. T., Papinchock, J. M. et Solberg, E. C. (2010). Decisions in developing and selecting assessment tools. Dans J. L. F. N. T. Tippins (Éd.), *Handbook of employee selection* (pp. 363-376). New York, NY, US: Routledge/Taylor & Francis Group.
- Tremblay, M. et Wills, T. (2005). La mobilisation des ressources humaines : une stratégie de rassemblement des énergies de chacun pour le bien de tous. *Gestion* (2), 37.
- Truxillo, D. M. et Bauer, T. N. (2011). Applicant reactions to organizations and selection systems. Dans *APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization* (pp. 379-397). Washington, DC: American Psychological Association.
- Tziner, A., Joanis, C. et Murphy, K. R. (2000). A comparison of three methods of performance appraisal with regard to goal properties, goal perception, and ratee satisfaction. *Group & Organization Management*, 25 (2), 175-190.
- Vallerand, R. J. et Thill, E. (1993). *Introduction à la psychologie de la motivation*. Laval, Québec: Éditions Études Vivantes.
- Van Jaarsveld, D., Frost, A. C. et Walker, D. (2007). The Canadian Contact Centre Industry : Strategy, Work Organization and Human Resource Management. *Vancouver : The Global Call Centre Industry Project*.
- Vincent, R. et McKeown, L. (2008). Tendances dans l'industrie des centres d'appels. *Ottawa : Statistiques Canada*.

- Vinchur, A. J. (2007). A History of Psychology Applied to Employee Selection. Dans *Historical perspectives in industrial and organizational psychology* (pp. 193-218). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Vinchur, A. J., Schippmann, J. S., Switzer, F. S., III et Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology*, 83 (4), 586-597.
- Viswesvaran, C., Ones, D. S. et Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81 (5), 557-574.
- Viswesvaran, C., Schmidt, F. L. et Ones, D. S. (2005). Is There a General Factor in Ratings of Job Performance? A Meta-Analytic Framework for Disentangling Substantive and Error Influences. *Journal of Applied Psychology*, 90 (1), 108-131.
- Weekley, J. A., Harding, R., Creglow, A. et Ployhart, R. E. (2004). Scoring situational judgment tests : Does the middle matter ? [Paper presented at the 19th annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL].
- Weekley, J. A. et Jones, C. (1997). Video-Based Situational Testing. *Personnel Psychology*, 50 (1), 25-49.
- Weekley, J. A. et Jones, C. (1999). Further Studies of Situational Tests. *Personnel Psychology*, 52 (3), 679-700.
- Weekley, J. A. et Ployhart, R. E. (2005). Situational Judgment: Antecedents and Relationships with Performance. *Human Performance*, 18 (1), 81-104.
- Weekley, J. A. et Ployhart, R. E. (2006). *Situational judgment tests: Theory, measurement, and application*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Weekley, J. A., Ployhart, R. E. et Harold, C. M. (2004). Personality and Situational Judgment Tests Across Applicant and Incumbent Settings: An Examination of Validity, Measurement, and Subgroup Differences. [Article] *Human Performance*, 17 (4), 433-461.
- Weekley, J. A., Ployhart, R. E. et Holtz, B. C. (2006). On the Development of Situational Judgment Tests: Issues in Item Development, Scaling, and Scoring. Dans J. A. W. R. E. Ployhart (Éd.), *Situational judgment tests: Theory, measurement, and application* (pp. 157-182). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Wernimont, P. F. et Campbell, J. P. (1968). SIGNS, SAMPLES, AND CRITERIA. *Journal of Applied Psychology*, 52 (5), 372-376.

Whetzel, D. L. et McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, 19 (3), 188-202.

Whetzel, D. L., McDaniel, M. A. et Nguyen, N. T. (2008). Subgroup Differences in Situational Judgment Test Performance: A Meta-Analysis. [Article] *Human Performance*, 21 (3), 291-309.

Winsted, K. F. (2000). Service behaviors that lead to satisfied customers. *European Journal of Marketing*, 34 (3/4), 399-417.

Wright, P. M., Dunford, B. B. et Snell, S. A. (2001). Human resources and the resource based view of the firm. *Journal of Management*, 27 (6), 701-721.

Zapf, D., Isic, A., Bechtoldt, M. et Blau, P. (2010). What is typical for call centre jobs? Job characteristics, and service interactions in different call centres. *European Journal of Work and Organizational Psychology*, 12 (4), 311-340.

## ANNEXE A : DIMENSIONS EVALUEES PAR LE T3C

### Dimensions évaluées par le Test pour Centre Contact Client (T3C) basées sur le dictionnaire de compétences Compmetrica ©

#### Aspect fonctionnel du service à la clientèle

58	<b>Orientation vers la clientèle</b>	Vise à établir des relations positives à long terme avec les clients et à maintenir un haut niveau de satisfaction en offrant des services ou des produits de qualité répondant à leurs désirs, à leurs besoins et à leurs attentes
12	<b>Savoir transmettre l'information</b>	Sait transmettre efficacement de l'information en utilisant des techniques de communication en fonction du contexte et de l'auditoire
42	<b>Orientation vers l'action</b>	Prend des décisions et agit rapidement malgré les contraintes auxquelles il doit faire face

#### Aspect personnel du service à la clientèle

14	<b>Communication interpersonnelle</b>	Démontre de la considération, de la compréhension et du respect envers les gens qu'il côtoie dans son milieu de travail
35	<b>Contrôle de soi</b>	Sait maîtriser ses émotions et ses impulsions, ainsi que d'agir de façon appropriée dans des situations difficiles ou désagréables
17	<b>Empathie</b>	Est sensible aux difficultés rencontrées et aux besoins exprimés par les gens.
11	<b>Savoir écouter</b>	Écoute attentivement et activement afin de permettre aux gens d'exprimer en toute liberté et avec confiance leur point de vue et de leur transmettre sa volonté de vouloir les comprendre.
18	<b>Tolérance</b>	Respect et interagit facilement avec les personnes qui ont des opinions, des valeurs et des approches différentes des siennes.

