

DOMINIQUE J. FERRAND

JEAN-MARC MARTEL

Le choix multicritère des items d'une échelle de mesure

Mathématiques et sciences humaines, tome 89 (1985), p. 35-59

http://www.numdam.org/item?id=MSH_1985__89__35_0

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1985, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

LE CHOIX MULTICRITERE DES
ITEMS D'UNE ECHELLE DE MESUREDOMINIQUE J. FERRAND¹
JEAN-MARC MARTEL^{2*}

La recherche est sans cesse confrontée dans ses tentatives d'opérationnalisation à l'épineuse question de la mesure de ses concepts. La construction d'une échelle à partir d'items multiples constitue une des méthodes privilégiées parmi les plus usitées de conception d'un instrument de mesure.

La démarche de construction d'une échelle (peu importe d'ailleurs qu'elle soit de type Guttman, Likert, sémantique différentielle ou tout autre modèle similaire) consiste, dans un premier temps, à générer une multitude d'items à partir de la définition du concept et, ensuite à chercher à conférer à l'instrument de mesure, constitué par un ensemble d'items, une qualité métrologique satisfaisante en procédant à une analyse des items.

L'analyse des items est le processus par lequel on sélectionne quelques items parmi un ensemble d'items candidats à partir de critères multiples de fiabilité et de validité. Les critères de validité testent si l'instrument

¹ Professeur à l'Université du Québec à Hull

² Professeur à l'Université Laval (Québec)

* Les auteurs remercient M. Jean Couillard pour sa collaboration à la programmation de l'algorithme ELECTRE. Cette recherche a été financée conjointement par le F.I.R. de l'Université du Québec à Hull et le Fonds F.C.A.C. (EQ-2594).

mesure ce qu'il est censé mesurer. Les critères de fiabilité analysent si, toutes choses étant par ailleurs égales, l'instrument est susceptible d'établir des mesures analogues d'une fois à l'autre.

L'analyse des items cependant pose un problème délicat. Certes, la fiabilité croît avec le nombre d'items, mais chaque nouvel item accroît la durée de l'expérimentation, et le gain marginal de fiabilité est largement décroissant. Pour dissiper cette difficulté, on cherche à retenir le plus petit nombre suffisant d'items.

En l'absence de méthode définie, le choix des items à partir des critères multiples de fiabilité et de validité était laissé à l'arbitraire de l'analyste. Pour remédier à cette lacune, nous proposons d'utiliser la méthode ELECTRE II qui permet de classer les items candidats à partir de relations de surclassement.

Cette méthode exige toutefois que soit précisé le rôle joué par chacun des critères dans l'analyse. La détermination de l'importance relative des critères permet alors de concevoir une procédure d'analyse et de choix systématique des items, que nous examinerons à travers une application.

I - FIABILITE, VALIDITE ET ANALYSE DES ITEMS

Selon la théorie de la fiabilité, l'utilité d'un instrument de mesure dépend de sa capacité à éviter d'enregistrer des erreurs. La fiabilité est une estimation de la performance de l'instrument de mesure évaluée à partir d'un coefficient de corrélation. Plus précisément, le coefficient de fiabilité calculé à partir de deux mesures parallèles X et X' s'établit comme le rapport de la covariance entre X et X' sur le produit de leur écart-type respectif:

$$\rho_{XX'} = \frac{\sigma_{XX'}}{\sigma_X \sigma_{X'}} = \frac{\sigma_R^2}{\sigma_X^2}, \quad (1)$$

Le rapport (1) montre que l'importance de l'erreur diminue la valeur du coefficient de fiabilité. De façon opérationnelle, on calcule le coefficient de fiabilité suivant la méthode du test-retest ou celle de l'équivalence (Bohrnstedt, 1969).

La méthode du test-retest est le plus souvent exclue de l'analyse en raison de ses nombreux inconvénients: il est en effet impossible de contrôler l'effet des changements qui peuvent intervenir entre chaque administration de l'instrument de mesure, et l'apprentissage provoqué par une première pratique de l'instrument. On préfère en général à cette méthode, celle de l'équivalence qui comporte deux types d'approche: la mesure de bissection (split-half) dont le coefficient de fiabilité est le plus souvent estimé à partir de la formule de Spearman-Brown (1910), et la mesure de la consistance interne qui, en général, s'effectue à partir du coefficient alpha de Cronbach (1951).

