

---

## Construction et validation d'une épreuve de groupe élaborée dans le cadre d'un « assessment center » pour des officiers de marine

*Construction and validity of a leaderless group discussion in an Assessment Center for naval officers*

**Xavier Borteyrou, Nicole Rasclé, Marilou Bruchon-Schweitzer et Pierre Collomb**

---



### Édition électronique

URL : <http://journals.openedition.org/osp/1190>  
DOI : 10.4000/osp.1190  
ISSN : 2104-3795

### Éditeur

Institut national d'étude du travail et d'orientation professionnelle (INETOP)

### Édition imprimée

Date de publication : 15 décembre 2006  
Pagination : 535-554  
ISSN : 0249-6739

### Référence électronique

Xavier Borteyrou, Nicole Rasclé, Marilou Bruchon-Schweitzer et Pierre Collomb, « Construction et validation d'une épreuve de groupe élaborée dans le cadre d'un « assessment center » pour des officiers de marine », *L'orientation scolaire et professionnelle* [En ligne], 35/4 | 2006, mis en ligne le 15 décembre 2009, consulté le 19 avril 2019. URL : <http://journals.openedition.org/osp/1190> ; DOI : 10.4000/osp.1190

---

Ce document a été généré automatiquement le 19 avril 2019.

© Tous droits réservés

---

# Construction et validation d'une épreuve de groupe élaborée dans le cadre d'un « assessment center » pour des officiers de marine

*Construction and validity of a leaderless group discussion in an Assessment Center for naval officers*

Xavier Borteyrou, Nicole Rascle, Marilou Bruchon-Schweitzer et Pierre Collomb

---

- 1 Les « Assessment Center » (A.C.) sont des dispositifs mis en place pour recruter des candidats ou, plus fréquemment, pour effectuer des bilans de carrière. Ils permettent d'évaluer des caractéristiques dispositionnelles comme l'intelligence et la personnalité, mais aussi des compétences professionnelles plus spécifiques<sup>1</sup>. Le principe essentiel d'un A.C. est de placer les évalués dans des situations professionnelles aussi proches que possible des situations professionnelles réelles. Les A.C. actuels comportent deux types d'épreuves : des tests d'intelligence et de personnalité, qui appréhendent des dispositions stables et générales (décontextualisées) ; des mises en situation (ou échantillons de travail), qui permettent d'évaluer des conduites spécifiques (contextualisées). Les A.C. comprennent généralement plusieurs types de mises en situation : l'épreuve *in basket*<sup>2</sup>, le jeu de rôle<sup>3</sup>, l'entretien structuré<sup>4</sup> et l'épreuve de groupe<sup>5</sup>. Dans le cadre de cet article, nous allons décrire la construction et la validation d'une épreuve de groupe.
- 2 Pour Bohem (1981), les A.C. peuvent être décrits par six caractéristiques : système mis en place pour évaluer le potentiel des cadres, recourant à des techniques variées, mettant l'accent sur l'évaluation du comportement en situation, confié à plusieurs spécialistes de l'évaluation, formés pour observer et évaluer le comportement de plusieurs candidats et dans lequel les phases d'observation et d'évaluation sont bien séparées.

## La validité des mises en situation : un constat paradoxal

- 3 D'une manière générale, un outil est considéré comme valide lorsqu'il mesure ce qu'il prétend mesurer. Cette définition est évidemment trop vague et générale, mais permet de comprendre l'objectif de ce qui ne constitue pas simplement une qualité des méthodes d'évaluation (leur validité), et correspond à l'ensemble d'une démarche, que l'on appelle aujourd'hui « processus » de validation (Dickès, Tournois, Flieller & Kop, 1994). Celui-ci ne se borne pas à construire un outil ; c'est le fait d'accumuler des preuves visant à en démontrer l'utilité et la pertinence (Anastasi, 1990). Un outil d'évaluation est considéré comme valide s'il satisfait trois types d'exigences : la validité de contenu, la validité de construit et la validité de critère, comprenant les validités prédictive et concurrente (Cronbach, 1984 ; Laveault & Grégoire, 1997).

### La validité de contenu

- 4 La validité de contenu d'une épreuve consiste à se demander dans quelle mesure les divers aspects de cette épreuve sont représentatifs du concept visé (Laveault & Grégoire, 1997). Byham (1980) cité par Sackett et Ryan (1985) suggère que la validité de contenu des mises en situations est recherchée à trois niveaux : les caractéristiques évaluées correspondent aux exigences du poste futur, les exercices sont représentatifs des postes à pourvoir, les caractéristiques jugées importantes sont évaluées lors de ces exercices. La validité de contenu est certainement le point fort des A.C. qui sont fondés sur le principe de la cohérence comportementale (Wernimont & Campbell, 1968). Ce principe repose sur le fait que le comportement suscité chez le candidat au cours d'une mise en situation est représentatif du comportement qu'il adoptera dans son travail futur dans des situations semblables. La validité de contenu correspondant à l'échantillonnage exhaustif des conduites et des situations, les mises en situations sont, par rapport à d'autres méthodes d'évaluation du personnel, des outils remplissant bien ces conditions.

### La validité de construit

- 5 La validité de construit consiste à se demander quels sont les construits psychologiques appréhendés par un outil. La littérature relative aux mises en situation fait apparaître que la validité de construit est le « talon d'Achille des A.C. » (Lievens, 1996). Ce problème a fait l'objet d'un très grand nombre de publications depuis une vingtaine d'années (Lance, Lambert, Gewin, Lievens & Conway, 2004 ; Lievens, 2001 ; Sackett, 1982 ; Sackett & Dreher, 1982). La validité de construit des mises en situation est mise à l'épreuve grâce à deux techniques : les matrices multi-traités multi-méthodes (M.T.M.M.) et les analyses factorielles.
- 6 La M.T.M.M. est une matrice de corrélations dans laquelle on met en relation des dimensions supposées similaires mais mesurées avec des techniques différentes (la résistance au stress mesurée par une épreuve de groupe, un jeu de rôle et un test *in basket* par exemple). Les études ayant utilisé cette technique aboutissent à des coefficients de validité convergente faibles (Chan, 1996 ; Fleenor, 1996 ; Robertson, Gratton & Sharpley, 1987), sauf pour la dimension « communication orale » (Kauffman, Jex, Love &

Libkuman, 1993)<sup>6</sup>. Autrement dit, les mises en situation ne permettent pas de mesurer des dimensions (leadership, résistance au stress par exemple) indépendamment des épreuves qui ont servi à les évaluer. La technique M.T.M.M. n'a donc pas démontré la validité de construit des mises en situation.

- 7 Pour éprouver la validité de construit des tests en situation, Sackett & Dreher (1982) proposent de conduire des analyses factorielles (A.F.) à partir des scores obtenus sur toutes les conduites des candidats dans tous les exercices, afin d'isoler des facteurs qui pourraient nous permettre d'identifier d'éventuels construits psychologiques sollicités par les mises en situation. Les conclusions de ces auteurs sont que les facteurs ainsi obtenus correspondent aux exercices plus qu'à des dimensions psychologiques communes. Ceci sera retrouvé dans des A.C. différents et avec des populations différentes (Kleinmann & Koller, 1997 ; Robertson, Gratton & Sharpley, 1987 ; Silverman, Dalessio, Woods & Johnson, 1986) : l'analyse des conduites évaluées dans les mises en situation (dans des exercices et par des évaluateurs différents) ne donne pas des construits psychologiques clairement identifiables (Klimoski & Brickner, 1987).