#### Aspect vente

60	<b>Savoir négocier</b>	Sait utiliser de très bonnes connaissances des approches et des techniques de négociation afin de parvenir à des ententes qui sont avantageuses pour toutes les parties impliquées
23	<b>Persuasion</b>	Sait convaincre les gens d'adopter une ligne de pensée, une position ou façon de faire
43	<b>Persévérance</b>	Démontre de la persévérance dans son travail, malgré les difficultés et les obstacles, afin d'atteindre les objectifs fixés

#### Aspect relié aux résultats et à la performance

31	<b>Rigueur</b>	Respecte soigneusement et entièrement les procédures et les méthodes de travail prescrites afin de garantir un niveau de qualité et de précision supérieurs
32	<b>Autonomie</b>	Sait travailler seul, avec un minimum d'encadrement, tout en utilisant à bon escient la marge de manœuvre qui lui est donnée
38	<b>Savoir s'adapter</b>	Accueille positivement les changements et ajuste son comportement de façon appropriée.
44	<b>Résistance au stress</b>	Sait garder son calme et être objectif et efficace malgré la présence de circonstances difficiles ou de stress intense.
34	<b>Confiance en soi</b>	Avoir confiance en soi, incluant avoir confiance en ses habiletés, ses capacités et son jugement
58	<b>Orientation vers les résultats</b>	Consacre des efforts importants afin de réaliser les objectifs fixés et considère que l'atteinte des résultats représente une préoccupation primordiale

## ANNEXE B : STATISTIQUES DESCRIPTIVES POUR LE T3C

### Statistiques descriptives pour le T3C Tronc commun et modules complémentaires (Partie 1) (N = 423)

Compétences	Nombre d'items	Minimum	Maximum	Moyenne	Écart-type	Alpha	Symétrie	Kurtose
Autonomie	8	6	40	20,5	7,2	0,60	0,09	-0,08
Communication interpersonnelle	8	2	35	17,7	7	0,45	0,41	0,06
Confiance en soi	8	0	36	18,5	7,8	0,36	0	-0,34
Contrôle de soi	6	7	30	22,8	6,3	0,61	-0,81	-0,33
Empathie	7	2	35	13,9	6,2	0,58	0,92	0,86
Orientation client	8	7	35	22	6,3	0,25	-0,09	-0,23
Orientation résultats	5	0	25	11,3	6,2	0,68	-0,08	-0,93
Orientation vers l'action	8	2	35	15,5	7,6	0,65	0,6	-0,15
Persévérance	6	0	30	18,7	7,1	0,52	-0,14	-0,72
Persuasion	7	0	35	21,4	6,3	0,26	-0,18	-0,18
Résistance au stress	6	0	30	14,7	6,1	0,22	0,06	-0,12
Rigueur	11	0	55	32	10,2	0,41	-0,18	-0,19
Savoir écouter	7	2	35	21,9	6,6	0,38	-0,25	0,32
Savoir informer	9	0	45	21,1	8,2	0,50	-0,02	-0,57
Savoir négocier	5	2	25	15,4	4,5	0,15	0,06	-0,49
Savoir s'adapter	11	0	55	14	6,8	0,24	0,74	0,26
Tolérance	8	5	40	22,1	7,2	0,50	0,03	-0,4

## ANNEXE C : STRUCTURE FACTORIELLE DU T3C

### Structure factorielle du T3C Tronc commun et modules complémentaires (Partie 1)

	Aspects reliés à la performance	Aspects personnels	Orientation clientèle
<b>Compétences</b>			
Persévérance	0,836		
Persuasion	0,827		
Gestion du stress	0,764		
Rigueur	0,757		
Savoir négocier	0,814		
Orientations résultats	0,56		
Savoir s'adapter	0,545		
Empathie		0,726	
Orientations vers l'action		0,781	
Savoir écouter		0,843	
Tolérance		0,831	
Communication interpersonnelle		0,643	
Confiance en soi		0,688	
Orientations client			0,852
<b>Eigenvalue</b>	3,965	3,511	1,064

61% de variance expliquée

Extraction Method: Principal Component Analysis.

Rotation Method: Oblimin with Kaiser Normalization.

## ANNEXE D : CONTRE-VALIDATION DE L'APPROCHE EMPIRIQUE

---

### Contre-validation de l'approche empirique par le biais d'un échantillon de trente grilles de correction différentes

Essai	R	R <sup>2</sup>	Z Fisher
D6	0,022	,000	0,02
D17	0,029	,001	0,03
D9	0,052	,003	0,05
D26	0,054	,003	0,05
D18	0,058	,003	0,06
D23	0,079	,006	0,08
D8	0,085	,007	0,09
D25	0,09	,008	0,09
D12	0,1	,010	0,10
D15	0,112	,013	0,11
D13	0,114	,013	0,11
D30	0,126	,016	0,13
D24	0,128	,016	0,13
D27	0,133	,018	0,13
D2	0,138	,019	0,14
D22	0,143	,021	0,14
D19	0,154	,024	0,15
D21	0,158	,025	0,16
D29	0,162	,026	0,16
D5	0,165	,027	0,17
D10	0,168	,028	0,17
D4	0,175	,030	0,18
D11	0,188	,035	0,19
D3	0,188	,035	0,19
D20	0,189	,036	0,19
D1	0,19	,036	0,19
D16	0,197	,039	0,20
D7	0,206	,043	0,20
D14	0,207	,043	0,20
D28	0,278	,077	0,21

## ANNEXE E : CRITERES D'ÉVALUATION DU RENDEMENT SUPERIEUR

---

### Critères permettant d'identifier les employés ayant un rendement supérieur

#### *Démarche suggérée*

Afin d'identifier les employés ayant un rendement supérieur, nous vous suggérons d'utiliser deux approches :

Dans le cas des employés pour lesquels vous disposez de données objectives (ex : quantité de ventes, nombre d'appels moyens, etc.), identifiez un ou deux critères qui vous apparaissent essentiels.

À partir de ces critères, identifiez la liste des personnes qui ont obtenu les meilleurs résultats.

Examinez ensuite cette liste et déterminez, pour chacune des personnes identifiées, si elle se situe dans le 3<sup>e</sup> ou le 4<sup>e</sup> groupe (voir l'échelle d'évaluation plus bas). Si c'est le cas, cette personne peut être considérée comme ayant un rendement supérieur. En effet, non seulement elle obtient des résultats supérieurs mais en plus, elle se démarque par rapport aux compétences jugées essentielles pour le poste.

Dans le cas des employés pour lesquels vous ne disposez pas de données objectives, nous vous suggérons de n'utiliser que l'échelle d'évaluation présentée plus bas.

