

# **Validité de construit d'une épreuve de jeu de rôle dans le cadre d'un centre d'évaluation chez des officiers de marine.**

## **The construct validity of a role play exercise in a navy officer's assessment center**

Borteyrou<sup>1</sup>, X., Rasclé, N., Collomb, P. & Bruchon-Schweitzer, M.

[xavier.borteyrou@univ-fcomte.fr](mailto:xavier.borteyrou@univ-fcomte.fr)

### *Résumé :*

Cette étude porte sur la validité d'une épreuve de jeu de rôle. L'échantillon se compose de 125 officiers de marine, évalués à la fois par psychologues et des officiers de marine. Tous les évaluateurs ont participé à des séances de formation (utilisation de grilles comportementales, formation sur les biais de l'évaluation, entraînement sur des séquences vidéo). La procédure d'évaluation incluait une épreuve de jeu de rôle, des questionnaires d'intelligence (test de facteur g) et de personnalité (cinq grands). Dans un premier temps, des analyses factorielles exploratoires ont été conduites à partir de la grille d'observation du jeu de rôle et 4 facteurs ont été identifiés (attitude autoritaire, considération, franchise, communication orale). Dans une perspective nomologique, nous avons établi des liens entre ces facteurs et les scores obtenus en personnalité et en intelligence. Les résultats suggèrent que le jeu de rôle et les questionnaires évaluent des construits psychologiques différents.

### *Abstract :*

The present study was designed to obtain validity estimates for a role-play test. Participants were 125 french Navy officers who were rated by a pool of professional assessors and psychologists. All the assessors received recurring training sessions, focusing on the behavioral checklist, on rating errors, and on share frame of reference. The assessment procedure included role play exercise, cognitive ability scale (g factor) and personality scale (big five factors). First, exploratory factor analyses were conducted on the data gathered, and four factors were identified (authoritarianism, oral communication, consideration with others and frankness). In a nomological perspective, we also analysed the links between the exercises dimensions, personality inventory and intelligence scale. The findings suggest that role play dimension, personality and intelligence seem to measure different constructs.

**Mots clefs :** Epreuve de jeu de rôle, personnalité, intelligence, validité de construit, techniques de recrutement.

**Key-words:** Role play test, personality, intelligence, construct validity, personnel selection techniques.

---

<sup>1</sup> Laboratoire de Psychologie EA 4139. 3 ter place de la Victoire. 33 076 Bordeaux cedex.

## 1. Introduction

Les mises en situations (MS) sont de plus en plus utilisées dans le cadre du recrutement professionnel et/ou des bilans de compétences. Le principe de la mise en situation consiste à mettre un individu dans une situation professionnelle fictive mais réaliste, et à recueillir des informations sur son comportement. La situation correspond à un échantillon de travail c'est-à-dire à une situation représentative d'une activité professionnelle donnée. Aujourd'hui, on ne peut pas parler de mises en situations professionnelles sans évoquer les centres d'évaluation (CE), qui comprennent en général plusieurs types de MS individuelles et collectives (in basket, épreuve de groupe, jeu de rôle) ainsi que des épreuves psychométriques (personnalité et intelligence). Les épreuves “ *in basket* ” sont des épreuves individuelles dans lesquelles les évalués doivent répondre (par écrit) à un certain nombre de sollicitations en temps limité. Dans ces épreuves, les demandes faites aux candidats sont plus ou moins complexes et urgentes. D'une manière générale, un test *in basket* est une épreuve qui sert à appréhender la capacité à hiérarchiser et à traiter les priorités (Baron & Janman, 1996). L'épreuve de groupe est une épreuve collective dans laquelle les candidats doivent atteindre un objectif commun et qui permet notamment d'évaluer les qualités relationnelles et le travail en équipe (Baron & Janman, 1996).

Enfin, le jeu de rôle est une situation dans laquelle le candidat doit interagir avec une autre personne, les deux protagonistes ayant généralement des objectifs différents. Cette épreuve est systématiquement citée dans les CE car elle correspond à des situations d'évaluation et d'apprentissage privilégiées. Malgré un usage très fréquent, on possède peu d'information sur la nature de ce qui est évalué dans le cadre d'un jeu de rôle qui est un cas particulier des épreuves de mise en situation. Bien que les jeux de rôle aient déjà fait l'objet de rares tentatives de validation (Squires, Torkel, Smither, & Ingate, 1991), on ne sait pas bien si ces épreuves permettent d'évaluer des aspects cognitifs, conatifs ou alors des compétences spécifiques.

## 2. Problématique

### 2. 1. La validité des mises en situation

Si notre travail se rapporte à la validité de construit d'une épreuve de jeu de rôle, il nous semble important d'aborder rapidement la notion de validité de contenu car il est difficile d'envisager ces deux validités de manière séparée (Binning & Barrett, 1989).