### La validité prédictive ou validité critérielle<sup>7</sup>

- 8 La validité prédictive des A.C. a longtemps été considérée de manière holistique. Il s'agissait alors de mettre en relation une évaluation globale issue d'un A.C. avec un critère de réussite professionnelle. Gaugler, Rosenthal, Thornton et Bentson (1987) ont ainsi obtenu un coefficient de validité prédictive moyen de .37. Cette démarche a été contestée (Arthur, Day, McNelly & Edens, 2003). En effet, les construits psychologiques évalués par les A.C. n'ayant pas été identifiés, un score global n'a guère de signification psychologique (il supposerait que tous les exercices qu'il comprend sollicitent une même compétence générale). En outre, il serait paradoxal de calculer un score total pour chaque candidat évalué par un A.C., si celui-ci s'avère d'une validité prédictive inférieure à celle des épreuves qu'il contient (Hoeft & Schuler, 2001). En effet, si le coefficient de validité prédictive des A.C. est en moyenne de .37, il est de .51 pour les entretiens structurés, de .51 pour les tests d'intelligence (aptitude mentale générale) et de .54 pour les échantillons de travail (Schmidt & Hunter, 1998). Pourquoi utiliser conjointement des outils, s'ils donnent, isolément, de meilleurs résultats ?
- 9 Le recours à des méthodes d'évaluation différentes dans les A.C. est considéré aujourd'hui comme souhaitable (Lievens, Harris, Van Keer & Bisqueret, 2003). Ceci à condition que les caractéristiques mesurées aient été bien définies, opérationnalisées puis validées (validité prédictive, notamment). Des dimensions comme « l'analyse des problèmes » et « la considération pour autrui » (évaluées dans des mises en situation) prédisent ensemble 20 % de la variance de la réussite professionnelle (Arthur *et al.*, 2003), alors qu'une évaluation globale ne permet de prédire que 14 % de ce critère (Gaugler *et al.*, 1987). Ces résultats montrent que la prise en compte des dimensions spécifiques aux mises en situation peut se révéler plus valide (en termes de validité prédictive) qu'une évaluation globale dans un A.C. En outre, ces dimensions ont une signification psychologique, ce qui n'est pas le cas pour des évaluations globales.
- 10 Pour conclure quant à la validité des mises en situation, on peut affirmer que celles-ci possèdent une bonne validité prédictive. En revanche, leur validité de construit n'a pas été démontrée, ce qui fait dire à certains auteurs : « ça marche, mais l'on ne sait pas ce que l'on mesure » (Russel & Domm, 1995).

## Les moyens pour améliorer la validité de construit des mises en situation

- 11 Ces dernières années, de nombreuses tentatives ont été conduites pour améliorer la validité de construit des mises en situation (Lievens, 1998). Ces efforts portent principalement sur les caractéristiques des évaluateurs et les caractéristiques des grilles utilisées.
- 12 L'expérience de l'évaluateur aurait une influence sur les processus perceptivo-cognitifs que sollicite une évaluation des candidats (Lievens, 2002). Ces processus concernent l'attention, la catégorisation, l'intégration et le rappel (Srull & Wyer, 1989). Lievens (2002) constate que les évaluations faites par des évaluateurs novices sont moins « discriminatives » que celles faites par des évaluateurs expérimentés. La mise en œuvre de processus cognitifs plus spécialisés et différenciés pour évaluer autrui pourrait s'améliorer avec la pratique. Aussi, des observateurs ayant suivi une formation (de type « *frame of reference training* ») feraient des évaluations plus différenciées des conduites des candidats que des novices (Schleicher, Day, Mayes & Riggio, 2002)<sup>8</sup>. Chez les observateurs ayant participé à une formation de ce type, les coefficients de validité convergente sont plus élevés.
- 13 Les caractéristiques des grilles comportementales utilisées par les évaluateurs peuvent également influencer la validité de construit d'un A.C. Reilly, Henry et Smither (1990) comparent deux types de grilles permettant d'évaluer 120 candidats sur des échelles de Likert en 5 degrés. L'une permet d'évaluer 8 dimensions de manière globale (par exemple la « résistance au stress » de chaque candidat). L'autre permet d'évaluer les mêmes 8 dimensions, chacune étant définie et illustrée par des comportements précis dont il faut évaluer l'occurrence. La validité convergente moyenne (mise en relation des dimensions identiques mesurées lors d'épreuves différentes) des grilles « globales » est de .24, celle des grilles « comportementales » de .43. Une telle procédure d'opérationnalisation permet donc d'améliorer la validité de construit des mises en situation. Dans cette optique, Lowry (1995) préconise l'utilisation de grilles comportementales spécifiques à chaque exercice. Ces grilles sont réalisées « sur-mesure » avec une liste de comportements précisément définis et issus des analyses de poste préalables.
- 14 Comparée aux autres mises en situation, l'épreuve de groupe est souvent considérée comme l'épreuve reine des A.C. (Baron & Janman, 1996). Si la littérature portant sur la validité des A.C. est extrêmement riche, nous n'avons recensé que très peu d'études ayant trait à la validité des épreuves de groupe, pourtant utilisées de manière systématique (Herriot, 2003). En France, nous n'avons recensé qu'une étude. Cette dernière porte sur la validité de construit des épreuves *in basket* (Rolland, 1999). À notre connaissance, seule l'étude de Chan (1996) a mis à l'épreuve à la fois la validité de construit et la validité prédictive des A.C. L'objectif de cet article est d'éprouver la validité de construit et la validité prédictive d'une épreuve de groupe. Quels sont les construits psychologiques évalués par une épreuve de groupe ? Cette méthode permet-elle de prédire la performance professionnelle ultérieure des candidats évalués ? Pour y répondre, nous avons tenu compte des propositions récentes visant à améliorer la validité de construit des mises en situation : utilisation de grilles comportementales construites sur mesure et formation des évaluateurs.

## Méthodologie

### Participants

- 15 Notre échantillon comprend 125 officiers de la Marine Nationale (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 (C.I.D.) de l'école militaire de Paris. Le C.I.D. est une formation de 1 an qui donne ensuite aux stagiaires la possibilité d'accéder à des postes de haut niveau. L'accès au C.I.D. 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. Les stagiaires occupent des niveaux hiérarchiques assez proches (la majorité sont « capitaines de frégate », ce qui correspond au 2<sup>e</sup> échelon du grade de Commandant). Ils ont tous un niveau scolaire élevé (formation initiale à l'École Navale). Les évaluateurs occupent des niveaux hiérarchiques supérieurs (capitaine de vaisseau ou amiral) et n'ont jamais travaillé avec les stagiaires évalués. Notre A.C. était présenté comme une étude expérimentale visant à améliorer les pratiques d'évaluation au sein de la Marine Nationale. Nous nous sommes engagés à ne pas communiquer les résultats de cette évaluation à la hiérarchie. Par contre, les résultats ont été restitués aux stagiaires de manière individuelle dans le cadre d'un entretien. Faute de place, nous ne présentons ici qu'une partie de notre recherche. Les stagiaires ont participé à une épreuve « in basket », un jeu de rôle et un entretien structuré. Nous avons également utilisé un test de personnalité en cinq facteurs, des tests d'intelligence fluide (aptitude mentale générale) et cristallisée (aptitudes verbales et numériques). Tous ces éléments ont été portés à la connaissance des stagiaires lors d'un entretien de restitution. Chacun d'eux a reçu un document de synthèse.