#### *Échelle d'évaluation*

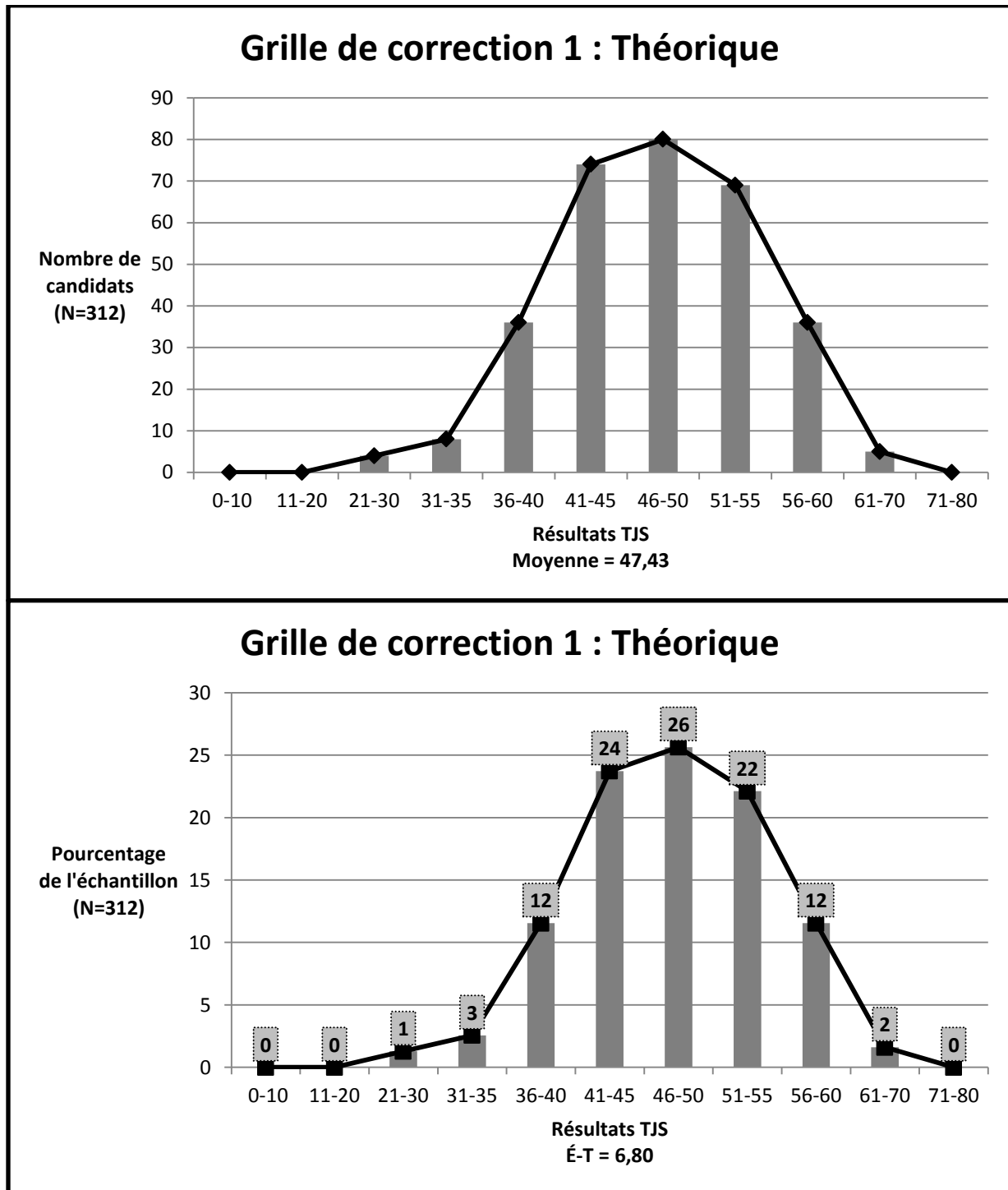
En tenant compte de l'échelle d'évaluation suivante, identifiez où se situe l'employé par rapport à des personnes que vous avez pu observer et qui œuvrent dans des emplois similaires en Centre d'appel.

1	2	3	4
Performance dans le quart inférieur	Performance un peu sous la moyenne	Performance un peu au-dessus de la moyenne	Performance dans le quart supérieur
Par rapport à des employé(e)s que vous avez observés et qui occupent des emplois similaires, l'employé(e) se situe dans le 25 % inférieur.	Par rapport à des employé(e)s que vous avez observés et qui occupent des emplois similaires, l'employé(e) se situe dans la moyenne inférieure.	Par rapport à des employé(e)s que vous avez observés et qui occupent des emplois similaires, l'employé(e) se situe dans la moyenne supérieure.	Par rapport à des employé(e)s que vous avez observés et qui occupent des emplois similaires, l'employé(e) se situe dans le 25 % supérieur.