Ces coefficients, parmi les plus usuels, montrent que la fiabilité s'accroît avec le nombre d'items. Il faut néanmoins remarquer que le gain marginal de fiabilité par item supplémentaire est décroissant, et il est d'autant plus décroissant que la fiabilité des premiers items est élevée. Bohrnstedt (1969) rapporte un diagramme sur l'évolution du coefficient de Spearman-Brown en fonction du nombre d'items pour des coefficients initiaux de fiabilité différents:

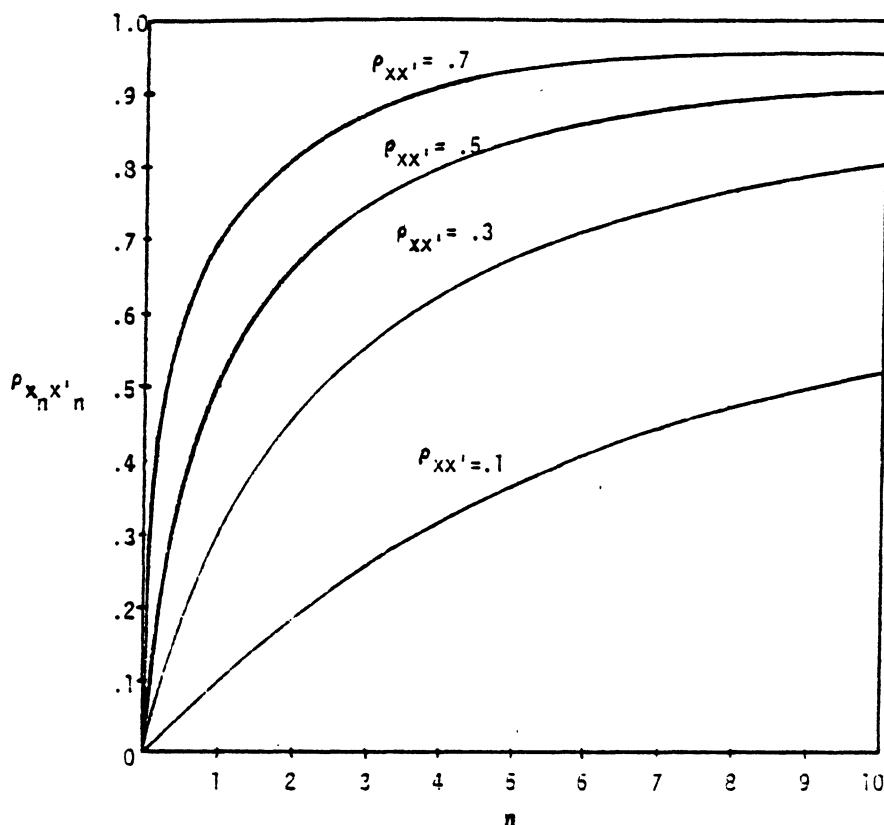


Figure 1. Accroissement de fiabilité en fonction du nombre d'items (corrélacion inter-item constante) d'après Borhnstedt (1969)

Ce schéma montre qu'il suffirait d'introduire un grand nombre d'items pour s'assurer d'un coefficient de fiabilité satisfaisant. Mais une telle démarche se heurte à des obstacles d'ordre opérationnel et théorique. L'addition d'items a un coût habituellement estimé élevé qui s'exprime en temps d'expérimentation, et le gain marginal de fiabilité par item supplémentaire est largement décroissant. Il s'avère alors important de sélectionner le plus petit ensemble suffisant d'items qui garantiront une fiabilité et une validité satisfaisantes de l'échelle. L'analyse s'appuie alors sur de multiples critères à partir desquels peut s'effectuer un choix. Les critères sont essentiellement définis à partir des différents types de validité (contenu, critère, construit), la fiabilité n'en introduisant qu'un seul.

La corrélation ρ_{ih} d'un item candidat au total de l'échelle H, composée de tous les items candidats, constitue le principal critère relié à la fiabi-

lité. Toutefois, l'item étant inclus dans l'échelle H, on estime également un coefficient de corrélation corrigé ρ_{ih} , qui mesure la corrélation d'un item candidat à l'échelle modifiée H' composée de tous les autres items candidats (Cureton, 1966). La différence $\rho_{ih} - \rho_{ih}'$ est alors un indice de contribution de l'item à l'échelle*.

La validité de contenu réfère au degré auquel l'échelle couvre les diverses facettes du concept qu'elle mesure. Comme il n'existe pas de critère statistique pour évaluer la validité de contenu, on recourt à une procédure qui offre une garantie relative de couverture du concept. Le champ du contenu est stratifié, et chaque strate (ou dimension) est divisée en sous-strates (ou sous-dimensions) jusqu'à l'obtention des strates (ou composantes) élémentaires. Il s'agit alors de déterminer si les items appartenant à une même strate ou dimension présentent entre eux des corrélations plus élevées que celles enregistrées avec les items des autres strates. Le critère ρ_{id} mesure la relation linéaire entre l'item i et une dimension d, et ρ_{id}' correspond à sa corrélation corrigée. La différence $\rho_{id} - \rho_{id}'$ est alors un indice de contribution de l'item à la dimension. Si $\rho_{i\bar{d}}$ représente la corrélation de l'item aux autres dimensions, alors $1 - \rho_{i\bar{d}}$ est un indice d'indépendance de l'item aux k-1 autres dimensions.