### 2. 1. 1. La validité de contenu des MS

La validité de contenu consiste à se demander si les comportements évalués dans le cadre d'une mise en situation sont représentatifs des conduites professionnelles. Le postulat qui sous-tend la méthode des MS est la théorie de la cohérence comportementale (Wernimont & Campbell, 1968). Selon cette théorie, les comportements mis en place dans ces mises en situation sont représentatifs des comportements adoptés ultérieurement dans des situations professionnelles similaires. Dans une MS, il s'agit donc de recréer des échantillons professionnels les plus proches possibles des situations professionnelles futures du candidat. Etant donné que les MS sont construits sur le principe de la cohérence comportementale, la validité de contenu est le point fort de cette technique (Sackett, 1987). Il s'agit une étape qualitative du processus de validation qui repose sur un échantillonnage pertinent des conduites. Ainsi, la mise en place d'analyses de postes minutieuses ainsi que le « réalisme » des mises en situation sont considérés comme des pré requis nécessaires pour une validité de contenu satisfaisante (Callinan & Robertson, 2000).

### 2. 1. 2. La validité de construit des MS

La validité de construit consiste à se demander à quoi correspondent les construits psychologiques évalués dans le cadre d'une MS. Or, depuis 1979, différents chercheurs ne parviennent pas à décrire précisément ces construits (Archambeau, 1979; Lievens, 2002; Lievens & Klimoski, 2001; Sackett & Dreher, 1982). Pour cela, ils ont eu recours aux matrices multi-traits multi-méthodes (MTMM) ou aux analyses factorielles (AF).

L'objectif des MTMM est de vérifier si des construits psychologiques (leadership par exemple) mesurés lors d'épreuves différentes (jeu de rôle, épreuve de groupe par exemple) sont bien corrélées entre eux (Archambeau, 1979; Joyce, Thayer, & Pond, 1994; Turnage & Muchinsky, 1982). Les études utilisant ces techniques aboutissent toutes aux mêmes résultats avec des coefficients de validité convergente faibles et des coefficients de validité divergente qui sont élevés. Autrement dit, les corrélations entre des construits différents mesurés par une même épreuve sont fortes, alors que les corrélations entre les mêmes construits mesurés avec des épreuves différentes sont faibles. Ces premières recherches indiquent donc une faible validité de construit des MS car on ne parvient pas à identifier des construits psychologiques indépendants des épreuves utilisées (on parle dans ce cas de variance instrumentale).

D'autres chercheurs (Sackett & Dreher, 1982) ont tenté de décrire l'existence de construits psychologiques évalués avec des MS par analyse factorielle exploratoire. Ils mènent des AF à partir des scores obtenus sur toutes

les dimensions, ceci pour l'ensemble des épreuves de mise en situation d'un CE. Dans leur article, Sackett & Dreher (1982) mettent en évidence que les composantes issues des AF ne correspondent pas à des dimensions interprétables (communication orale, leadership ou résistance au stress par exemple) mais aux épreuves qui ont servi à les évaluer (épreuve de groupe, in basket par exemple). Ces derniers résultats seront ensuite contre-validés sur des CE différents avec des populations différentes (Kleinmann & Koller, 1997; Robertson, Gratton, & Sharpley, 1987; Silverman, Dalessio, Woods, & Johnson, 1986). Récemment, ces résultats ont été confirmés par des études ayant recours à des analyses factorielles confirmatoires. Lance, Lambert, Gewin, Lievens, & Conway (2004) conduisent des analyses confirmatoires à partir de résultats issus de 34 CE. En moyenne, ils observent que la part de variance expliquée par les épreuves ( $\lambda^2=.52$ ) est supérieure à celle expliquée par les dimensions censées y être évaluées ( $\lambda^2=.14$ ).

Les MTMM et les analyses factorielles suivent des logiques différentes mais ces deux techniques ont permis de mettre en évidence la faible validité de construit des MS utilisées conjointement dans le cadre d'un CE.

## 2. 2. Une solution pour améliorer la validité de construit des CE

Dans le cadre de la présente recherche, nous avons pris en compte des solutions méthodologiques proposées récemment par la littérature pour améliorer la validité de construit des MS. Ces solutions incluent le recours à des grilles comportementales qui vont permettre de rendre la procédure d'évaluation plus descriptive qu'évaluative (Thornton & Zorich, 1980). Des chercheurs ont déjà mis en évidence que le recours à des grilles comportementales augmentait la validité convergente qui passe de .24 à .43 (Reilly, Henry, & Smither, 1990) ce qui permet de réduire la variance due aux exercices (variance instrumentale). Il ne s'agit plus d'évaluer le candidat de manière globale mais d'établir des listes très précises où les comportements ne sont pas inférés mais observés (Hennessy, Mabey, & Warr, 1998). Cette démarche implique de connaître très précisément les comportements qui sont pertinents à évaluer pour un poste donné. En effet, si l'utilisation des grilles comportementales rend l'évaluation plus rigoureuse, elle suppose que tous les comportements pertinents figurent bien dans cette grille. Si cette dernière condition n'est pas remplie, les évaluateurs peuvent « manquer » des informations importantes (Donahue, Truxillo, Cornwell, & Gerrity, 1997). L'utilisation de grilles comportementales implique donc un travail d'analyse de poste très précis et spécifique à chaque MS.