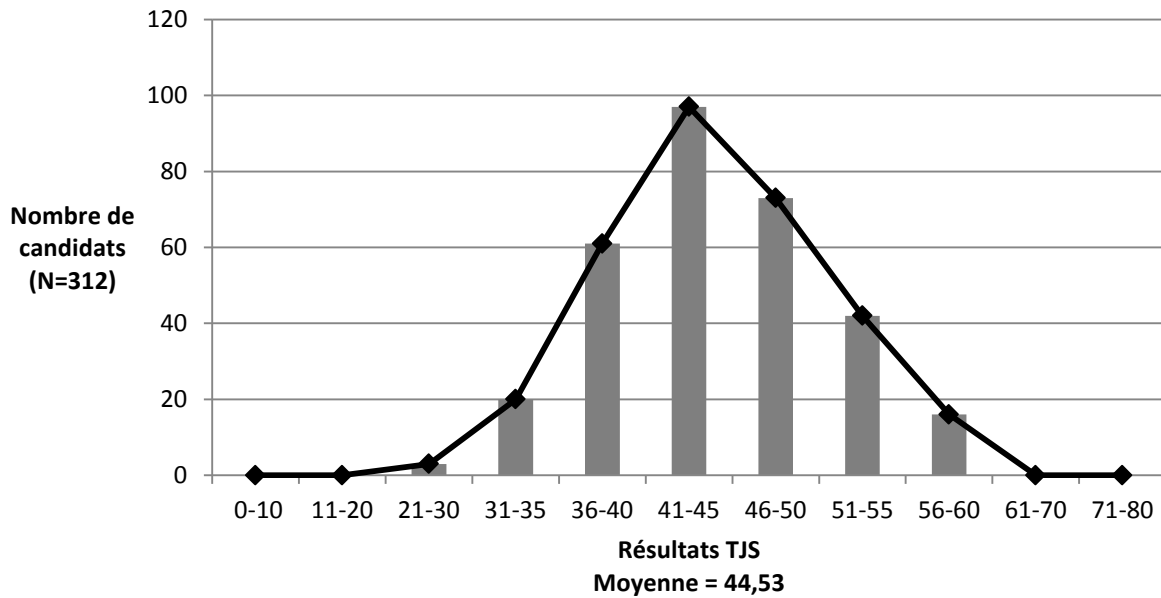


## ANNEXE F : HISTOGRAMMES DE DISTRIBUTION DES RESULTATS

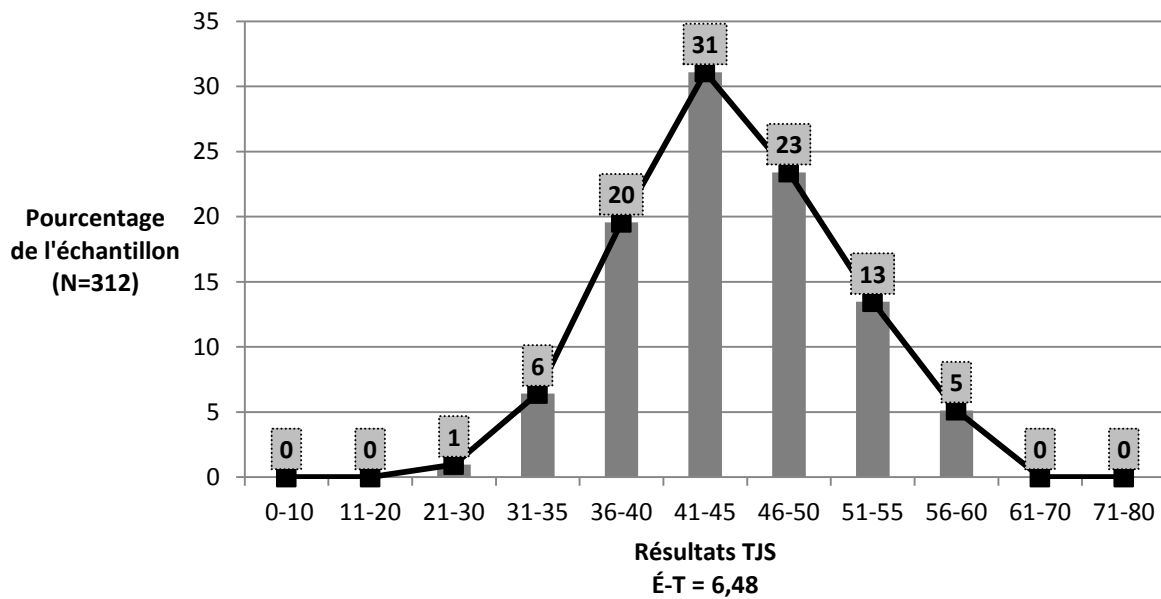
### Histogrammes de distribution des résultats par grille de correction



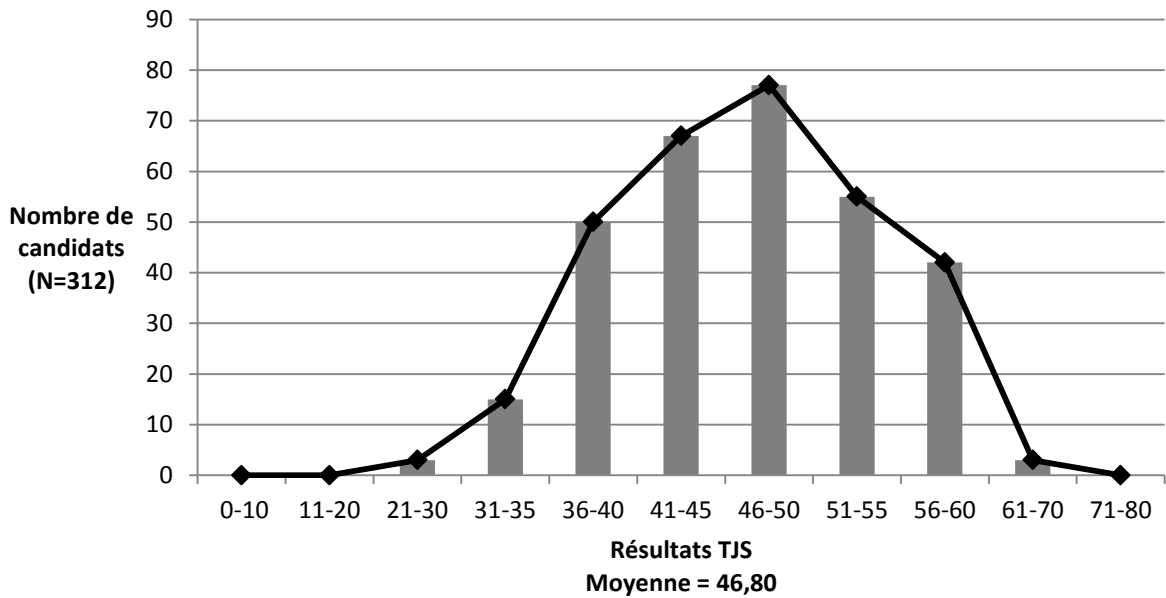
## Grille de correction 2 : Experts financiers



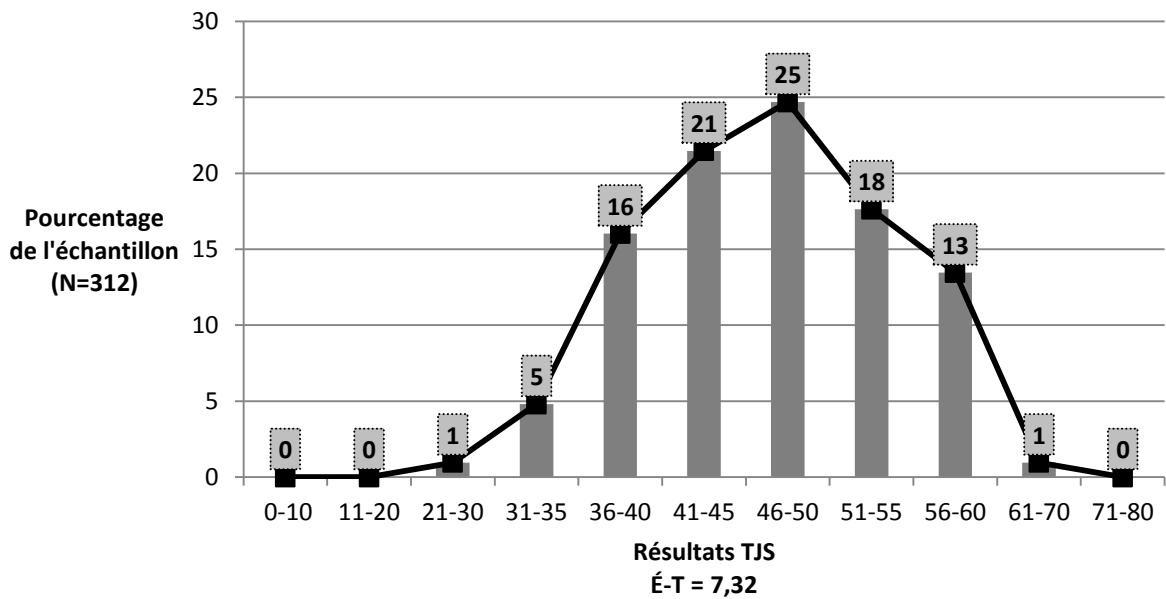
## Grille de correction 2 : Experts financiers