### Procédure

- 16 Notre étude s'est déroulée sur une période totale de 6 ans, d'août 1998 à novembre 2004 (tableau 1).

Tableau 1/Table 1

	EFFECTIF	1998	1999	2000	2001	2004
Promotion 1	30	T1				T2
Promotion 2	30		T1			T2
Promotion 3	33			T1		T2
Promotion 4	32				T1	T2

Les quatre groupes de sujets distingués selon le moment de l'évaluation  
*The four groups of subjects classified according to the time of the evaluation*

- 17 Les évaluations faites en T1 (épreuve de groupe) se répartissent sur une durée totale de 4 ans. Les évaluations faites en T2 (critère de réussite professionnelle) ont toutes eu lieu en 2004. L'intervalle entre T1 et T2 n'est donc pas identique pour toutes les promotions. Cet intervalle varie de 3 ans (promotion 2001) à 6 ans (promotion 1998). Le fait que le délai entre l'évaluation des prédicteurs et celle du critère ne soit pas le même pour tous les sujets est regrettable mais des contraintes organisationnelles nous ont obligé à

collecter tous les critères de réussite en même temps. La réussite professionnelle de ces officiers est une information confidentielle et difficile à obtenir (il a fallu attendre novembre 2004 pour les obtenir en totalité). Il est cependant appréciable d'avoir pu réaliser une étude longitudinale pour valider notre A.C. (intervalle de 3 à 6 ans entre T1 et T2) car les études longitudinales effectuées dans ce domaine sont très rares (McEvoy & Beatty, 1989).

## Construction des grilles d'évaluation

- 18 Les analyses de poste ont été menées en 1997 selon la méthode des incidents critiques (Flanagan, 1954). Nous avons analysé 24 « postes clés » en suivant cette méthode. Tous les postes analysés correspondent aux affectations que les candidats occuperont à la sortie du C.I.D. Les stagiaires que nous avons rencontrés sont par la suite affectés dans des postes à responsabilité de type « commandant de navire », « responsable du recrutement des officiers », « commandant d'une école de formation » ou « chef du service de communication d'une préfecture maritime » par exemple. Les analyses de poste ont permis d'identifier des catégories de comportements caractérisant les futures affectations (communication orale, analyse des problèmes ou des dossiers, adaptabilité, décision, leadership ou commandement, persuasion, tolérance à la pression extérieure, jugement, organisation et planification, capacités de travail, travail en équipe, adhésion à l'institution, résistance au stress). Bien entendu, ces catégories n'ont pas été évaluées directement, mais opérationnalisées sous forme d'unités de comportement plus spécifiques définies de façon précise et quantifiable. La communication orale, par exemple, comprend les comportements suivants : parler de manière concise et claire, se perdre dans les détails, se baser sur des faits pour argumenter, présenter des ratés dans le discours. Ce sont ces comportements qui seront évalués lors de l'épreuve de groupe. Les évaluateurs (2 par candidat) remplissent d'abord une grille d'observation individuelle, puis une grille commune appelée « grille consensuelle » qui sera utilisée pour noter les stagiaires.

## Formation des évaluateurs

- 19 Chaque évaluateur a été formé aux pratiques de l'évaluation par nos propres soins pendant deux demi-journées. Nous nous sommes inspirés de plusieurs méthodes décrites par Woehr et Huffcut (1994). Dans un premier temps, nous avons sensibilisé les évaluateurs aux distorsions pouvant affecter l'hétéro-évaluation (Laberon, 2001). Nous leur avons décrit l'effet de halo, les corrélations illusives, l'économie cognitive, la confirmation d'hypothèse, l'effet de similitude ainsi que celui de première impression, tout en expliquant de quelle manière ces processus peuvent biaiser l'évaluation d'autrui.
- 20 Enfin, nous avons « entraîné » les observateurs à effectuer des évaluations sur des cas réels à l'aide d'un support vidéo<sup>9</sup>. Il s'agissait d'une formation participative à l'issue de laquelle les différents observateurs ont été invités à confronter leurs évaluations. Cette formation leur a aussi permis de se familiariser avec les épreuves et les grilles utilisées. Les évaluateurs qui ont participé à notre recherche étaient soit capitaine de vaisseau, soit amiral. Nous disposions chaque année d'un vivier de 15 évaluateurs que nous invitions en fonction de leur disponibilité et de la connaissance qu'ils avaient des stagiaires (chaque stagiaire était évalué par un évaluateur qui ne le connaissait pas).

## La construction de l'épreuve de groupe

- 21 Les épreuves de groupe sont de divers types, la plus prédictive de la réussite professionnelle semblant être l'épreuve de groupe avec des rôles assignés (Thornton & Byham, 1982). Elle consiste à donner des rôles préétablis à chacun des protagonistes. Nous avons donc mis en place une épreuve réunissant trois participants pour régler par exemple un problème de « déficit de ligne budgétaire des appelés ». Chaque protagoniste appartient à un bureau différent : l'un est chargé de la politique d'emploi du personnel (il doit gérer les affectations), un autre est chargé du budget (il doit respecter une enveloppe budgétaire) et le troisième est chargé du recrutement du personnel (il doit donc rendre les postes attractifs). Le but de la réunion est d'aboutir à une solution quant à la répartition des lignes budgétaires qui puisse convenir à l'ensemble des participants, sachant que les objectifs de chacun sont différents au départ. Chaque stagiaire dispose d'argumentaires et d'objectifs précis. Le scénario a été conçu de manière à ce que les arguments soient équilibrés et que la connaissance des dossiers (qui peut varier d'un individu à l'autre) ne favorise pas l'un des protagonistes. En effet, l'objectif de cette épreuve n'est pas d'évaluer des connaissances professionnelles mais d'appréhender le comportement d'un individu en groupe et en particulier la façon dont il défend son point de vue et prend en compte celui d'autrui.
- 22 Les stagiaires disposent de 45 minutes pour prendre connaissance des dossiers. Chaque stagiaire possède un objectif chiffré spécifique à son bureau et auquel il doit parvenir à l'issue de la négociation. On lui indique également les seuils financiers correspondant à un « accord parfait », un « accord avec concessions », « accord inacceptable ». Les dossiers sont accompagnés d'une lettre de leur supérieur hiérarchique leur demandant d'obtenir le meilleur résultat possible (accord parfait). Le scénario, ainsi que les objectifs, sont tirés d'une situation réelle. La négociation dure 1 heure. Chaque candidat est observé par deux évaluateurs et la séquence est filmée. Un système de projection vidéo permet aux évaluateurs de visionner certaines séquences de l'épreuve, autant de fois que nécessaire.