### 2. 3. La nécessité d'identifier ce qui est mesuré lors d'une mise en situation

Il existe aujourd'hui un certain consensus concernant la dénomination des construits psychologiques évalués dans le cadre des MS d'un CE. Les dimensions les plus fréquemment citées sont la communication écrite et orale, la considération, l'énergie, l'influence, l'organisation ou la résistance au stress par exemple (Arthur, Day, McNelly, & Edens, 2003). Par contre, les recherches donnent peu de précision concernant la nature ou le contenu de ces dimensions. Ces caractéristiques sont régulièrement citées mais nous possédons peu d'informations sur ce qu'elles recouvrent. Notre étude a donc pour objectif d'éclairer la nature des construits psychologiques évalués dans le cadre d'une épreuve de jeu de rôle afin de mieux comprendre ce qui est réellement appréhendé lors d'une épreuve de mise en situation.

Pour remplir cet objectif, nous avons dans un premier temps décidé d'adopter une démarche à visée exploratoire qui consiste à ne pas postuler a priori l'existence de dimensions pré identifiées pour noter les candidats. Dans un premier temps, il s'agit de décrire comment les évaluations se structurent dans le cadre d'une mise en situation afin de noter les candidats en conséquences, ceci sans *a priori* sur le nombre et le contenu des caractéristiques évaluées.

Enfin, nous souhaitons éclairer la nature de ces construits dans une perspective nomologique. Ce principe qui a été proposée très récemment, consiste à comparer les dimensions obtenues lors d'une MS à des construits psychologiques déjà clairement identifiés comme la personnalité et l'intelligence par exemple (Salgado & Moscoso, 2002; Shore, Thornton, & Shore, 1990). Cette procédure est considérée comme une méthode alternative aux techniques d'analyse factorielle et de MMTM pour mettre à l'épreuve la validité de construit des MS, afin d'identifier plus précisément la nature des construits évalués.

## 3. Méthode

### 3. 1. Participants

Notre échantillon comprend 125 officiers de la Marine Nationale ayant le grade de commandant (âge moyen : 41.2 ans, écart-type = 4.6 ans). Les stagiaires font partie de quatre promotions (1997, 1998, 1999 et 2000) du Collège Interarmées de Défense (CID) de l'école militaire de Paris. Le CID est une formation de 1 an qui donne ensuite aux stagiaires la possibilité d'accéder à des postes de haut niveau. L'accès au CID est extrêmement sélectif et se fait sur des critères d'expérience et de connaissances professionnelles. La sélection est

interne et fait suite à des examens écrits et oraux particulièrement exigeants à l'issue desquels seulement 5% des candidats sont retenus. Les stagiaires occupent des niveaux hiérarchiques assez proches (la majorité d'entre eux est "capitaine de frégate", ce qui correspond au 2<sup>ème</sup> échelon du grade de Commandant). Ils ont tous un niveau scolaire élevé (formation initiale à l'Ecole Navale). Les participants sont invités à participer à un CE expérimental qui comprend une épreuve in basket, un jeu de rôle, un entretien semi structuré centré sur le travail et des tests psychométriques (intelligence et personnalité). Notre recherche est présentée comme « un moyen d'améliorer la politique de gestion du personnel de la Marine » et nous nous engageons à restituer les résultats à chaque participant. De plus, nous insistons sur le fait que les données recueillies ne sont pas communiquées à la hiérarchie

### 3. 2. Procédure

L'épreuve du jeu de rôle est évaluée par deux observateurs (un observateur psychologue et un observateur de la Marine Nationale qui ont été préalablement formés par nos soins). Ce dernier est plus gradé que le stagiaire et ne le connaît pas (nous disposons d'un vivier de 10 évaluateurs formés à l'évaluation que nous invitons en fonction de leur disponibilité). Le scénario est tiré d'une situation réelle dans laquelle un officier chargé de la gestion du personnel (le stagiaire) doit recevoir un officier qui n'est pas satisfait de l'affectation qu'on lui propose (ce rôle est joué par un compère). Ce dernier a eu un déroulement de carrière exemplaire et il trouve que le poste proposé ne correspond pas à ses compétences. Le candidat dispose de deux documents (une description très brève de la carrière de l'officier, une injonction de son supérieur hiérarchique qui lui intime l'ordre d'être persuasif). Ces documents sont fournis au dernier moment et le candidat décide du moment où il va accueillir le demandeur. Ce dernier possède un scénario très précis sur ce qu'il doit dire : rapide présentation de sa carrière, évocation d'une situation familiale difficile, possibilité de quitter la Marine Nationale si on le force à accepter cette affectation. Le candidat est donc mis dans une situation difficile où il doit défendre l'intérêt de l'institution tout en respectant la légitimité de la demande de l'impétrant. Sa marge de manœuvre est donc très réduite.

La négociation est limitée à 30 minutes à partir du moment où le stagiaire prend connaissance du dossier. L'épreuve est filmée afin que les deux observateurs puissent revenir sur des moments importants de la négociation. Chaque observateur remplit sa grille individuellement puis les deux grilles sont comparées par les observateurs afin de renseigner une grille consensuelle remplie conjointement.

### 3. 3. Outils

#### **La grille d'évaluation du jeu de rôle**

La grille d'observation utilisée pour évaluer le jeu de rôle a été construite suite à des analyses de postes menées selon la méthode des incidents critiques (Flanagan, 1954). Cette méthode consiste à relever tous les comportements qui sont utilisés dans une situation particulière et reconnus comme efficaces ou inefficaces. Ainsi, plusieurs entretiens ont été menés auprès de responsables de la gestion du personnel afin d'établir une grille pertinente et exhaustive de ces comportements. Pour chacun d'eux, les évaluateurs doivent cocher la case correspondante : comportement absent (jamais observé), comportement présent (observé une ou deux fois), comportement marqué (trois fois ou plus). Ces réponses ont été codées 1 (Absent), 2 (Observé) et 3 (Marqué).