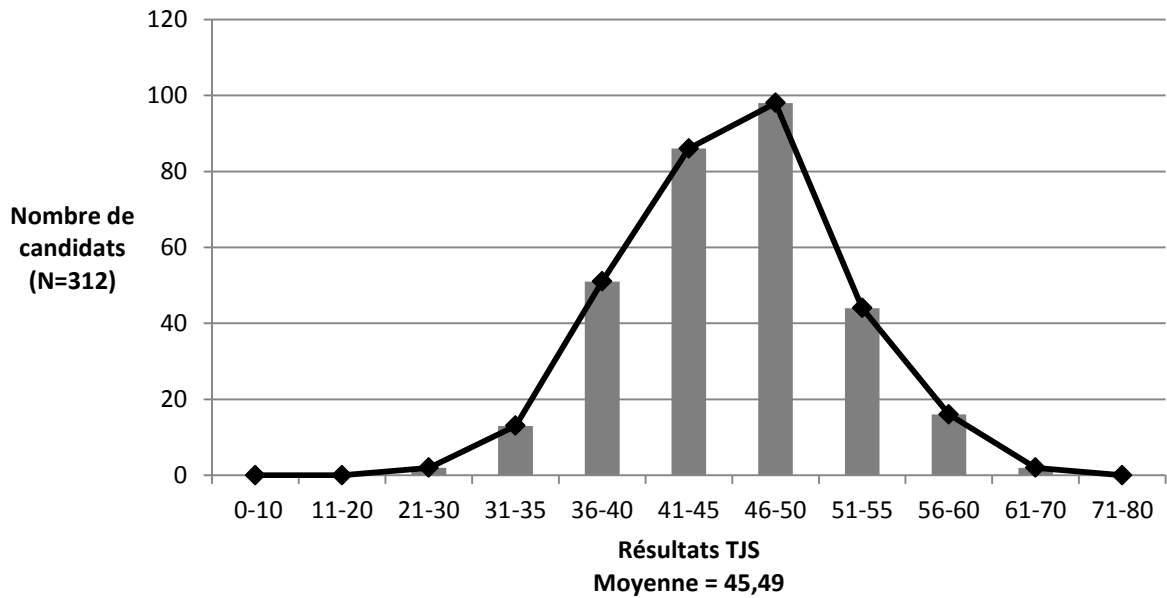
### Grille de correction 3 : Experts énergie



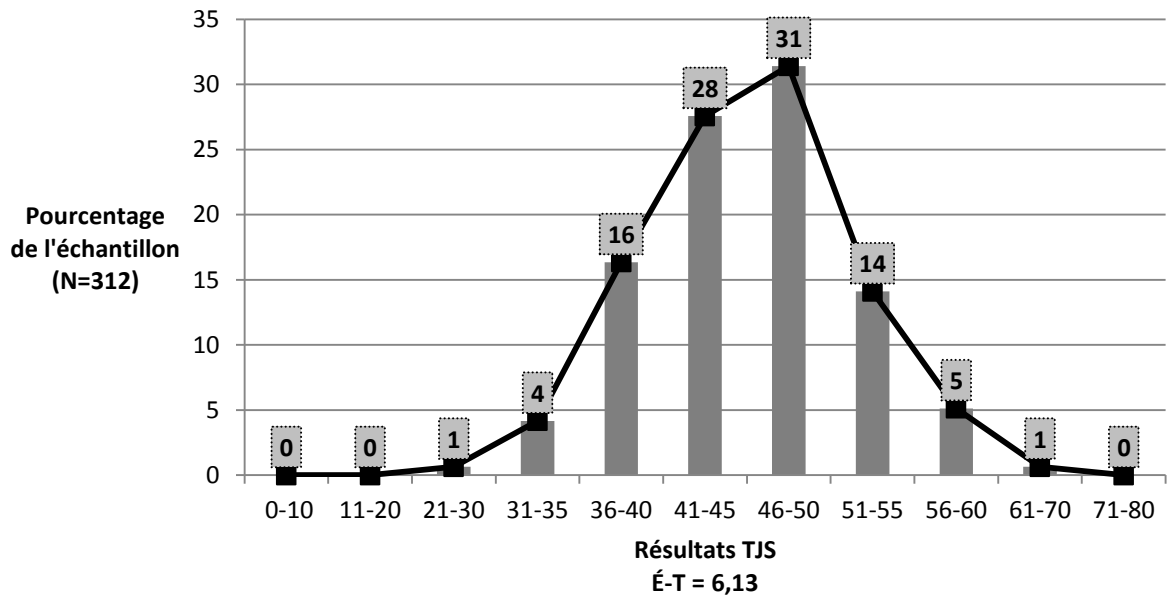
### Grille de correction 3 : Experts énergie



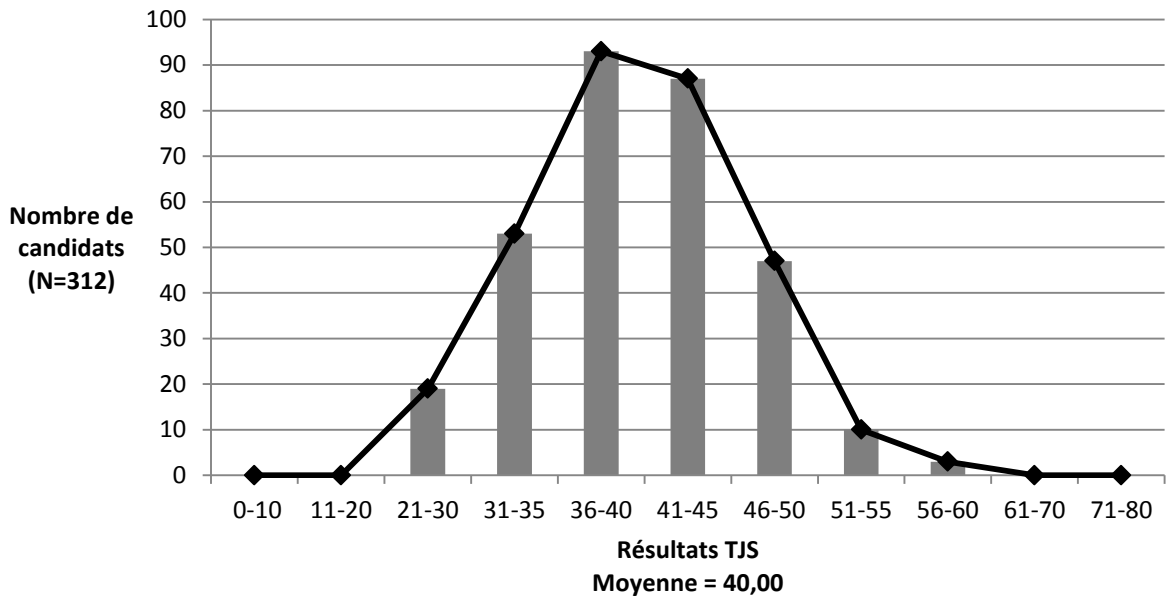
### Grille de correction 4 : Experts publics