La validité associée à un critère vérifie si l'échelle est corrélée à une mesure directe du concept. Le critère ρ_{ic} estime l'intensité de la relation linéaire entre un construit-type et l'item i, et $1 - \rho_{i\bar{c}}$ constitue un indice d'indépendance de l'item aux autres construits-types.

La validité de construit examine la correspondance entre la théorisation et l'opérationnalisation de l'échelle: il s'agit de savoir si l'item mesure,

* Si le nombre d'items est grand, cette différence peut être assez faible.

et avant tout autre, le concept pour lequel il a été conçu. La différence $\rho_{id} - \rho_{ih}$ représente alors un indice d'appartenance de l'item à la dimension qu'il est censé mesurer. Mais cette convergence établie, il faut également s'assurer que l'échelle est capable de discriminer. Le critère de différenciation est un test du t qui examine s'il existe une différence significative pour chaque item sur un échantillon donné entre les 27% des réponses enregistrant les scores les plus élevés et les 27% des réponses ayant les scores les plus faibles (Edwards, 1957). On peut enfin adjoindre à ce critère une mesure de la centralité des réponses g_i qui indique la qualité de la relation entre l'énoncé de l'item et sa gradation: elle correspond à la valeur absolue de la moyenne des différences entre les scores et la valeur centrale de la gradation de l'item.

Chaque item, étant évalué sur ces multiples critères, se pose un problème de choix: il consiste à identifier le plus petit ensemble suffisant d'items offrant à la fois les qualités métrologiques indispensables de fiabilité et de validité et la représentation la plus fidèle de l'ensemble des dimensions composant le concept à mesurer. La complexité de la tâche nécessiterait, dans la plupart des cas, l'intervention d'un outil d'aide à la décision pour que s'opère de façon adéquate ce choix. La méthode ELECTRE (Elimination Et Choix Traduisant la Réalité) semble la plus appropriée pour répondre à ce besoin.

II - LA METHODE ELECTRE

Pour résoudre le problème du choix des items, nous pouvons envisager d'agréger les appréciations enregistrées par les divers critères. Cette solution s'avère peu acceptable: bien que les valeurs attribuées aux critères soient de nature statistique, ceux-ci reflètent des points de vue différents. Il est donc préférable d'opter pour une méthode qui respecte cette diversité. ELECTRE II permet d'établir un classement sur un ensemble fini

d'objets en tenant compte simultanément de plusieurs critères non aisément réductibles en un seul.

Dans le cas présent, nous disposons d'un ensemble de n items candidats, et pour chaque item i , $i = 1, 2, \dots, n$, d'une série $\gamma_1(i), \gamma_2(i), \dots, \gamma_m(i)$ de scores sur chacun des j , $j = 1, 2, \dots, m$, critères. Nous obtenons alors un tableau à double entrée tel qu'illustré (Figure 2).

		CRITERES				
		1	j	m		
ITEMS	1	$\gamma_1(1)$	\dots	$\gamma_j(1)$	\dots	$\gamma_m(1)$
	.					
	.					
	i	$\gamma_1(i)$	\dots	$\gamma_j(i)$	\dots	$\gamma_m(i)$
	.					
	.					
	n	$\gamma_1(n)$	\dots	$\gamma_j(n)$	\dots	$\gamma_m(n)$

Figure 2. Formalisation canonique du problème de choix multicritère des items d'une échelle de mesure

Comme, en général, aucun item n'apparaît meilleur que tous les autres sur l'ensemble des critères, nous sommes conduits à rechercher les items qui, à la fois, réalisent un "bon" score sur le plus grand nombre possible de critères (en considérant leur importance relative) et n'affichent pas de trop "mauvais" scores sur les autres. Aussi, importe-t-il de définir les conditions qui autorisent à affirmer, étant donné deux items, ou que l'un est globalement préférable à l'autre, ou qu'il est équivalent, ou qu'il s'avère impossible de conclure le risque d'erreur étant trop grand. La notion de "globalement préférable" est formalisée par une relation dite de surclassement notée S .

Une relation binaire S_I définie sur l'ensemble I des items est appelée relation de surclassement si, pour toute paire d'items, on se prononce pour une situation:

- d'indifférence lorsque $iS_I i'$ et $i'S_I i$,
- de préférence stricte ou large lorsque $iS_I i'$ et non $i'S_I i$,
- d'incomparabilité lorsque non $iS_I i'$ et non $i'S_I i$, où i et $i' \in I$

(Roy, 1974).

La décision d'opter pour l'une de ces trois situations pour une paire d'items (i, i') dépend des vecteurs de scores $\underline{\gamma}(i) = (\gamma_1(i), \dots, \gamma_j(i), \dots, \gamma_m(i))$ et $\underline{\gamma}(i') = (\gamma_1(i'), \dots, \gamma_j(i'), \dots, \gamma_m(i'))$, et c'est à partir de ces scores que s'élabore la relation de surclassement S_I . L'idée de base de la méthode ELECTRE est sur ce point relativement intuitive et naturelle: si un item i est au moins aussi bon que i' sur la plupart des critères (concordance) et si par ailleurs il n'existe pas de critère stipulant que i est beaucoup plus mauvais que i' (discordance), alors les chances sont grandes pour que i soit préférable à i' , c'est-à-dire que i surclasse i' . La relation de surclassement est ainsi conçue à partir de deux indices: un indice de concordance et un indice de discordance*.