#### **Les indicateurs de personnalité**

Le NEO PI-R (Costa, McCrae, & Rolland, 1998) est l'instrument le plus utilisé pour évaluer les cinq grands facteurs (Salgado, 2003). Il s'agit d'une version révisée du NEO PI, éditée aux Etats Unis en 1985 par Costa & McCrae. Ce questionnaire comporte 240 items qui évaluent 5 dimensions générales (Névrosisme, Extraversion, Ouverture, Agréabilité, Conscience) et 30 facettes plus spécifiques de la personnalité. Pour chaque proposition, le répondant doit indiquer son degré d'accord sur une échelle de Likert en 5 points. La validité de construit de ce questionnaire a été éprouvée (Rolland, 2004) ainsi que sa validité prédictive (Lévy-Leboyer, 2004; Salgado, 2003).

#### **Les indicateurs d'efficacité mentale**

Le test de facteur g de R. B. Cattell (1974) vise à mesurer l'aptitude générale de déduction de relations. Il peut être défini comme “ *Une capacité mentale générale qui, entre autres choses, implique la faculté de raisonner, planifier, résoudre des problèmes, penser de manière abstraite, comprendre des idées complexes et apprendre rapidement* ” (Gottfredson, 2002). En matière de recrutement, les tests de facteur g sont reconnus comme étant les plus prédictifs de la performance professionnelle (Schmidt, 2002). De plus, le facteur g d'intelligence est bien identifié comme étant au sommet d'une organisation hiérarchique de l'intelligence (Huteau & Lautrey, 1999).

## 4. Résultats

### 4.1. La structure factorielle de l'épreuve de jeu de rôle : une approche exploratoire

Les épreuves de Jeu de Rôle, ainsi que les grilles utilisées ont été les mêmes entre 1999 et 2001. Chaque épreuve est évaluée par 2 personnes qui remplissent d'abord séparément une grille d'observation puis se réunissent pour coter de manière consensuelle. Nous disposons donc de 291 observations sur l'ensemble de ces années.

Dans un premier temps, nous avons réalisé une analyse descriptive des items afin de ne retenir que ceux qui se distribuent sur les trois modalités. Dans une étude similaire, Kendall et Stuart (1958) ont proposé d'éliminer les items dont le skewness (indice d'asymétrie) approche 2 (ou -2) et dont le kurtosis (indice d'aplatissement) est supérieur à 4. Ce critère sera pris en compte ici bien qu'il soit utilisé assez rarement dans les recherches actuelles, car il est très contraignant (Gorsuch, 1997). Nous avons cependant revu à la hausse le seuil d'acceptabilité des items afin de ne pas nous priver d'un trop grand nombre d'informations. Nous avons finalement éliminé les items dont le skewness approchait de 3 ou -3 et le kurtosis était supérieur à 4. Cette précaution est en effet nécessaire afin d'éliminer tous les items qui ne différencient pas suffisamment les individus observés. Les items ayant reçu systématiquement le score 1 ou le score 3 n'apportent effectivement aucune information en termes de différences individuelles. De plus, ces items seraient artificiellement corrélés entre eux. Dans le cadre de l'analyse factorielle, cela donnerait lieu à des pseudo-facteurs constitués d'items dont la variance est nulle. Cette première étape nous a conduit à supprimer 8 items non discriminatifs sur 48 (*ne se baser que sur des a priori ; Etre distrait ; fuir le regard de l'interlocuteur ; utiliser des comportements non verbaux non adaptés à la situation ; s'exprimer de manière laconique ; présenter un discours illogique ; rester passif ; faire preuve de malveillance pour aboutir*).

Nous avons soumis les 40 items restants à une analyse factorielle en maximum de vraisemblance suivie de rotations obliques de type promax. Nous avons choisi une analyse factorielle « pure » car notre démarche est entièrement exploratoire et une analyse en Composantes Principales serait dans notre cas trop contraignante. Nous avons choisi des rotations obliques car il n'y a pas de raison de penser que les dimensions recherchées sont indépendantes. Cette méthode d'extraction nous a permis d'identifier 6 facteurs dont la valeur propre est supérieure à 1 (critère de Kaiser). Ces résultats sont présentés au tableau 1.

Tableau 1 : Saturation des items de la grille du jeu de rôle sur les 6 facteurs extraits par analyse factorielle par maximum de vraisemblance, suivie de rotations promax.