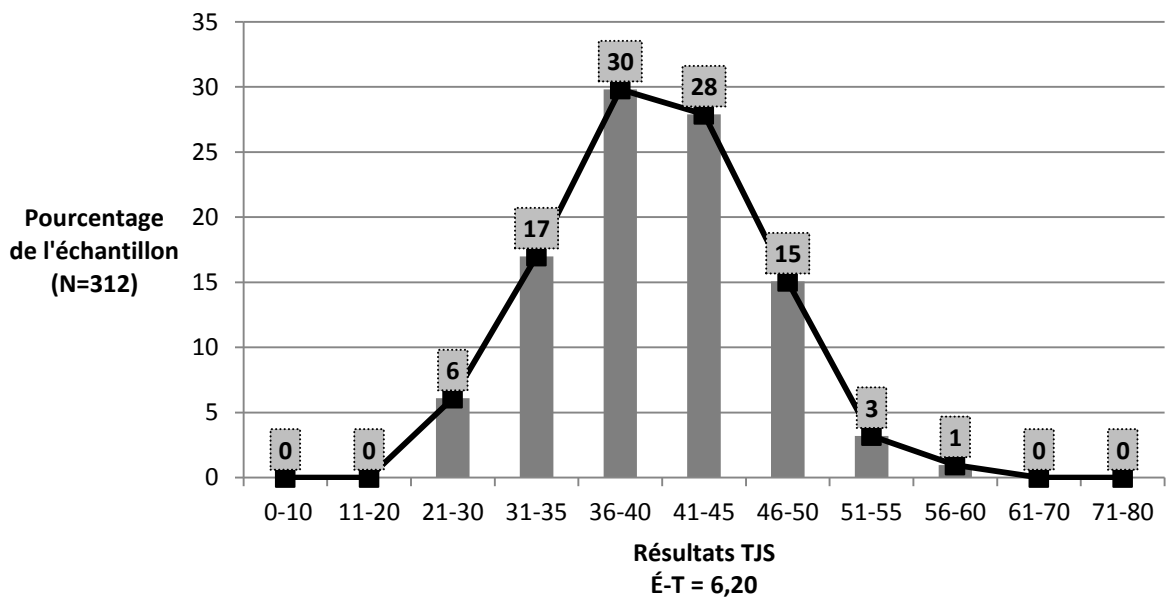
### Grille de correction 4 : Experts publics