Une façon simple de calculer un indice de concordance consisterait à retenir le nombre de critères pour lesquels i est au moins aussi bon que i' . Cependant, tous les critères ou points de vue ne présentent pas nécessairement le même intérêt. Aussi, associe-t-on aux critères des poids P_j exprimant leur importance relative. On peut alors calculer l'indice de concordance en effectuant la somme des poids pour lesquels i est au moins aussi bon que i' , somme qui se normalise en divisant le résultat obtenu par la somme de

* Le lecteur ayant une bonne connaissance d'Electre II pourra passer directement à la section III: "Application à la Construction d'une Echelle de Mesure".

tous les poids. Ainsi, pour la paire d'items (i, i') l'indice de concordance $C(i, i')$ est tel que:

$$C(i, i') = \frac{1}{P} \sum_{j \in C_{ii'}} p_j, \quad (2)$$

où $P = \sum_{j=1}^m p_j$, et où $C_{ii'} = \{j: \gamma_j(i) \geq \gamma_j(i')\}$ est l'ensemble des critères procurant un score au moins aussi élevé à l'item i qu'à l'item i' . On peut vérifier que cet indice varie de façon non décroissante entre 0 et 1 avec l'enrichissement de l'ensemble $C_{ii'}$.

Il est en général nécessaire, mais non suffisant, d'atteindre un degré relativement élevé de concordance pour déclarer que l'item i surclasse l'item i' . En effet, certains parmi les critères j vont parfois à l'encontre de la relation "i est au moins aussi bon que i' ". Aussi, faut-il introduire des conditions supplémentaires sur ces écarts négatifs pour décider du surclassement de i' par i . Il semble fondé de refuser ce surclassement dès l'instant où $\gamma_j(i)$ apparaît trop "discordant" comparé à $\gamma_j(i')$ pour $j \in C_{ii'}^-$, $C_{ii'}^- = \{j: \gamma_j(i) < \gamma_j(i')\}$. Ainsi, pour chaque paire d'items (i, i') l'indice de discordance selon le critère j est établi de la manière suivante:

$$D_j(i, i') = \begin{cases} 0 & \text{si } \gamma_j(i) \geq \gamma_j(i') \\ \gamma_j(i') - \gamma_j(i) & \text{si } \gamma_j(i) < \gamma_j(i') \end{cases} \quad (3)$$

Ce n'est pas la seule façon de définir ces indices et il pourrait être utile de la modifier en fonction du contexte.

Les indices de concordance et de discordance calculés, on pourra dire qu'un item i surclasse un item i' , $iS_I i'$, où I est l'ensemble des items analysés, si les trois conditions suivantes sont réunies:

1) Le rapport de la somme des poids des critères pour lesquels i est considéré comme au moins aussi bon que i' sur la somme des poids des critères pour lesquels i est considéré moins bon que i' doit être plus grand que 1; soit:

$$\frac{P(i, i')}{P^-(i, i')} > 1, \text{ où } P(i, i') = \sum_{j \in C_{ii'}} P_j \text{ et } P^-(i, i') = \sum_{j \in C_{ii}^-} P_j .$$

2) La somme des poids des critères pour lesquels i est considéré comme au moins aussi bon que i' , $P(i, i')$ doit être "suffisamment élevée": cette condition estime l'importance du nombre de relations favorables au surclassement de i' par i .

3) Pour tout critère j ne vérifiant pas la relation i est moins bon que i' , la différence de score $D_j(i, i') = \gamma_j(i') - \gamma_j(i)$ ne doit pas être "trop importante". C'est une condition statuant sur l'ampleur de la discordance.

L'estimation d'un niveau "suffisamment élevé" ou celle d'un niveau "pas trop important" sont spécifiées par des seuils dits de concordance et de discordance. Le niveau des seuils dépend à la fois des exigences de l'analyste et du taux de risque qu'il est prêt à accepter dans la construction des relations de surclassement*.

La combinaison des seuils avec les conditions de concordance et de discordance permet d'interpréter la force de la relation de surclassement. Dans ELECTRE II, le surclassement est évalué comme fort ou faible en fonction des relations observées entre trois seuils de concordance ($\{c_1, c_2, c_3\}$)

* Le choix de ces seuils relève autant d'une appréciation subjective que d'un calcul d'erreur au sens de la physique. Il ne s'agit pas ici de grandeurs expérimentales, mais de grandeurs d'opportunité utiles pour scinder l'approximatif du significatif dans les données. Aussi, les valeurs attribuées aux seuils contiennent-elles une part inévitable d'arbitraire.

avec $1 > c_1 > c_2 > c_3 > 0$) et deux seuils de discordance ($\{d_j(1), d_j(2)\}$ avec $d_j(1) < d_j(2)$) dont le niveau dépend des échelles choisies pour enregistrer les appréciations des critères*.