	F 1	F 2	F 3	F 4	F 5	F 6
<b>Valeur propre</b>	<b>6.75</b>	<b>3.70</b>	<b>1.89</b>	<b>1.69</b>	<b>1.17</b>	<b>1.02</b>
<b>Pourcentage de variance expliquée</b>	<b>15.70</b>	<b>8.60</b>	<b>4.40</b>	<b>3.92</b>	<b>2.71</b>	<b>2.37</b>
Chercher à gagner en s'opposant à l'autre	0,72					
Agresser	0,65					
Se poser comme détenteur de la vérité	0,62					
Montrer que l'autre a tort	0,61					
Réfuter d'emblée	0,52					
Chercher à aboutir en dominant	0,51					
Contester le propos de l'autre	0,48					
Montrer que l'on a raison	0,43					
Chercher à aboutir en affaiblissant	0,41					
Faire preuve de malveillance pour aboutir	0,30					
Orienter son esprit vers la compréhension de		0,74				
Tenir compte des réactions des interlocuteurs		0,68				
Accorder de l'estime à ce qu'il dit		0,63				
Avoir une attitude d'ouverture		0,61				
Fait preuve d'une conception interactive		0,60				
Accepter et intégrer d'autres idées		0,60				
Chercher à comprendre		0,60				
Entrer dans la logique de l'autre		0,58				
Engager une discussion constructive		0,55				
Avoir le réflexe de questionner		0,52				
Etre chaleureux		0,41				
Acquiescer d'emblée		0,34				
Faire preuve de ruse pour aboutir			-0,64			
Chercher à aboutir en manipulant			-0,56			
Se perdre dans les détails			-0,54			
Esquiver la logique de l'interlocuteur			-0,47			
Encaisser ce que dit l'autre, subir				-0,57		
Etre pris au dépourvu par les questions				-0,49		
S'engager à douter				-0,48		
Présenter des signes d'anxiété				-0,39		
S'exprimer avec aisance					0,63	

	F 1	F 2	F 3	F 4	F 5	F 6
Être attentif et concentré					0,63	
Présenter un discours structuré					0,56	
Parler de manière concise et claire					0,56	
S'exprimer en développant son discours					0,55	
Chercher à expliquer plutôt qu'à influencer					0,50	
Se baser sur des faits pour argumenter					0,45	
Réaliser un dosage équitable du temps de parole					0,30	
Parler de manière confuse						0,76
Ratés dans le discours						0,57
<b>Alpha de Cronbach</b>	<b>0.81</b>	<b>0.87</b>	<b>0.64</b>	<b>0.49</b>	<b>0.78</b>	<b>0.36</b>

Le premier facteur explique 15.40% de la variance totale des items. Les items saturés par le premier facteur sont relativement homogènes sur le plan sémantique. Ils renvoient à une attitude déterminée, à vouloir gagner en force, en dominant son interlocuteur. Ce premier facteur correspond clairement à une attitude de confrontation agressive et à une attitude **autoritaire** (*chercher à gagner en s'opposant à l'autre, agresser, se poser comme détenteur de la vérité*) qui évoque le type de leadership « autoritaire » décrit par Lewin (1975). L'alpha de Cronbach de ce facteur est de 0.81 ce qui est satisfaisant.

Les items saturés par le deuxième facteur font référence à une attitude d'ouverture et d'acceptation vis-à-vis de l'autre (*orienter son esprit vers la compréhension de l'autre, accorder de l'estime à ce qu'il dit*). Ce deuxième facteur qui explique 8.60% de la variance totale des items, est également très homogène sur le plan sémantique et proche de la dimension « **considération** » évoquée par Bass (1990). Ce facteur décrit dans quelle mesure un leader se montre concerné par le bien être d'autrui. Le leader considérant exprime sa reconnaissance du travail bien fait, met l'accent sur l'importance de la satisfaction au travail, maintient et renforce l'estime de soi de ses subordonnés en les traitant comme ses égaux. Nous pouvons d'ailleurs retrouver cette dimension dans les facteurs de deuxième ordre identifiés par Arthur et al (2003) sous le nom de considération/conscience des autres. Cette dimension décrit dans quelle mesure le stagiaire prend en compte les sentiments et les besoins de l'autre. En outre, ce facteur est très homogène avec un alpha de Cronbach de 0.87.

Le troisième facteur ne comporte que des items saturés négativement Il renvoie à une attitude correspondant à la volonté d'éviter le problème, de « noyer le poisson », attitude qui avait également été mise en évidence par les analyses de poste. L'intitulé du pôle positif de cette dimension devrait donc correspondre à une attitude **franche** et directe dans le cadre de la négociation. L'homogénéité de ce troisième facteur est moins satisfaisante avec un alpha de

Cronbach de 0.64 mais il s'agit là d'une dimension originale et il nous semble intéressant de la conserver.

Le quatrième facteur comporte également des items saturés négativement dont le contenu renvoie à un manque d'assurance, à une remise en question voire à une certaine anxiété. Le pôle positif de ce facteur correspondrait au contraire à une attitude **affirmée**, sereine et solide. Avec un alpha de Cronbach de 0.49, on ne peut pas considérer qu'il s'agit d'un facteur homogène et il faut en toute rigueur l'exclure de nos analyses (et ne pas s'en servir dans la mesure finale).

Le cinquième facteur correspond clairement à une expression orale en termes d'aisance, de clarté. Ce facteur est fréquemment retrouvé dans les CE sous la dénomination de « communication orale » (Arthur et al., 2003). Seul un item (*être attentif et concentré*) ne correspond pas précisément à la thématique de ce facteur. L'homogénéité de ce facteur est satisfaisante avec un alpha de Cronbach de 0.87.

Le sixième facteur fait état de difficultés dans le discours, de **confusion** et comprend seulement deux items qui apparaissent sémantiquement cohérents. Une vérification de la consistance interne de ce facteur indique cependant qu'il n'est pas homogène avec un alpha de Cronbach de 0.36. Nous avons donc décidé d'exclure ce facteur du reste de nos analyses.