## Le critère de réussite professionnelle

- 23 La première possibilité était d'utiliser le système d'évaluation annuel effectué par le supérieur hiérarchique ( $n + 1$ ). Chaque officier est effectivement évalué localement par son supérieur hiérarchique dans le cadre de la notation annuelle. De nombreuses recherches ont montré que la notation faite par le  $n + 1$  reste la plus pertinente pour évaluer la performance professionnelle (Borman, White, Pulakos & Opler, 1991 ; Harris & Schaubroeck, 1988 ; Johnson, 2001). Pourtant, ces évaluations se sont avérées peu discriminatives pour notre population (la variance des notes obtenues étant très faible). Comme nous l'avons souligné, notre échantillon est composé d'officiers faisant partie de l'élite de la Marine Nationale et qui sont tous très bien notés par leur supérieur. Les différences inter-individuelles observées au niveau de cette variable se sont avérées très faibles, ce qui nous a conduit à ne pas prendre en compte ce critère de performance professionnelle.
- 24 Nous avons donc décidé de faire appel à la commission d'évaluation mise en place par la Marine Nationale (novembre 2004). Cette commission est chargée de centraliser toutes les feuilles de notation rédigées par les  $n + 1$ . Chaque officier est donc évalué par le même

groupe d'experts au lieu d'être évalué chacun par un supérieur différent. Le problème principal ayant trait à la variance des notations, nous avons demandé à cette commission d'évaluer chaque sujet de l'échantillon sur une échelle en 3 degrés : réussite normale, réussite supérieure à la moyenne et réussite inférieure à la moyenne. Pour décider, la commission disposait d'un ensemble d'éléments qualitatifs (appréciations du  $n + 1$  qui figurent sur le formulaire de notation annuelle) et quantitatifs (notation, rapidité d'avancement de carrière) mais prenait aussi en compte les contraintes spécifiques des postes occupés (certains étant plus difficiles que d'autres, ceci variant selon l'actualité). Seule une commission d'experts a le recul nécessaire pour pouvoir classer ces officiers entre eux et peut utiliser des critères d'appréciation communs.

## Résultats

### La validité de construit de l'épreuve de groupe

- 25 Pour éprouver la validité de construit de l'épreuve de groupe, nous avons mené des analyses factorielles exploratoires à partir des données notées par les observateurs dans les grilles comportementales. Nous avons préféré les analyses factorielles au M.T.M.M., car notre objectif était d'explorer de quelle manière les évaluations se structurent dans cette épreuve. Pour chaque comportement, les évaluateurs étaient invités à 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 grilles comportementales utilisées pour l'épreuve de groupe comportaient initialement 84 items. En ajoutant les grilles individuelles et les grilles consensuelles, nous obtenions 292 observations.
- 26 Dans un premier temps, nous avons supprimé 13 items qui ne présentaient pas des caractéristiques psychométriques satisfaisantes en termes de distribution. Nous avons éliminé les items dont le *skewness* (indice d'asymétrie) était trop élevé (à partir de + 3 ou de - 4) et dont le *kurtosis* (indice d'aplatissement) était supérieur à 4 (Gorsuch, 1997). Cette étape nous a conduit à éliminer 13 items (exemple : fuir le regard de l'interlocuteur, rester passif, être distrait).
- 27 Nous avons ensuite conduit des A.F. (méthode du maximum de vraisemblance), suivie de rotations promax. Cette analyse a donné 12 facteurs de valeur propre supérieure à 1 (critère de Kaiser). Les facteurs retenus selon ce critère étant trop nombreux et pour certains, non interprétables, nous avons décidé d'utiliser des critères d'extraction plus exigeants, le scree-test et le critère d'interprétabilité<sup>10</sup>. Nous avons retenu ainsi une solution à 6 facteurs, qui est présentée au *tableau 2* qui explique 41,40 % de la variance totale. Les items dont les saturations sont inférieures à .30, ainsi que les items à double saturation (figurant en italiques en bas du *tableau 2*) ont été éliminés. La solution factorielle présentée ici est obtenue après élimination des 22 items défaillants (13 qui ne se distribuent pas de manière normale et 9 qui présentent des saturations trop basses). Cette solution présentée au *tableau 2* est obtenue après rotation, à partir de 292 observations et 62 variables, ce qui constitue un ratio nombre d'observations/ nombre de variables de 4,7. Ce ratio est acceptable mais limite, sachant qu'il doit être supérieur à 5 pour être satisfaisant (Kline, 1993).
- 28 *Tableau 2/*Table 2

	F 1	F 2	F 3	F 4	F 5	F 6
<b>Pourcentage de variance expliquée</b>	<b>17.85</b>	<b>7.38</b>	<b>5.30</b>	<b>4.32</b>	<b>3.82</b>	<b>2.74</b>
<b>Alpha de Cronbach</b>	<b>.90</b>	<b>.13</b>	<b>.86</b>	<b>.90</b>	<b>.75</b>	<b>.53</b>
46 Proposer une solution adéquate (efficace)	0,74					
44 Prendre des initiatives, des mesures à son niveau	0,72					
50 Décider au moment opportun (ni trop tôt, ni trop tard)	0,65					
48 Proposer différentes alternatives	0,64					
47 Proposer une solution réalisable	0,63					
3 Prendre en compte la totalité des objectifs	0,61					
53 Proposer une solution en la formulant de façon précise	0,58					
41 Prendre ses responsabilités	0,56					
27 Aller à l'essentiel	0,55					
16 Utiliser l'humour	0,50					
45 Savoir enfreindre les règles à bon escient	0,50					
54 Traiter le problème, répondre à la question posée	0,49					
8 Effectuer une synthèse claire et concise	0,46					
66 Céder sur certains points pour pouvoir avancer sur d'autres	0,43					
14 Faire apparaître les points positifs d'un dossier	0,42					
42 Prendre une solution innovante	0,38					
2 Faire des choix dans le traitement de l'information	0,38					
40 Émettre un choix dans les propositions	0,37					
43 Demander un arbitrage supérieur	- 0,39					
56 Laisser les autres décider à sa place	- 0,51					
51 Laisser une situation se détériorer ; laisser traîner	- 0,67					

30 Tenir compte des réactions des interlocuteurs		0,69				
37 Présenter un discours structuré		0,61				
31 Être attentif et concentré		0,50				
38 Accepter et intégrer d'autres idées que les siennes		0,48				
34 S'exprimer avec aisance		0,35				
33 S'exprimer de manière laconique		-0,32				
29 Ne se baser que sur des opinions, des <i>a priori</i>		-0,57				
36 Ratés dans le discours		-0,60				
12 Convaincre les interlocuteurs que l'on a raison			0,70			
71 Montrer que l'on a raison			0,67			
69 Montrer que l'autre a tort			0,59			
9 Démontre l'argumentaire inverse			0,58			
13 Prouver à l'interlocuteur que sa méthode n'est pas bonne			0,57			
7 Construire un argumentaire pour soutenir ses choix			0,55			
68 Chercher à gagner en s'opposant à l'autre			0,53			
11 Présenter un argumentaire détaillé			0,52			
15 Utiliser et démonter l'antithèse pour appuyer sa thèse			0,49			
70 Chercher à aboutir en dominant			0,47			
74 Chercher à aboutir en affaiblissant			0,44			
35 S'exprimer en développant son discours			0,41			
4 Avoir en mémoire les informations pertinentes			0,41			
26 Parler de manière concise et claire			0,38			
5 Reformuler la problématique globale			0,34			
72 Chercher à aboutir en manipulant				-0,98		

73 Faire preuve de ruse pour aboutir				-0,95		
21 Organiser un travail en équipe					0,71	
59 Répartir les tâches					0,56	
17 Donner les orientations à ses interlocuteurs					0,47	
23 Traduire les directives en ordres précis					0,46	
20 Recadrer les actions ou comportements					0,42	
24 Rechercher l'adhésion de son équipe					0,36	
19 Effectuer un remue-ménages					0,33	
61 Réalise un dosage équitable du temps de parole					0,33	
58 Maintenir ses objectifs						0,43
39 Être pris au dépourvu par les questions						-0,35
65 Recherche l'accord						-0,42
49 Manquer de précision dans ses propositions						-0,42
57 Aller dans le sens d'une décision jugée mauvaise						-0,42
18 Laisser ses interlocuteurs s'exprimer						-0,49
60 Laisser les autres s'exprimer						-0,53
25 Prendre en compte un avis contraire						-0,58
28 Se baser sur des faits pour argumenter	0,25					
1 Demande de clarification		0,26				
63 Chercher à expliquer plutôt qu'à influencer		0,26				
32 Présenter des signes d'anxiété		-0,18				
22 Décider d'utiliser un argument d'autorité		-0,28				
6 Faire une enquête			0,23			
67 Engager une discussion constructive					0,28	
62 Orienter son esprit dans la dynamique de la compréhension de l'autre						-0,26
64 S'engager à douter, voire à accepter changer d'opinion						-0,28