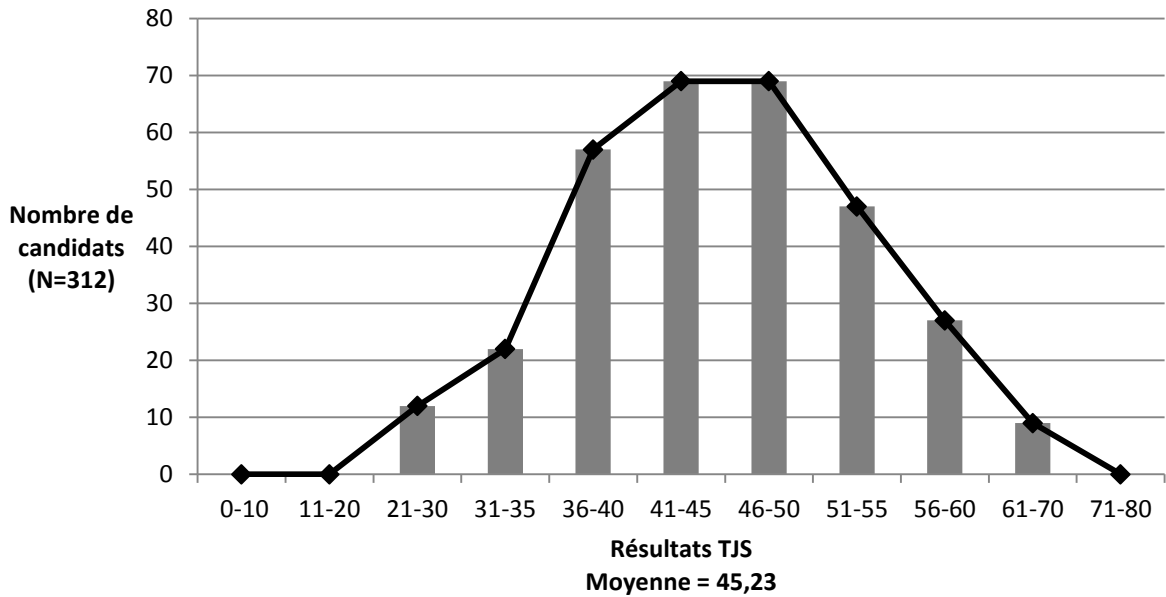
### Grille de correction 6 : Empirique



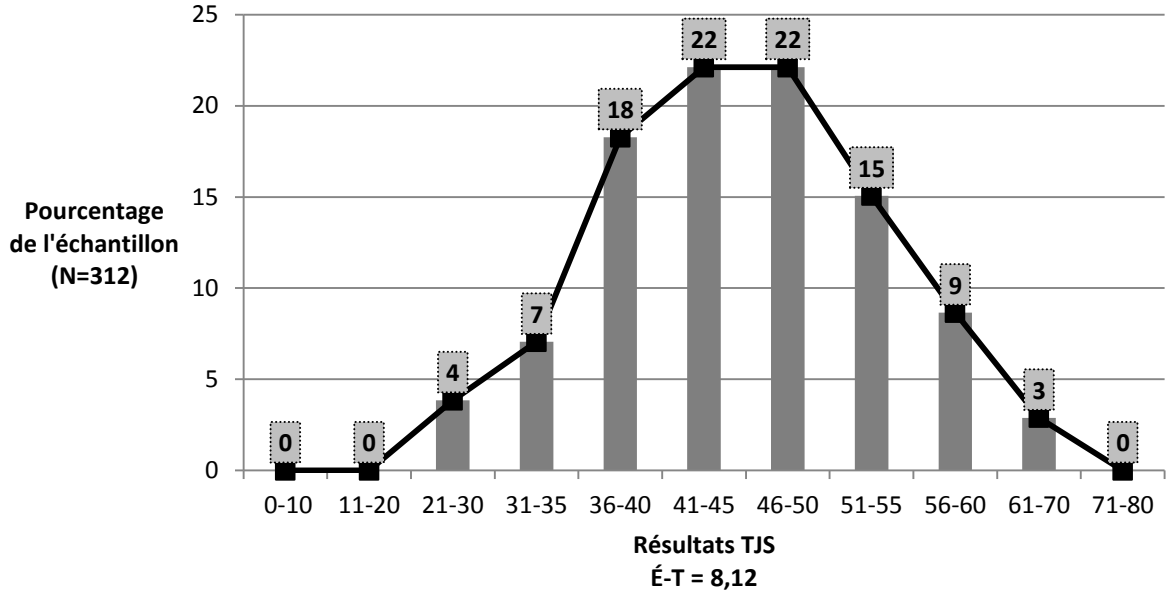
### Grille de correction 6 : Empirique



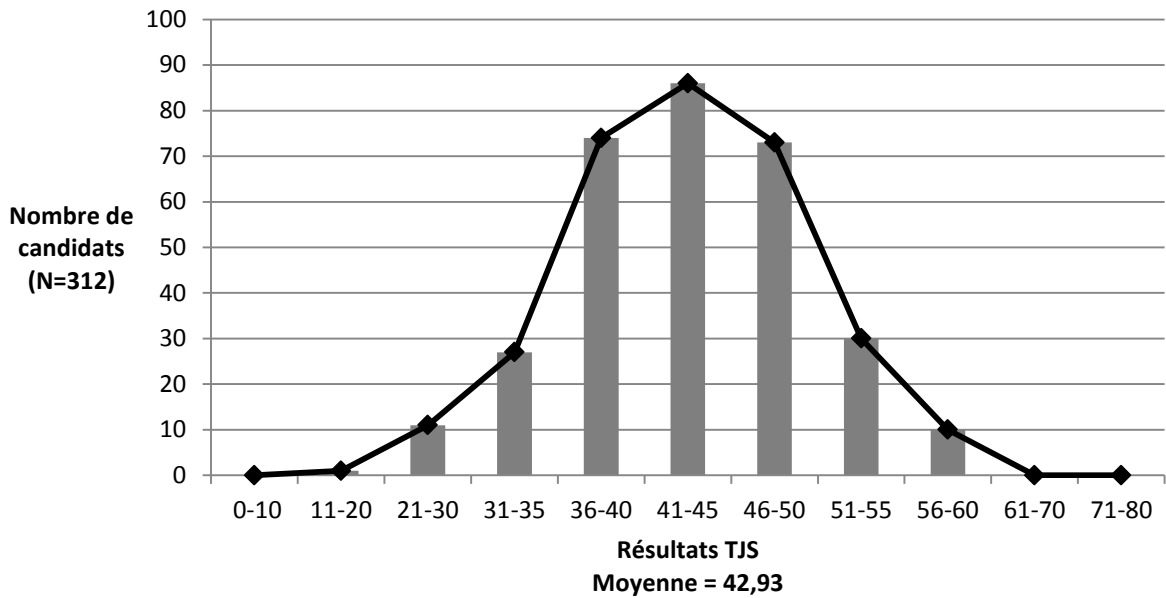
### Grille de correction 19 : Empirique



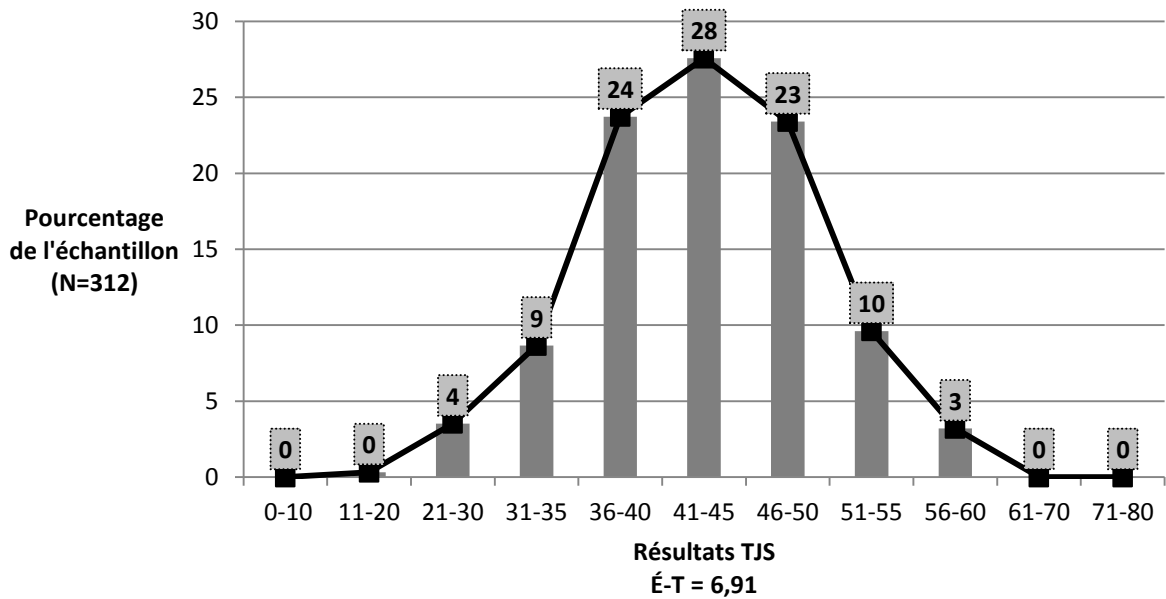
### Grille de correction 19 : Empirique



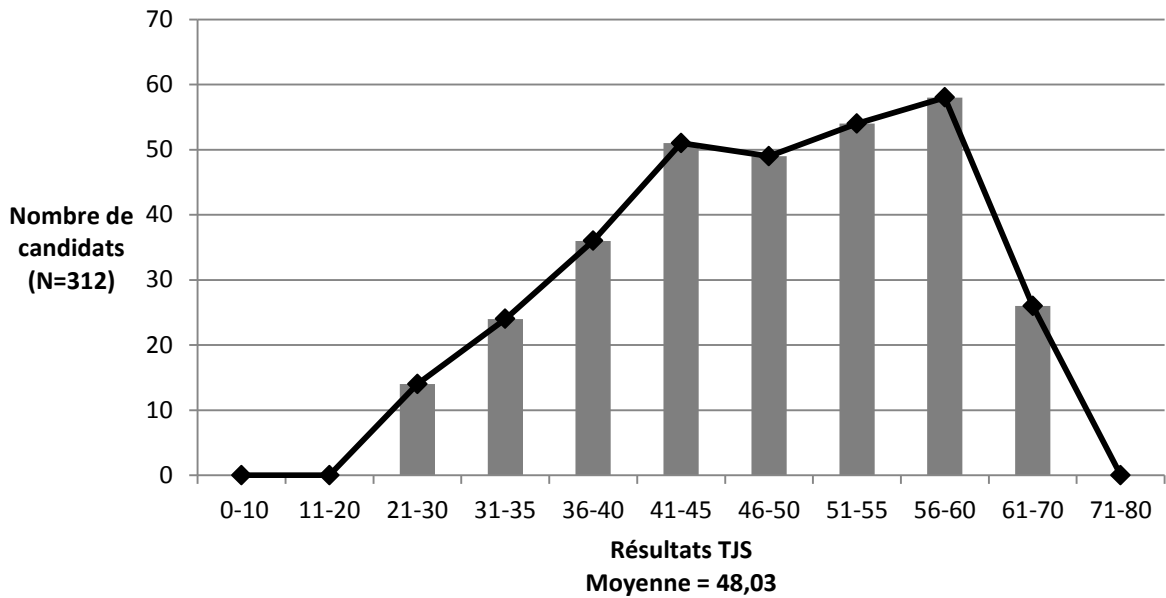
### Grille de correction 28 : Empirique



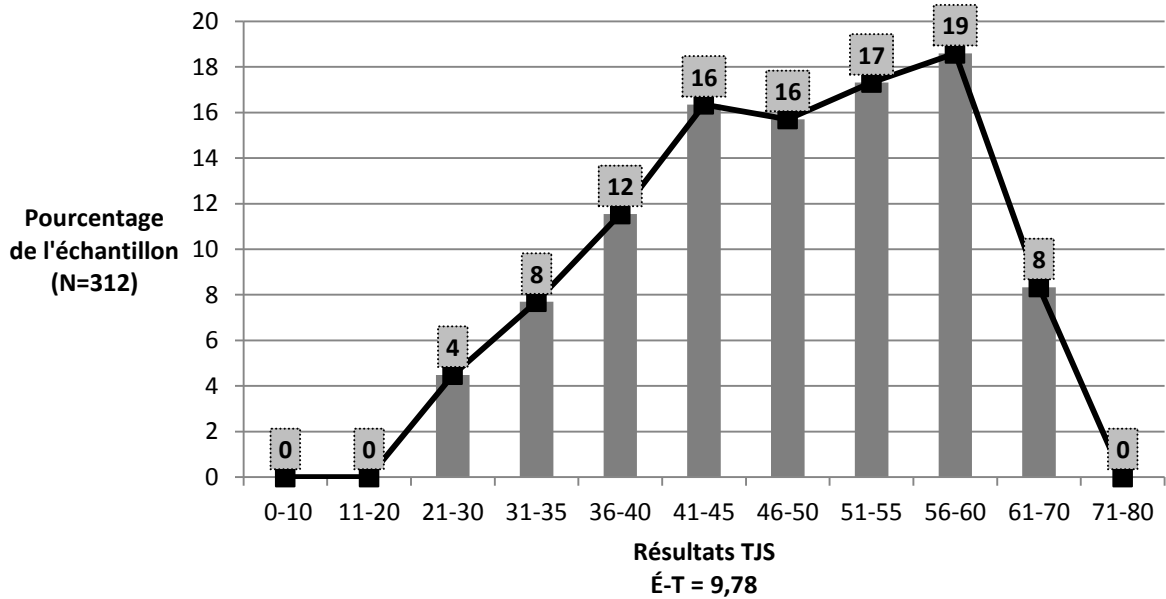
### Grille de correction 28 : Empirique



### Grille de correction 8 : Hybride



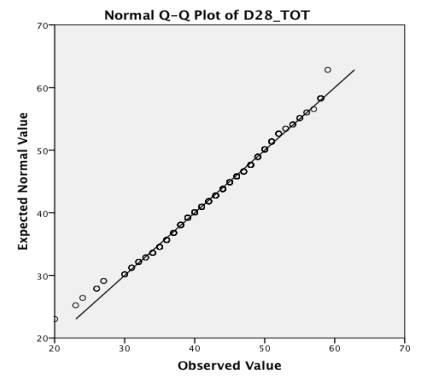