Pour résumer, les quatre facteurs obtenus (attitude autoritaire, considération, franchise, communication orale) présentent une bonne homogénéité et sont clairement interprétables. Des facteurs dont le contenu est proche ont déjà été identifiés dans la littérature (Arthur et al., 2003), mais jamais en utilisant des analyses factorielles exploratoires. Notre épreuve de jeu de rôle permet d'évaluer ces quatre dimensions du comportement de façon satisfaisante et en respectant des étapes de la validation d'un outil. Par contre, 2 de ces facteurs ne sont pas homogènes et ne peuvent donc pas être considérées comme des dimensions interprétables.

#### 4. 2. Mise en relation avec des indicateurs de personnalité et d'intelligence : une approche nomologique

Afin de nous éclairer davantage sur la nature des construits identifiés dans notre première hypothèse, nous les avons mis en relation avec des indicateurs de personnalité et d'intelligence. Nous avons calculé les notes en facteurs pour les quatre dimensions identifiées précédemment afin de noter chaque participant à partir de la grille consensuelle (remplie conjointement pas les deux observateurs). Ces indicateurs n'ont pas été utilisés au hasard mais correspondent à des outils dont la validité de construit a été établie (Furnham, Forde, & Cotter, 1998; Goldberg, 1990).

Tableau 2 : Corrélations de Bravais-Pearson entre les dimensions issues du jeu de rôle et les échelles de personnalité et d'intelligence

	N	E	O	A	C	G
Attitude autoritaire	0,00	0,03	0,03	0,02	0,02	-0,05
Considération	-0,04	0,06	0,12	0,06	-0,07	0,09
Franchise	-0,11	0,01	-0,04	0,01	0,10	<b>-0,22*</b>
Communication Orale	<b>0,22*</b>	0,08	<b>0,19*</b>	-0,01	0,03	0,04

En gras, valeurs significatives à  $P < 0,05$

N : Névrosisme, E : Extraversion, O : Ouverture, A : Agréabilité, C : Conscience, G : facteur g.

En premier lieu, nous constatons sur le tableau 2 que seuls 3 coefficients de corrélation sur 24 calculés sont significatifs. Les dimensions « Attitude autoritaire » et « Considération » ne corrélerent avec aucun des indicateurs de personnalité ou d'efficacité mentale retenus.

La dimension « franchise » est corrélée négativement avec le score obtenu au facteur g de Cattell. Les individus adoptant une attitude franche dans une relation duelle obtiendraient donc des scores plus bas en intelligence fluide. Toutefois, ces résultats doivent être considérés avec prudence car la dimension « franchise » est moyennement homogène.

La dimension Communication Orale est corrélée positivement avec deux variables de personnalité (le Névrosisme et l'Ouverture). Ceci indique que les personnes obtenant des scores élevés en Névrosisme et en Ouverture sont mieux évaluées sur leur communication orale.

Si nous comparons nos résultats à ceux obtenus avec d'autres études, les conclusions sont contrastées. Lance, Newbolt, Gatewood, Foster, & French (2000) avancent que les « in basket » sollicitent plutôt des caractéristiques cognitives alors que les jeux de rôle sollicitent plutôt des conduites conatives. Nos résultats ne permettent pas de corroborer leurs hypothèses dans la mesure où les variables de personnalité sont peu corrélées avec les dimensions évaluées dans notre épreuve de jeu de rôle (seulement 2 corrélations significatives sur 24). Scholz & Schuler (1993) trouvent que certaines dimensions sont liées à certains traits de personnalité (la communication orale avec l'extraversion par exemple) alors que d'autres seraient liées à des aspects intellectuels (délégation et contrôle et GMA par exemple). Or, nous ne retrouvons pas ces résultats avec nos données puisque dans notre cas, la communication orale est corrélée avec le Névrosisme et l'Ouverture.

Au regard de ces résultats, on peut se demander si la démarche nomologique adoptée ici contribue vraiment à éclairer la nature des construits psychologiques évalués dans le cadre de notre jeu de rôle. En effet, nous ne

sommes pas parvenus à répliquer les quelques résultats décrits par la littérature selon lesquels le jeu de rôle mettrait plutôt en jeu des compétences intellectuelles. Au contraire, nos résultats vont dans le sens d'une relation très faible entre les compétences évaluées dans le cadre de notre épreuve, des indicateurs de personnalité et des indicateurs d'efficacité mentale. Sur un plan psychométrique, cela revient à dire que les dimensions que nous avons identifiées, la personnalité et l'efficacité mentale ne se recouvrent pas.