Résultats de l'analyse factorielle (méthode en maximum de vraisemblance suivie de rotations promax) menée sur les items de la grille comportementale. Les items éliminés sont en italiques. Les résultats présentés sont obtenus après rotation  
*Results of the factorial analysis (maximum likelihood followed by rotation promax) of the items of the behavioral checklist. Items dropped are written in italic. Results presented here are obtained after rotation*

- 29 Le 1<sup>er</sup> facteur comprend 21 items et explique 17,85 % de la variance. Les saturations positives correspondent clairement au fait de prendre des initiatives efficaces (*proposer des solutions adéquates, prendre des initiatives, décider au moment opportun*) alors que les saturations négatives correspondent à un manque d'initiative (*laisser les autres décider à sa place, laisser une situation se détériorer*). Les items qui saturent ce facteur décrivent tous la prise (ou l'absence) d'initiatives sauf l'item : « *va à l'essentiel* », qui reste cependant cohérent avec la signification de ce facteur. Le contenu de ce facteur est proche de la dimension « structuration » du leadership décrite par Bass (1990)<sup>11</sup>. La structuration correspond au fait qu'un leader initie et organise l'activité d'un groupe et définit la façon dont le travail doit être fait. Une structuration faible est le fait d'hésiter à prendre des initiatives dans le groupe et de ne pas prendre des mesures adéquates. Or les items constituant ce facteur correspondent exactement à ces caractéristiques. En outre, il possède un alpha de Cronbach élevé (.90). Nous avons donc conservé ce facteur et l'avons dénommé « prise d'initiatives efficaces ».
- 30 Le deuxième facteur est saturé par 8 items et explique 7,38 % de la variance. Il est difficile à identifier car il renvoie à la fois à la sensibilité interpersonnelle (*tenir compte des réactions de son interlocuteur, être attentif et concentré*) et à la communication orale (*s'exprimer avec aisance, présenter des ratés dans le discours*). Ce facteur étant difficilement interprétable et hétérogène (alpha de Cronbach égal à .13), nous ne l'avons pas conservé.
- 31 Le troisième facteur explique 5,30 % de la variance et sature 15 items très cohérents sur le plan sémantique. Ces items décrivent un mode de persuasion autoritaire (*convaincre les interlocuteurs que l'on a raison, montrer que l'on a raison, démonter l'argumentaire inverse*). Ce facteur est très homogène (alpha de Cronbach de .86). Nous l'avons conservé et l'avons dénommé : « autoritarisme ».
- 32 Le quatrième facteur n'est saturé que par deux items (*cherche à aboutir en manipulant, fait preuve de ruse pour aboutir*). Il explique 4,32 % de la variance. Les coefficients de saturation sont négatifs et élevés. Le pôle positif de ce facteur correspond à l'absence de stratagème dans les communications en groupe (naïveté), le pôle négatif à de la manipulation (ruse). Ce résultat est étonnant et doit être considéré avec prudence, les deux items étant extrêmement corrélés (doublons ?). Nous savons que des inter-corrélations très fortes peuvent aboutir à des solutions factorielles artificielles (Nunnally & Bernstein, 1994). Or ce risque est particulièrement important si les variables comportent peu de degrés comme c'est le cas ici. Nous avons donc supprimé ce facteur pour la suite de nos analyses.
- 33 Le cinquième facteur comprend 8 items et explique 3,82 % de la variance. Son contenu évoque clairement des comportements de travail en équipe et correspond à un leadership à la fois démocratique (*organiser un travail en équipe, recherche l'adhésion de son équipe*) et structurant (*recadre les actions ou les comportements, répartir les tâches, donner des orientations à son interlocuteur*). Ce facteur a un alpha de .75 ce qui est satisfaisant compte tenu de son petit nombre d'items (8). Nous avons dénommé ce facteur « travail en équipe ».
- 34 Enfin, le sixième facteur est saturé par 7 items et explique 2,74 % de la variance. Il correspond à une attitude opposante, consistant à ne pas prendre en compte l'avis de ses interlocuteurs (*ne prend pas en compte un avis contraire, ne laisse pas les autres s'exprimer, ne*

*recherche pas l'accord*). Les saturations des items sur ce facteur sont faibles. La signification psychologique de ce facteur semblait assez claire, mais son alpha de Cronbach étant faible (.53), nous avons donc décidé de le supprimer pour la suite de l'étude.

- 35 Pour résumer, les trois facteurs obtenus (prise d'initiatives efficaces, autoritarisme et travail en équipe) présentent une bonne homogénéité et sont clairement interprétables. Des facteurs dont le contenu est proche ont déjà été identifiés dans la littérature (Arthur, Day, McNelly & Edens, 2003), mais jamais en utilisant des analyses factorielles exploratoires. Notre épreuve de groupe permet d'évaluer ces trois dimensions du comportement de façon satisfaisante même si la variance expliquée par ces trois facteurs reste modeste avec 26,97 % de la variance totale des items.

## La validité prédictive de l'épreuve de groupe

- 36 Pour établir le poids respectif des compétences évaluées dans le cadre de l'épreuve de groupe dans la prédiction de la réussite professionnelle future, nous avons conduit une régression linéaire multiple entre les trois dimensions de l'épreuve de groupe identifiées ci-dessus, considérées à présent comme des prédicteurs, et le critère de réussite professionnelle. Les candidats ont reçu une note en facteur sur ces trois dimensions à partir des scores qu'ils ont obtenus à la grille comportementale consensuelle (remplie après confrontation des observations notées par deux évaluateurs différents).

Tableau 3/Table 3

		BÊTA STANDARDISÉ	T	p	R <sup>2</sup>
	(constante)	0,00	- 4,41	0,00	
F 1	Prise d'initiatives efficaces	0,43	5,89	0,00	.31
F 3	Travail en équipe	0,34	5,38	0,00	.12
F 2	Autoritarisme	0,01	0,17	0,43	.00

R<sup>2</sup> ajusté pour l'ensemble des variables : 0,43.

Résultats des analyses de régression multiple calculées entre les dimensions de l'épreuve de groupe et le critère de réussite professionnelle

*Results of the regression analyses carried out between group exercise dimensions and the criteria of work performance*

- 37 La dimension « prise d'initiatives efficaces » (bêta standardisé de 0,39) explique à elle seule 31 % de la variation du critère. Ce résultat est d'autant plus intéressant que cette dimension est très homogène (alpha de Cronbach = .90)<sup>12</sup>. Les items qui composent cette dimension renvoient pour la plupart à l'*efficacité* des initiatives prises (proposer une solution *adéquate*, décider au moment *opportun*, trouver une solution *innovante*). Certains auteurs ont déjà montré que la prise d'initiatives (comportements dits « proactifs ») prédisent la performance professionnelle à long terme (Seibert, Crant & Kraimer, 1999).
- 38 La dimension « travail en équipe » (bêta standardisé de 0,34) explique 12 % de la variation du critère de réussite professionnelle. Ce facteur correspond au premier des trois déterminants de l'avancement de carrière, dénommé « efficacité interpersonnelle » par Hogan, Curphy et Hogan (1994). Selon ces auteurs, il comprend la capacité à travailler en réseau, les relations humaines et la communication orale<sup>13</sup>.
- 39 La dimension « autoritarisme » ne participe pas de manière significative à la prédiction du critère (bêta standardisé de 0,01, non significatif).