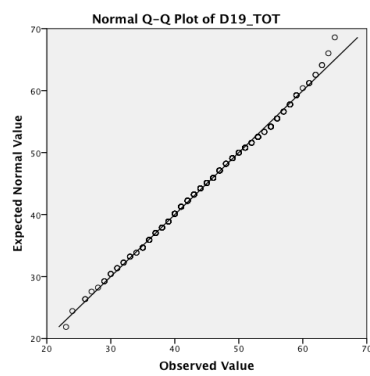
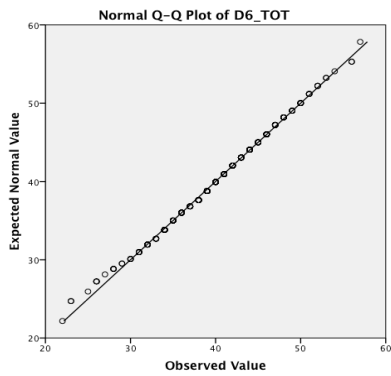
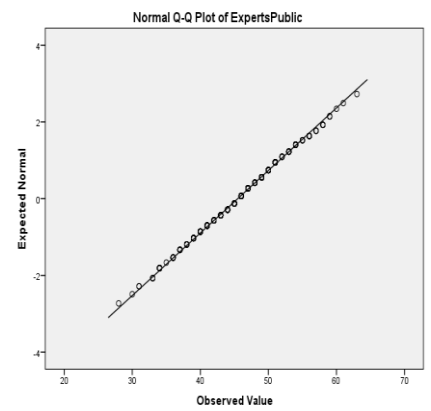
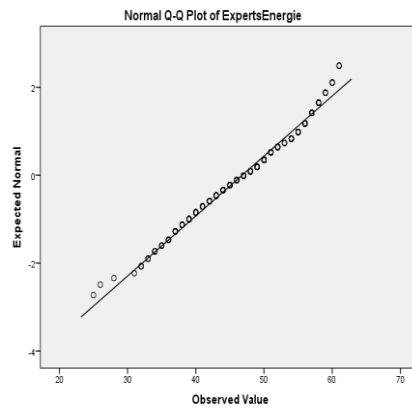
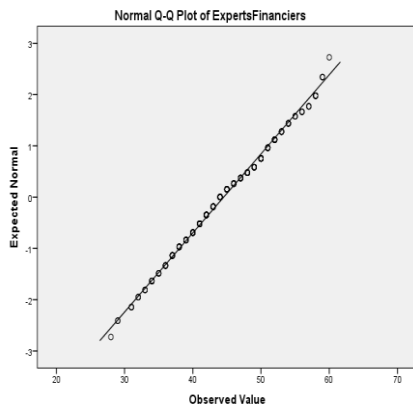
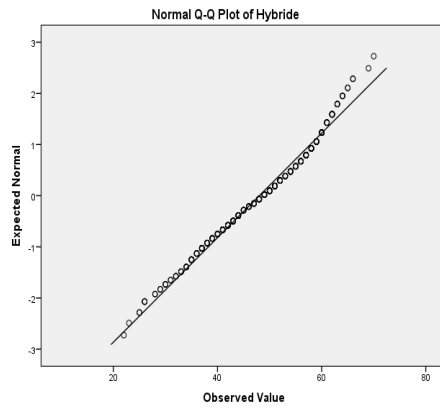
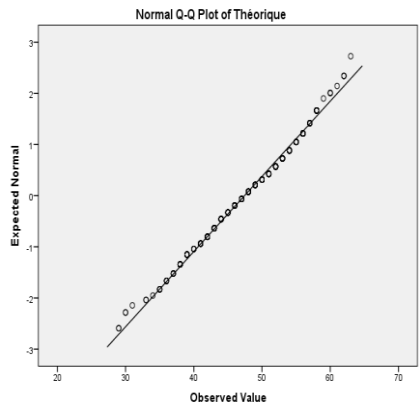
### Grille de correction 8 : Hybride





# ANNEXE G : DIAGRAMMES QUANTILE-QUANTILE

## Diagrammes quantile-quantile pour l'ensemble des grilles de correction



ANNEXE H : DIAGRAMMES DE DISPERSION

Diagrammes de dispersion pour l'ensemble des grilles de correction