## **5. Discussion**

Durant ce travail, nous avons tenté de suivre deux objectifs. Le premier objectif était d'identifier de manière plus précise quels sont les construits psychologiques évalués dans le cadre d'une épreuve de jeu de rôle. Pour cela, nous avons d'abord recensé les comportements pertinents puis nous avons construit des grilles d'observation qui ont été renseignées par des observateurs formés aux biais de l'évaluation. Dans notre démarche, nous avons essayé de respecter toutes les étapes de validation de la construction d'un outil d'évaluation (Anastasi, 1990), sachant que ces précautions ne sont pas toujours prises lors de la mise en place d'une MS. Cela comprend les analyses de poste, la standardisation des épreuves, la formation des évaluateurs et l'utilisation de grilles comportementales. Nous sommes parvenus à isoler quatre dimensions interprétables et qui présentent des qualités psychométriques satisfaisantes. Ces dimensions (attitude autoritaire, considération, franchise, communication orale) ont été identifiées de manière exploratoire ce qui est nouveau. Notre travail trouve cependant des limites dans le fait que deux dimensions sur six (attitude affirmée et confusion) ne présentent pas les qualités psychométriques requises en termes d'homogénéité. De plus, nous avons mené des analyses factorielles sur l'ensemble des observations alors qu'il aurait fallu dans l'idéal mener des analyses séparées pour les évaluateurs psychologues et les évaluateurs de la Marine Nationale. On peut effectivement penser, avec Sagie & Magnezy (1997) que les évaluations se structurent différemment selon le type d'observateurs (les managers identifieraient moins de dimensions que des psychologues car plus soumis à l'effet de halo). Malheureusement, notre effectif ne nous permettait pas de mener des analyses factorielles séparément. Ces résultats sont malgré tout encourageants puisque la validité de construit est considérée comme le « talon d'Achille des CE » depuis le début des années 80. Nous avons mis en évidence que sous certaines conditions, la validité de construit d'une épreuve de jeu de rôle peut être améliorée. Nous pouvons donc affirmer que le premier objectif de notre travail est en grande partie atteint.

Notre deuxième objectif était de situer, d'éclairer la nature des construits psychologiques identifiés en les mettant en relation avec des indicateurs de la personnalité et de l'intelligence. Cette démarche est très

récente et correspond à un questionnaire plus fondamental sur la nature de ce qui est réellement évalué dans le cadre d'une MS. Il est notamment intéressant de se demander si les conduites mises en place par les évalués relèvent plus de la performance (c'est-à-dire de l'intelligence) ou de conduites habituelles (c'est-à-dire de la personnalité). On peut également se demander si l'évaluation porte sur des caractéristiques dispositionnelles (relatives à l'individu uniquement) ou sur des caractéristiques situationnelles (relatives aux situations uniquement). Selon les résultats que nous obtenons, les caractéristiques évaluées dans le cadre de notre jeu de rôle sont peu corrélées avec des éléments de la personnalité et de l'intelligence. Autrement dit, les dimensions identifiées dans le cadre de notre épreuve apportent une variance spécifique par rapport à des variables dispositionnelles telles que la personnalité ou l'intelligence. Ceci explique certainement la validité prédictive élevée des MS qui apporte des informations supplémentaires à celles obtenues avec les tests psychométriques. On parle dans ce cas de validité incrémentielle avec Sechrest (1963), cité par Hunsley & Meyer (2003). Selon nous, l'intérêt principal d'une MS réside dans la possibilité d'évaluer l'individu dans le cadre d'une interaction avec les exigences professionnelles alors que les tests psychométriques traditionnels ne prennent pas en compte le contexte.

Notre travail montre qu'il est possible d'évaluer des caractéristiques individuelles à la fois homogènes, spécifiques et contextualisées. Ces caractéristiques individuelles ne peuvent être évaluées que dans le cadre des MS, et ne sont appropriées que si ces dernières restituent les situations professionnelles de manière fidèle. Les dimensions que nous avons retrouvées ne sont pertinentes que dans le cadre des épreuves qui ont servi à les évaluer. On ne peut donc pas donner de réponse générale à la question « que mesure une épreuve de jeu de rôle ? ». Par contre, nous sommes capables de dire ce qu'évalue l'épreuve de jeu de rôle que nous avons mise en place.

Nous espérons que ces résultats permettront de développer l'utilisation de grilles comportementales qui permettent de mieux connaître ce qui est réellement sollicité par les mises en situation. Des efforts doivent être poursuivis en ce sens. Nous espérons que ce travail encouragera les organismes de recrutement à mettre en place des techniques de sélection plus proches des réalités professionnelles et moins aprioristes, en utilisant des outils d'une bonne validité (de contenu, de construit et de critère), afin de mieux servir les intérêts des services qui recrutent mais aussi les intérêts des candidats.