- 40 Enfin, le coefficient de régression multiple ajusté, calculé sur nos trois prédicteurs est satisfaisant ( $R^2$  ajusté = .43)<sup>14</sup>. Ainsi deux des dimensions évaluées par l'épreuve de groupe de notre A.C. permettent à elles seules de prédire 43 % de la variation du critère de réussite professionnelle.

## Discussion

- 41 Cette étude s'était fixé deux objectifs : établir la validité de construit de l'épreuve de groupe réalisée en T1 (hypothèse 1) et démontrer sa capacité à prédire la réussite professionnelle ultérieure (en T2) des officiers de Marine ainsi évalués (hypothèse 2).
- 42 Nous avons tout d'abord identifié par des analyses factorielles trois dimensions des comportements sollicités par l'épreuve de groupe : prise d'initiatives, travail d'équipe, autoritarisme. Il est intéressant de noter que ces trois dimensions ont déjà été identifiées dans la littérature sur le leadership. Le premier facteur a été décrit par Bass (1998) qui distingue la considération (décrivant dans quelle mesure un leader se montre concerné par le bien-être des autres membres du groupe) de la structuration (indiquant dans quelle mesure un leader organise l'activité d'un groupe). Le deuxième facteur a été identifié par Hogan, Curphy et Hogan (1994) qui parlent d'efficacité interpersonnelle plutôt que de travail en équipe. Enfin, la troisième dimension rappelle fortement des travaux plus anciens sur le leadership décrit dans la littérature par Lewin *et al.* (1939) cités par Mahfooz (1990) qui renvoie surtout à des techniques de persuasion autoritaires ou de commandement.
- 43 Dans notre recherche, ces dimensions n'ont pas été évaluées de manière globale, mais par des grilles comprenant des petites unités de comportements spécifiques et précises. Cette précision ainsi que toutes les précautions méthodologiques prises ont probablement facilité la découverte de « construits psychologiques » clairs et dont les qualités psychométriques (consistance interne, validité de construit, validité prédictive) sont satisfaisantes. Il est par ailleurs intéressant de noter que les dimensions mesurées par notre épreuve de groupe (prise d'initiatives, travail d'équipe, autoritarisme) sont des caractéristiques individuelles difficiles (voire impossibles) à évaluer par des tests psychométriques classiques<sup>15</sup>.
- 44 Nos résultats montrent que la validité de construit des mises en situation peut être démontrée à condition d'utiliser des grilles comportementales permettant de définir opérationnellement les dimensions évaluées, de former les évaluateurs et de standardiser les épreuves (et les rôles assignés à chacun des protagonistes)<sup>16</sup>. La validité de construit de cette épreuve de groupe est encore perfectible. En effet, certains des facteurs obtenus ne sont pas interprétables et/ou ne sont pas homogènes et la variance expliquée par ces trois facteurs est modeste. Il serait intéressant de mener des analyses factorielles séparément chez les évaluateurs non professionnels et les évaluateurs professionnels, afin de vérifier si leurs évaluations se structurent différemment, l'évaluation devant être plus différenciée dans le second cas (Sagie & Magnezy, 1997)<sup>17</sup>. Ces résultats sont malgré tout encourageants puisque la validité de construit est considérée comme le « talon d'Achille des C.E. » depuis le début des années 80. Nous avons mis en évidence que sous certaines conditions, la validité de construit des A.C. peut être améliorée sans toutefois être comparable à celle des tests de personnalité ou d'intelligence.

- 45 En ce qui concerne la validité prédictive des dimensions de notre épreuve de groupe, les pourcentages de variance expliquée sont importants, ce qui est en accord avec les quelques résultats disponibles concernant la validité prédictive de l'épreuve de groupe, ces pourcentages variant de 34 % (Bobrow & Leonards, 1997 ; Lievens, Harris, Van Keer & Bisqueret, 2003 ; Robertson & Kandola, 1982) à 39 % (Lievens, Harris, Van Keer & Bisqueret, 2003).
- 46 La validité prédictive de notre épreuve de groupe est donc relativement importante (elle permet de prédire 43 % de la réussite professionnelle ultérieure), ce qui pourrait s'expliquer de quatre façons.
- 47 Premièrement, les mises en situation sont calquées sur les situations effectives de travail, ce qui est impossible à proposer dans des tests psychométriques traditionnels, qui doivent rester génériques. Ces mises en situation permettent d'évaluer directement des compétences qui ne peuvent pas être appréhendées avec des outils psychométriques classiques, à la fois décontextualisés et trop généraux (plutôt que spécifiques).
- 48 Deuxièmement, les mises en situation permettent de « capturer » une interaction entre les caractéristiques des candidats (personnalité, intelligence) et les caractéristiques du contexte (choisies comme représentatives des situations professionnelles futures). Ceci explique en grande partie la validité prédictive élevée des mises en situation.
- 49 Troisièmement, il a déjà été démontré que la validité prédictive des A.C. est plus élevée lorsque ces derniers sont utilisés dans le cadre de l'avancement de carrière, plutôt que dans le cadre de recrutement (Gaugler, Rosenthal, Thornton & Bentson, 1987).
- 50 Enfin, il semblerait que la validité prédictive des A.C. augmente avec le temps (McEvoy & Beatty, 1989). Dans notre étude, le laps de temps médian entre le recueil des prédicteurs et celui du critère est de trois ans et demi ce qui est relativement élevé par rapport aux recherches menées dans ce domaine. Ceci constitue une raison supplémentaire pour expliquer un coefficient de validité prédictive si élevé.
- 51 Si les épreuves de groupe sont des tests « sur mesure » (et non du « prêt-à-porter comme les tests classiques), les compétences qu'elles permettent d'évaluer ne sont pas généralisables, les conduites évaluées ne correspondant qu'à la gamme de situations proposées. Une telle spécificité explique l'absence d'outils génériques dans ce domaine, outils qui existent pour évaluer la personnalité ou l'intelligence. Elle nuit à la généralisation des résultats et au progrès des connaissances scientifiques dans le champ des A.C. Aujourd'hui, les chercheurs travaillant dans ce domaine semblent faire le deuil de la possibilité d'élaborer des mises en situation relativement généralisables (Lowry, 1995).
- 52 Cette étude montre que l'épreuve de groupe que nous avons mise en place possède une bonne validité de construit (ce qui est nouveau) et une excellente validité prédictive. Nous espérons que ces résultats encourageants inciteront d'autres chercheurs à construire, valider et utiliser des grilles comportementales du même type (à la fois structurées et « situationnelles »). De telles mises en situation permettent des évaluations non seulement très valides mais en outre non explorées par les tests classiques. Elles complètent la gamme habituelle des outils utilisés dans le recrutement en apportant des évaluations différentes, complémentaires et sensibles ce qui est particulièrement appréciable pour des individus ayant de hautes qualifications professionnelles. Nous souhaitons que ce travail puisse encourager les organismes responsables de l'évaluation et l'orientation du personnel à mettre en place des procédures de recrutement plus

proches des réalités professionnelles, en utilisant des outils d'une validité bien établie, afin de mieux défendre non seulement les intérêts des recruteurs mais aussi ceux des candidats.