L'item i surclasse fortement i' , noté $iS_I(F)i'$, si en plus de la condition (1) énumérée précédemment:

$$\text{soit } \begin{cases} C(i, i') \geq c_1 \\ \text{et} \\ D_j(i, i') \leq d_j(2) \\ \text{pour tout critère } j \end{cases} \quad \text{soit } \begin{cases} C(i, i') \geq c_2 \\ \text{et} \\ D_j(i, i') \leq d_j(1) \\ \text{pour tout critère } j \end{cases} \quad (4)$$

L'item i surclasse faiblement i' noté $iS_I(f)i'$ si en plus de la condition (1):

$$C(i, i') \geq c_3 \text{ et } D_j(i, i') \leq d_j(2) \text{ pour tout critère } j. \quad (5)$$

L'évaluation, en terme de forte ou de faible, de la relation de surclassement pour chaque paire d'items permet de construire un graphe qui procure, en se basant sur la longueur des chemins, d'une part incidents et d'autre part issus, deux classements extrêmes** de l'ensemble des items: les relations de surclassement faibles sont introduites pour départager les items estimés équivalents selon le surclassement fort. Il est alors possible d'obtenir un ordre unique (préordre complet) en effectuant un classement final, situé en quelque sorte à égale distance des classements précédents, et dont les rangs correspondent à la moyenne des rangs obtenus dans les

* On retient habituellement comme valeurs standards pour les seuils de concordance $c_1 = 3/4$, $c_2 = 2/3$, $c_3 = 3/5$; les seuils de discordance sont fixés par l'analyste qui définit une amplitude de discordance acceptable.

** Un chemin incident à un sommet i est l'ensemble des sommets (i_1, i_2, \dots, i_k) vérifiant $i_1 S_F i_2, i_2 S_F i_3, \dots, i_{k-1} S_F i_k$, ou S_F signifie: surclassé fortement. Un chemin issu d'un sommet i est l'ensemble des sommets $(i^1, i^2, \dots, i^\ell)$ vérifiant $i S_F i^1, i^1 S_F i^2, \dots, i^{\ell-1} S_F i^\ell$. La longueur d'un chemin incident ou issu est le nombre de sommets figurant dans l'ensemble. La longueur des chemins incidents détermine le classement direct, et celle des chemins issus, le classement indirect.

classements extrêmes (Bertier et Roy, 1972). Ce classement final prend d'autant plus de sens que les classements extrêmes sont peu dissemblables; dans le cas de dissimilarité notoire, il existe de nombreuses incomparabilités insolubles selon la méthode, les données n'offrant pas une garantie suffisante.

III - APPLICATION A LA CONSTRUCTION D'UNE ECHELLE DE MESURE

L'utilité de cette procédure à la fois d'analyse et de choix des items apparaît clairement à la lumière d'un exemple. Nous nous servons à cette fin des données recueillies lors du prétest d'une échelle de type Likert baptisée "Indice de Motivation Technocratique et Politique" (Ferrand, 1982). L'objet de cette échelle est de déterminer dans quelle mesure les utilisateurs d'un système d'information attendent des récompenses de nature technocratique et/ou politique lorsqu'un projet de développement de ce système est engagé. Une récompense est dite de type technocratique lorsque sa finalité s'exprime en termes de moyens. Elle est dite de type politique lorsqu'elle se définit strictement en termes de finalité (Poitou, 1978).

Pour construire les items, nous avons divisé chaque dimension en ses composantes les plus élémentaires. La dimension "Politique" fut éclatée en quatre sous-dimensions correspondant aux quatre sources fondamentales du pouvoir (Shukla, 1977). La dimension "Technocratique" fut différenciée à partir des critères d'efficience et d'efficacité dans les systèmes d'information tels qu'exposés par Jenkins (1977) et Cheney (1977). Les diverses définitions attachées aux sous-dimensions permettent d'identifier des composantes élémentaires. Nous obtenons le tableau de la figure 3.