## Références

- Anastasi, A. (1990). *Psychological testing*. New York: MacMillan.
- Archambeau, D. J. (1979). Relationship among skill ratings assigned in an assessment center. *Journal of Assessment Center Technology*, 2, 7-20.
- Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion related validity of assessment center dimension. *Personnel Psychology*, 56, 125-154.
- Baron, H., & Janman, K. (1996). Fairness in the assessment center. *International Reviews of Industrial and Organizational Psychology*, 11, 60-112.
- Bass, B. M. (1990). *Bass and Stogdill's handbook of leadership*. New York: The free Press.
- Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decision: a conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, 74(3), 478-494.
- Callinan, M., & Robertson, I. T. (2000). Work sample testing. *International Journal of Selection and Assessment*, 8(4), 248-260.
- Cattell, R. B. (1974). *Manuel : Test de facteur "g" de R. B. Cattell*. Paris: ECPA.
- Costa, P. T., McCrae, R. R., & Rolland, J. P. (1998). *NEO PI-R. inventaire de personnalité révisé*. Paris: Les Editions du Centre de Psychologie Appliquée.
- Donahue, L. M., Truxillo, D. M., Cornwell, J. M., & Gerrity, M. J. (1997). Assessment center construct validity and behavioral checklists: Some additional findings. *Journal of Social Behavior and Personality*, 12(5), 85-108.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Furnham, A., Forde, L., & Cotter, T. (1998). Personality and intelligence. *Personality and Individual Differences*, 24(2), 187-192.
- Goldberg, L. R. (1990). An alternative description of personality : the Big Five factor structure. *Journal of Personality and Social Psychology*, 59(6), 1216-1229.
- Gorsuch, R. L. (1997). Exploratory factor analysis ; its role in item analysis. *Journal of Personality Assessment*, 68, 532-560.
- Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance*, 15(1/2), 25-46.
- Hennessy, J., Mabey, B., & Warr, P. (1998). Assessment Centre observation procedure: an experimental comparison of traditional, checklist and codind methods. *International Journal of Selection and Assessment*, 6(4), 222-230.
- Hunsley, J., & Meyer, G. J. (2003). The incremental validity of psychological testing and assessment : conceptual, methodological and statistical issues. *Psychological Assessment*, 15(4), 446-455.
- Huteau, M., & Lautrey, J. (1999). *Evaluer l'intelligence*. Vendome: PUF.
- Joyce, L. W., Thayer, P. W., & Pond, S. (1994). Managerial functions: an alternative to traditional assessment center dimensions. *Personnel Psychology*, 47, 109-121.
- Kendall, M. G., & Stuart, A. (1958). *The advanced theory of statistics, Volume 1 : distribution theory*. Londres: Ch. Griffin et Cie.

- Kleinmann, M., & Koller, O. (1997). Construct validity of assessment centers: Appropriate use of confirmatory factor analysis and suitable construction principles. *Journal of Social Behavior and Personality*, 12(5), 65-84.
- Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology*, 89(2), 377-385.
- Lance, C. E., Newbolt, W. H., Gatewood, R. D., Foster, M. S., & French, N. R. (2000). Assessment center exercises represent cross-situational specificity, not method bias. *Human Performance*, 13, 323-353.
- Lévy-Leboyer, C. (2004). *La personnalité. Un facteur essentiel de réussite dans le monde du travail*. Paris: Editions d'Organisation.
- Lewin, K. (1975). *Psychologie dynamique*. Paris: PUF.
- Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centers: An examination of assessor and assessee effects. *Journal of Applied Psychology*, 87(4), 675-686.
- Lievens, F., & Klimoski, R. J. (2001). Understanding the assessment centre process: where are we now. *International Reviews of Industrial and Organizational Psychology*, 16, 245-286.
- Reilly, R. R., Henry, S., & Smither, J. W. (1990). An examination of the effects of using behavior checklists on the construct validity of assessment center dimension. *Personnel Psychology*, 43, 71-84.
- Robertson, I. T., Gratton, L., & Sharpley, P. (1987). The psychometric properties and design of managerial assessment centres: Dimensions into exercises won't go. *Journal of Occupational Psychology*, 60, 187-195.
- Rolland, J. P. (2004). *L'évaluation de la personnalité: Le modèle en cinq facteurs*. Sprimont: Mardaga.
- Sackett, P. R. (1987). Assessment centers and content validity: some neglected issues. *Personnel Psychology*, 40, 13-25.
- Sackett, P. R., & Dreher, G. F. (1982). Construct and assessment center dimensions: some troubling empirical findings. *Journal of Applied Psychology*, 67(4), 401-410.
- Sagie, A., & Magnezy, R. (1997). Assessor type, number of distinguishable dimension categories, and assessment center construct validity. *Journal of Occupational and Organizational Psychology*, 70, 103-108.
- Salgado. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, 76, 323-346.
- Salgado, & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology*, 11(3), 299-324.
- Schmidt, F. L. (2002). The role of general cognitive ability and job performance: why there cannot be a debate. *Human Performance*, 15(1/2), 187-210.
- Scholz, C., & Schuler, H. (1993). Das nomologische netzwerk des Assessment Centre: Eine metaanalyse. *Zeitschrift für Arbeits und Organisationspsychologie*, 37, 73-85.

- Sechrest, L. (1963). Incremental validity : A recommendation. *Educational and Psychological Measurement*, 23, 153-158.
- Shore, T. H., Thornton, G. C., & Shore, L. M. (1990). Construct validity of two categories of assessment center dimension ratings. *Personnel Psychology*, 43(1), 101-116.
- Silverman, W. H., Dalessio, A., Woods, S. B., & Johnson, R. L. (1986). Influence of assessment center methods on assessors' ratings. *Personnel Psychology*, 39(3), 565-578.
- Squires, P., Torkel, S. J., Smither, J. W., & Ingate, M. R. (1991). Validity and generalizability of a role-play to select telemarketing representatives. *Journal of Occupational Psychology*, 64, 37-47.
- Thornton, G. C. I., & Zorich, S. (1980). Training to improve observer accuracy. *Journal of Applied Psychology*, 65, 351-354.
- Turnage, J. J., & Muchinsky, P. M. (1982). Transsituational variability in human performance within assessment centers. *Organizational Behavior and Human Performance*, 30, 174-200.
- Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, 52, 372-376.