---

## BIBLIOGRAPHIE

- Anastasi, A. (1990). *Psychological testing*. New York : MacMillan.
- 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.
- Bass, B. M. (1998). *Transformational Leadership : Industrial, Military, and Educational Impact*. Mahwah, NJ : Lawrence Erlbaum Associates.
- Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. *Academy of Management Review*, 6, 205-212.
- Bobrow, W., & Leonards, J. S. (1997). Development and Validation of an Assessment Center during organizational change. *Journal of Social Behavior and Personality*, 12, 217-236.
- Bohem, V. R. (1981). *Establishing the validity of an Assessment Center*. Pittsburg : Development Dimension Press, Monograph Series.
- Borman, W., White, L., Pulakos, E., & Opler, S. (1991). Models of supervisor job performance ratings. *Journal of Applied Psychology*, 75, 863-872.
- Byham, W. C. (1980). Starting an assessment center the right way. *Personnel Administrator*, February 1980, 27-32.
- Chan, D. (1996). Criterion construct validation of an assessment centre. *Journal of Occupational and Organizational Psychology*, 69, 167-181.
- Cronbach, L. J. (1984). *Essentials of Psychological Testing (4th)*. New York : Harper & Row.
- Dickès, P., Tournois, J., Flieller, A., & Kop, J.-L. (1994). *La psychométrie : théories et méthodes de la mesure en psychologie*. Paris : P.U.F.
- Doron, R., & Parot, F. (1991). *Dictionnaire de la psychologie*. Paris : P.U.F.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Fleenor, J. W. (1996). Constructs and developmental assessment centres : further troubling empirical findings. *Journal of Business and Psychology*, 10, 319-333.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493-511.
- Gorsuch, R. L. (1997). Exploratory factor analysis ; its role in item analysis. *Journal of Personality Assessment*, 68, 532-560.

- Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43-62.
- Herriot, P. (2003). Assessment by groups : Can value be added ? *European Journal of Work and Organizational Psychology, 12*, 131-145.
- Hoefl, S., & Schuler, H. (2001). The conceptual basis of assessment centre ratings. *International Journal of Selection and Assessment, 9*, 114-123.
- Hogan, R., Curphy, G. J., & Hogan, J. (1994). What we know about leadership. Effectiveness and personality. *American Psychologist, 49*, 493-504.
- Huffcutt, A. I., & Woerh, D. J. (1999). Further analysis of employment interview validity ; a quantitative evaluation of interviewer-related structuring methods. *Journal of Organizational Behavior, 20*, 549-560.
- Johnson, J. W. (2001). The relative importance of task and contextual performance dimensions to supervisor judgment of overall performance. *Journal of Applied Psychology, 86*, 984-996.
- Kauffman, J. R., Jex, S. M., Love, K. G., & Libkuman, T. M. (1993). The construct validity of assessment center performance dimension. *International Journal of Selection and Assessment, 1*, 213-223.
- 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*, 65-84.
- Klimoski, R., & Brickner, M. (1987). Why do assessment centers work ? The puzzle of assessment center validity. *Personnel Psychology, 42*, 243-260.
- Kline, P. (1993). *Handbook of psychological testing*. London : Routledge.
- Laberon, S. (2001). *Les processus psychologiques à l'œuvre dans la procédure de recrutement*. Université Bordeaux 2. Thèse de doctorat.
- 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*, 377-385.
- Laveault, D., & Grégoire, J. (1997). *Introduction aux théories des tests en sciences humaines*. Paris : De Boeck Université.
- Lévy-Leboyer, C. (1996). *La gestion des compétences*. Paris : Les Éditions d'Organisation.
- Lévy-Leboyer, C. (2000). *Le 360°. Outil de développement personnel*. Paris : Éditions d'Organisation.
- Lévy-Leboyer, C. (2001). *Évaluation du personnel. Quels objectifs ? Quelles méthodes ?* Paris : Éditions d'Organisation.
- Lewin, K., Lippitt, R., & White, R. K. (1939). Patterns of aggressive behavior in experimentally created social climates. *The Journal of Social Psychology, 10*, 271-299.
- Lievens, F. (1996). Évaluation des centres d'évaluation : indices de validité empirique. *Psychologie et Psychométrie, 17*, 63-74.
- Lievens, F. (1998). Factors which improve the construct validity of assessment centers : A review. *International Journal of Selection and Assessment, 6*, 141-152.
- Lievens, F. (2001). Assessors and use of assessment centre dimensions : a fresh look at a troubling issue. *Journal of Organizational Behavior, 22*, 203-221.

- 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*, 675-686.
- Lievens, F., Harris, M. M., Van Keer, E., & Bisqueret, C. (2003). Predicting cross-cultural training performance : the validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology, 88*, 476-489.
- Lowry, P. E. (1995). The assessment center process : Assessing leadership in the public sector. *Public Personnel Management, 25*, 307-322.
- Mahfooz, A. A. (1990). *Managing People at Work : Leadership Styles and Influence Strategies*. Newbury Park, CA : Sage Publications.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interview : a comprehensive review and meta-analysis. *Journal of Applied Psychology, 79*, 599-616.
- McEvoy, G. M., & Beatty, R. W. (1989). Assessment centers and subordinate appraisals of managers : a seven-year examination of predictive validity. *Personnel Psychology, 42*, 37-52.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York : McGraw-Hill.
- 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.
- Robertson, I. T., & Kandola, R. S. (1982). Work sample test : Validity, adverse impact and applicant reaction. *Journal of Occupational Psychology, 55*, 171-183.
- Rolland, J. P. (1999). Construct validity of In-Basket Dimension. *European Review of Applied Psychology, 49*, 251-259.
- Russel, C. J., & Domm, D. R. (1995). Two field tests of an explanation of assessment center validity. *Journal of Occupational and Organizational Psychology, 68*, 25-47.
- Sackett, P. R. (1982). A critical look at some common beliefs about assessment centers. *Public Personnel Management Journal, 11*, 140-147.
- Sackett, P. R., & Dreher, G. F. (1982). Construct and assessment center dimensions : some troubling empirical findings. *Journal of Applied Psychology, 67*, 401-410.
- Sackett, P. R., & Ryan, A. M. (1985). A review of recent assessment center research. *Assessment center Review, 4*, 13-27.
- 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.
- Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. (2002). A new frame for frame-of-reference training : Enhancing the construct validity of assessment centers. *Journal of Applied Psychology, 87*, 735-746.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology : Practical and theoretical implication of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

- Seibert, S. E., Crant, J. M., & Kraimer, M. L. (1999). Proactive personal and career success. *Journal of Applied Psychology, 84*, 416-427.
- Silverman, W. H., Dalessio, A., Woods, S. B., & Johnson, R. L. (1986). Influence of assessment center methods on assessors' ratings. *Personnel Psychology, 39*, 565-578.
- Srull, T. K., & Wyer, R. S., Jr. (1989). Person memory and judgment. *Psychological Review, 96*, 58-83.
- Thornton, G. C. I., & Byham, W. C. (1982). *Assessment Center and managerial performance*. San Diego, CA : Academic Press.
- Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*, 372-376.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal : a quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189-205.