Dimensions	Sous-dimensions	Composantes Élémentaires
Politique	Pouvoir Statutaire	Responsabilité, Autorité
	Pouvoir d'Expertise	Acquisition, Considération
	Pouvoir de Contrôle	Contrôle des Sources, Autonomie, Coordination
	Pouvoir Politique	Négociation, Coalition, Persuasion
Technocratique	Qualité de l'information (Efficience)	Rapidité, Compréhension, Utilité
	Qualité du système (Efficience)	Flexibilité, Intégration, Exhaustivité et Précision
	Qualité de la décision (Efficacité)	Pertinence, Temps, Conformité aux objectifs

Figure 3. Les composantes élémentaires des deux dimensions de l'échelle de mesure

Une trentaine d'items, issus des composantes élémentaires des deux dimensions, furent présentés séparément à trois experts du domaine des systèmes d'information qui en analysèrent la sémantique. Nous avons rejeté tout item qui ne recevait pas l'unanimité de compréhension des experts. Suite à leurs avis, 24 items furent retenus, parmi lesquels deux items correspondaient à des mesures directes de chacune des dimensions. L'échelle variait de 1 (absolument pas important) à 6 (extrêmement important). A titre d'exemple, elle comprenait pour la dimension politique des items comme: "j'ai préféré ce mode d'exploitation du système d'information parce qu'il correspond le mieux aux responsabilités liées à ma position hiérarchique", ou pour la dimension technocratique, des items comme: "j'ai préféré ce mode d'exploitation du système parce qu'il est le plus approprié pour conduire

à une amélioration dans la rapidité d'obtention de l'information". Les items formant l'échelle prétestée furent répartis suivant une suite de nombres aléatoires*. L'échelle fut administrée, afin d'être prétestée, à 54 utilisateurs importants de systèmes d'information ordinés.

Le coefficient de Spearman-Brown s'élevait à .84 et le coefficient α de Cronback était de .86, ce qui s'avérait d'excellents résultats, .80 étant considéré comme un taux de fiabilité très satisfaisant. Nous procédions alors au calcul des divers coefficients de corrélation, du test t et de la centralité de l'item. Les résultats apparaissent dans le Tableau 1.

Avant d'effectuer le traitement par l'algorithme ELECTRE II, il est nécessaire d'assigner un poids reflétant l'intérêt relatif de chaque critère. Six experts** ont accepté de participer à la définition, dans le cadre de la méthode ELECTRE II, de l'importance relative des poids des critères en suivant une procédure simplifiée de la méthode DELPHI: la consultation de panels (Helmer, 1966). Il importait notamment ici de préciser les mobiles qui légitiment le choix des poids relatifs de chaque critère. Nous avons enregistré les résultats suivants.

La corrélation de l'item à l'échelle ρ_{ih} a reçu un poids de (1). Bien qu'il soit important d'examiner l'incidence de l'item sur l'ensemble de l'échelle, celle-ci comprenant plusieurs dimensions, il est difficile d'en estimer l'impact réel. Le poids de l'indice de contribution de l'item à l'échelle $\rho_{ih} - \rho_{ih'}$ est seulement de (1) puisqu'il reprend en partie la corrélation de l'item à l'échelle.

* Provenant du générateur de nombres aléatoires APL-SV de l'Université Laval.

** Psychométriciens.

Il fut attribué un poids de (1) à la corrélation de l'item à la dimension ρ_{id} , les corrélations subséquentes $\rho_{id} - \rho_{id'}$, indice de contribution de l'item à la dimension, et $1 - \rho_{id}$, indice d'indépendance de l'item aux autres dimensions recevant un poids de (2). La distribution des poids de ces indices, important en raison de leur référence directe à la validité de contenu, s'explique par leur interdépendance.

La validité de construit reçue globalement, un poids de (4), répartie entre l'estimation de l'intensité de la relation entre le construit-type et l'item ρ_{ic} et l'indice d'indépendance de l'item aux autres construits-types $(1 - \rho_{ic})$.

Un poids de (2) a été accordé à l'indice d'appartenance de l'item à la dimension $\rho_{id} - \rho_{ih}$, la mesure de convergence de l'échelle étant un facteur important. Il fut admis que la capacité d'un item à discriminer s'avérait une caractéristique majeure de la qualité métrologique de l'échelle, puisqu'elle constitue sa raison d'être. Il a été de ce fait attribué un poids de (5) au critère de différenciation mesuré à partir d'un test du t. Enfin, le critère de centralité des réponses g_i a été doté d'un poids de (1).

Les indices de concordance sont obtenus simplement en appliquant aux données du tableau 1 la formule (2). Le tableau 2 donne les valeurs de ces indices pour les dimensions technocratique et politique.

La condition de discordance nécessite souvent (a) que soient opérés des ajustements préalables des échelles enregistrant les appréciations des critères ou (b) que les seuils de discordance soient établis en fonction des niveaux de l'échelle. On opte pour une transformation d'échelle lorsque des variations égales sur l'échelle n'ont pas une signification identique pour tous les niveaux de l'échelle. Pour la statistique t, par exemple,

Tableau 1. Scores sur chacun des critères

DI- MEN- SIONS	ITEM	ρ_{ih}	$\rho_{ih} - \rho_{ih'}$	ρ_{id}	$\rho_{id} - \rho_{id'}$	$i - \rho_{id}$	$\rho_{id} - \rho_{ih}$	ρ_{ic}	$1 - \rho_{ic}$	t	g_i
		(1)	(1)	(1)	(2)	(2)	(2)	(2)	(2)	(2)	(5)
T	1	.24	.08	.32	.14	.90	.08	.19	.97	2.51	.97
E	4	.34	.08	.47	.15	.81	.13	.14	1.00	3.00	1.01
C	5	.51	.09	.50	.17	.64	-.01	-.03	.78	3.36	1.33
H	6	.46	.05	.56	.07	.75	.10	.40	.96	4.30	.51
N	9	.40	.06	.61	.10	.85	.21	.36	1.13	2.06	.97
O	10	.46	.05	.68	.07	.84	.22	.27	1.02	3.22	.51
C	16	.49	.07	.69	.09	.80	.20	.37	.90	3.14	1.08
R	17	.61	.05	.51	.12	.46	-.10	.49	.77	3.23	.97
A	19	.26	.06	.57	.09	1.04	.31	.24	1.19	1.44	.76
T	21	.59	.06	.77	.07	.72	.18	.28	.99	4.14	.72
I	22	.51	.06	.67	.10	.76	.16	.42	.93	3.84	.79
Q	24	.51	.06	.57	.10	.71	.06	1.00	.69	3.61	.72
U											
E											
P	2	.28	.10	.41	.05	1.08	.13	.48	1.01	1.86	2.42
O	3	.21	.10	.20	.11	.92	-.01	.02	.90	1.75	1.93
L	7	.56	.07	.57	.09	.69	.01	.35	.74	5.75	1.83
I	8	.33	.10	.42	.09	.96	.09	.25	.82	1.75	2.86
T	11	.45	.08	.56	.19	.77	.10	.27	.81	2.26	1.61
I	12	.51	.06	.61	.12	.77	.10	.27	.74	3.20	2.01
Q	13	.66	.06	.75	.11	.67	.09	.50	.91	6.55	1.69
U	14	.52	.07	.55	.14	.68	.03	.22	.74	4.22	1.04
E	15	.48	.07	.65	.11	.89	.17	.62	.94	3.80	2.79
	18	.65	.07	.62	.17	.52	-.03	.28	.82	4.72	1.65
	20	.44	.07	.50	.08	.84	.06	.50	.90	2.64	1.97
	23	.54	.08	.67	.04	.90	.13	1.00	.69	4.53	2.99

Tableau 2. Matrice des indices de concordance

C(i,i') pour la dimension technocratique

	1	4	5	6	9	10	16	17	19	21	22	24
1	0.0	0.79	0.58	0.58	0.47	0.68	0.63	0.53	0.53	0.68	0.58	0.47
4	0.26	0.0	0.58	0.47	0.53	0.79	0.63	0.47	0.47	0.58	0.58	0.47
5	0.42	0.42	0.0	0.73	0.47	0.47	0.47	0.21	0.47	0.79	0.79	0.68
6	0.42	0.57	0.26	0.0	0.58	0.63	0.53	0.37	0.58	0.79	0.63	0.42
9	0.58	0.47	0.53	0.42	0.0	0.47	0.58	0.58	0.37	0.42	0.63	0.58
10	0.32	0.21	0.53	0.63	0.53	0.0	0.42	0.63	0.53	0.68	0.63	0.58
16	0.37	0.37	0.53	0.47	0.42	0.58	0.0	0.53	0.42	0.47	0.63	0.53
17	0.53	0.53	0.79	0.68	0.47	0.68	0.47	0.0	0.42	0.68	0.68	0.68
19	0.47	0.53	0.53	0.42	0.68	0.47	0.68	0.58	0.0	0.53	0.68	0.63
21	0.31	0.42	0.21	0.58	0.63	0.42	0.53	0.32	0.53	0.0	0.42	0.32
22	0.42	0.42	0.26	0.37	0.53	0.37	0.37	0.32	0.37	0.63	0.0	0.32
24	0.53	0.53	0.37	0.58	0.58	0.39	0.47	0.32	0.47	0.79	0.89	0.0

C(i,i') pour la dimension politique

	2	3	7	8	11	12	13	14	15	18	20	23
2	0.0	0.16	0.47	0.32	0.47	0.47	0.58	0.47	0.73	0.47	0.58	0.63
3	0.89	0.0	0.58	0.79	0.68	0.74	0.79	0.68	0.84	0.58	0.74	0.63
7	0.53	0.42	0.0	0.53	0.47	0.53	0.79	0.63	0.68	0.63	0.53	0.74
8	0.74	0.53	0.58	0.0	0.68	0.68	0.79	0.47	0.79	0.68	0.58	0.63
11	0.53	0.32	0.53	0.32	0.0	0.74	0.63	0.32	0.84	0.63	0.63	0.79
12	0.53	0.26	0.58	0.32	0.58	0.0	0.63	0.58	0.84	0.74	0.37	0.79
13	0.42	0.32	0.47	0.32	0.37	0.42	0.0	0.53	0.63	0.42	0.32	0.68
14	0.53	0.32	0.79	0.53	0.68	0.53	0.74	0.0	0.58	0.79	0.53	0.79
15	0.26	0.26	0.39	0.21	0.16	0.16	0.47	0.47	0.0	0.47	0.05	0.68
18	0.53	0.42	0.68	0.42	0.37	0.26	0.84	0.53	0.58	0.0	0.53	0.74
20	0.42	0.37	0.53	0.42	0.37	0.63	0.79	0.53	1.00	0.53	0.0	0.79
23	0.47	0.39	0.53	0.37	0.26	0.21	0.58	0.47	0.32	0.53	0.21	0.0