## NOTES

1. Nous appelons ici compétences « la mise en œuvre intégrée d'aptitudes, de traits de personnalité et aussi de connaissances acquises pour mener à bien une mission complexe dans le cadre de l'entreprise qui a chargé l'individu, et dans l'esprit de ses stratégies et de sa culture » (Lévy-Leboyer, 1996, p. 26).
2. 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. 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).
3. Le jeu de rôle est une situation dans laquelle le candidat doit interagir avec d'autres personnes, chacun des protagonistes ayant des objectifs différents. Il s'agit d'un exercice dans lequel la négociation joue un rôle prépondérant (recevoir une personne qui demande une augmentation, par exemple)
4. L'entretien structuré est la technique d'entretien la plus satisfaisante en termes de validité prédictive (McDaniel, Whetzel, Schmidt & Maurer, 1994). L'entretien structuré se distingue de l'entretien classique dans la mesure où les questions posées sont standardisées. Il existe cependant plusieurs niveaux de standardisation (Huffcutt & Woerh, 1999) qui vont d'une standardisation élevée (questions identiques et dans le même ordre) à une standardisation faible ou inexistante caractérisant les entretiens non structurés (lorsque les évaluateurs posent les questions qui leur viennent à l'esprit).
5. L'épreuve de groupe est un exercice durant lequel plusieurs candidats sont réunis dans le cadre d'une négociation.
6. La validité convergente peut être considérée comme une modalité particulière de la validité concurrente. Dans le cadre d'une M.T.M.M., les coefficients de validité convergente sont des corrélations entre des dimensions similaires mesurées par des exercices différents. Des coefficients élevés témoignent d'une validité de construit satisfaisante.
7. La validité prédictive est un cas particulier de la validité de critère (qui comprend aussi la validité concurrente). Dans le cadre de la validité concurrente, les deux mesures sont réalisées en même temps (le critère sert ici de référence). Dans le cadre de la validité prédictive, les deux mesures sont séparées par un laps de temps plus ou moins grand (le critère est ici une variable dépendante).
8. La notion de « *frame of reference training* » a été développée par Bernardin & Buckley (1981). Cette formation est conduite à partir d'exemples concrets. Elle a pour objectifs d'harmoniser les

pratiques de l'évaluation dans un contexte donné, de définir précisément les dimensions évaluées, d'opérationnaliser ces dimensions et de fournir aux observateurs un « *feedback* » sur la pertinence de leurs évaluations.

9. Cette méthode se retrouve dans la littérature sous le nom de « *frame of reference training* » (Schleicher, Day, Mayes & Riggio, 2002).

10. Selon le scree-test, on ne retient que les facteurs « à gauche » du point d'inflexion sur la courbe des valeurs propres. En effet, à mesure que la matrice se résidualise, les facteurs tendent à ne plus représenter que de la variance erreur.

11. Lévy-Leboyer utilise le terme de « structure d'initiative » (Doron & Parot, 1991)

12. L'erreur associée à la mesure de cette variable est donc réduite.

13. Les deux autres facteurs évoqués par les auteurs sont la persuasion (dominance, résistance au stress, confiance en soi, décision) et la réussite (ambition, niveau d'énergie).

14. Rappelons que la dimension « autoritarisme » n'apporte aucune variance expliquée dans la prédiction de notre critère de réussite professionnelle.

15. Il faut cependant noter que des outils récents comme par exemple l'évaluation à 360° permettent de pallier en partie cette difficulté (Lévy-Leboyer, 2000).

16. En outre, la définition opérationnelle de ces dimensions permet de faire un compte-rendu extrêmement détaillé aux sujets évalués, ceci en termes comportementaux, ce qui permet d'éviter certains biais courants dans ce domaine (dont la tendance à la catégorisation).

17. Des analyses factorielles exploratoires et confirmatoires menées par ces auteurs montrent que les évaluations faites par les managers se regroupent en deux facteurs (performance et relations interpersonnelles), alors que les évaluations faites par les psychologues se regroupent en 5 facteurs (relations humaines, sensibilité, initiative, organisation, analyse des problèmes), facteurs qui correspondent aux dimensions obtenues par l'analyse de poste. Malheureusement, notre effectif n'est pas suffisant pour mener des A.F. séparées.

---

## RÉSUMÉS

Si les études antérieures ont établi la validité prédictive des Assessment Center (A.C.) dans le domaine de l'évaluation du personnel et du recrutement, leur validité de construit n'a pas été démontrée. Dans le cadre d'un travail plus large visant à construire et valider un « Assessment Center » destiné à des officiers de la marine nationale française, nous avons élaboré une épreuve de groupe (discussion sans leader assigné), dont nous présenterons ici la construction et la validation (validité de contenu, de construit et validité prédictive) ; ces validités ont été testées sur un échantillon de 125 officiers de marine. Nous décrivons ici la construction de grilles d'observation « comportementales » et la formation des observateurs. Tout d'abord, des analyses factorielles exploratoires ont été menées sur les données collectées et trois dimensions furent identifiées (prise d'initiatives efficaces, travail en équipe, autoritarisme). Ensuite, des régressions linéaires multiples ont été calculées entre ces trois facteurs et un critère ultérieur de réussite professionnelle. Ces dimensions ont avec le critère un coefficient de régression multiple significatif ( $R^2 = .43$ ). Les mises en situation apportent des informations complémentaires précieuses que les tests psychométriques classiques ne permettent pas de capturer, soit les interactions entre individus et situations professionnelles.

If past research has established the predictive validity of Assessment Centers (A.C.) in the field of personnel assessment and recruitment, their construct validity has not been demonstrated. Within the framework of a larger study aiming to construct and validate an Assessment Center for French Navy officers, we elaborated a group situation (leaderless group discussion), whose construction and validation (content, construct and predictive validity) will be presented here. The validities were tested on a sample of 125 navy officers. We describe here the construction of behavioral checklists and the training of the observers. First, exploratory factor analyses were conducted on the data gathered, and three factors were identified (initiative, teamwork, authoritarianism). Then, linear multiple regressions were calculated between these three factors and a criterion of further professional performance. These factors have a significant multiple regression coefficient with the criterion ( $R^2 = .43$ ). The situational tests bring us precious complementary information that classical psychometrical tests cannot capture, e.g. the interactions between individuals and professional situations.

## INDEX

**Mots-clés :** Assessment center, Épreuve de groupe, Validité de construit, Validité prédictive

**Keywords :** Construct validity, Group exercise, Predictive validity

## AUTEURS

### XAVIER BORTEYROU

Xavier Borteyrou est ATER au département psychologie de l'université Bordeaux 2 Victor Segalen (2005-2006), (Laboratoire de psychologie EA 3662, université Bordeaux 2 Victor Segalen, 3 ter place de la Victoire, 33076 Bordeaux Cedex). Il vient de soutenir une thèse sur les liens entre personnalité, intelligence, mises en situation et réussite professionnelle chez des officiers de marine. Courriel : xavier.borteyrou@u-bordeaux2.fr.

### NICOLE RASCLE

Nicole Rascle est professeur de psychologie à l'université Bordeaux 2 Victor Segalen. Elle dirige actuellement le laboratoire de Psychologie EA 3662 ainsi que l'équipe de psychologie de la santé.

### MARILOU BRUCHON-SCHWEITZER

Marilou Bruchon-Schweitzer est professeur émérite à l'université Bordeaux 2 Victor Segalen. Elle a travaillé sur l'évaluation de la personnalité, sur les stéréotypes associés à l'apparence physique, puis a mené des recherches en psychologie de la santé et créé l'AFPSA (Association française de Psychologie de la Santé).

### PIERRE COLLOMB

Pierre Collomb est officier de la Marine nationale, docteur en psychologie. ancien chef de service d'études en ingénierie psychologique de la Marine nationale, il a mis en place un centre d'évaluation. Actuellement chargé de mission R.H. à la préfecture maritime de Brest.