une variation de 0.50 n'a pas la même signification si elle se situe autour d'une valeur t de 1.75 ou d'une valeur du t de 3.50; de plus, au delà d'un certain seuil, un accroissement marginal d'un score n'est pas significatif. Autre exemple: les valeurs des coefficients de corrélations ρ_{ih} , ρ_{id} , ρ_{ic} , utilisées comme critères, subissent la transformation de Fisher $\frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right)$, puis sont standardisées afin de se distribuer selon une normale (0,1). Les valeurs de ces trois variables standardisées z_{ij} , $j = h, d, c$, ainsi que les valeurs enregistrées par la statistique t sont alors transformées par la fonction de répartition de probabilité $\Phi(z)$ d'une normale (0,1). Ces diverses transformations conduisent au tableau 3.

On applique la condition de discordance à partir des données du tableau 3.

Les seuils retenus sont:

Critères j	1	2	3	4	5	6	7	8	9	10
$d_j(1)$.20	.016	.20	.03	.09	.055	.20	.07	.005	.25
$d_j(2)$.35	.032	.35	.06	.18	.11	.35	.14	.015	.50

L'algorithme ELECTRE II établit, à partir des indices de concordance et de discordance, un préordre partiel des items qui permet de construire un graphe de surclassement pour chacune des dimensions:

a) graphe de surclassement des items pour la dimension technocratique

(Figure 4),

b) graphe de surclassement des items pour la dimension politique (Figure 5).

Le classement issu de l'application de la méthode ELECTRE II est ainsi fonction, non seulement des évaluations des items par critère, mais aussi des

Tableau 3. Scores transformés sur chacun des critères

DI- MEN- SIONS	ITEM No:	ϕ_{ih}	$\rho_{ih}^{-\rho_{ih}}$	ϕ_{id}	$\rho_{id}^{-\rho_{id}}$	$1-\rho_{id}$	$\rho_{id}^{-\rho_{ih}}$	ϕ_{ic}	$1-\rho_{ic}$	$\phi(t)$	g_i
		(1)	(1)	(1)	(2)	(2)	(2)	(2)	(2)	(5)	(1)
T	1	.05	.08	.03	.14	.90	.08	.25	.97	.9940	.97
E	4	.17	.08	.18	.15	.81	.13	.16	1.00	.9987	1.01
C	5	.64	.09	.25	.17	.64	-.01	.02	.78	.9996	1.33
H	6	.48	.05	.41	.07	.75	.10	.78	.96	1.00	.51
N	9	.30	.06	.58	.10	.85	.21	.69	1.13	.9803	.97
O	10	.48	.05	.80	.07	.84	.22	.45	1.02	.9944	.51
C	16	.57	.07	.83	.09	.80	.20	.71	.90	.9992	1.08
R	17	.90	.05	.27	.12	.46	-.10	.93	.77	.9994	.97
A	19	.07	.06	.44	.09	1.04	.31	.37	1.19	.9251	.76
T	21	.86	.06	.97	.07	.72	.18	.47	.99	1.000	.72
I	22	.64	.06	.77	.10	.76	.16	.82	.93	.9999	.79
Q	24	.64	.06	.44	.10	.71	.06	1.00	.69	.9998	.72
U											
E											
P	2	.09	.10	.17	.05	1.08	.13	.78	1.01	.9686	2.42
O	3	.03	.10	.02	.05	1.08	.13	.04	1.01	.9599	1.93
L	7	.78	.07	.54	.09	.69	.01	.49	.74	1.000	1.83
I	8	.15	.10	.19	.09	.96	.09	.29	.82	.9599	2.86
T	11	.48	.08	.51	.19	.77	.10	.32	.81	.9881	1.61
I	12	.64	.06	.66	.12	.77	.10	.33	.74	.9993	2.01
Q	13	.96	.06	.96	.11	.67	.09	.82	.91	1.000	1.69
U	14	.67	.07	.48	.14	.68	.03	.23	.74	1.000	1.04
E	15	.54	.07	.77	.11	.89	.17	.96	.94	.9999	2.79
	18	.95	.07	.68	.17	.52	.03	.34	.82	1.000	1.65
	20	.41	.07	.35	.08	.84	.06	.82	.90	.9959	1.97
	23	.73	.08	.82	.04	.90	.13	1.00	.69	1.000	2.99

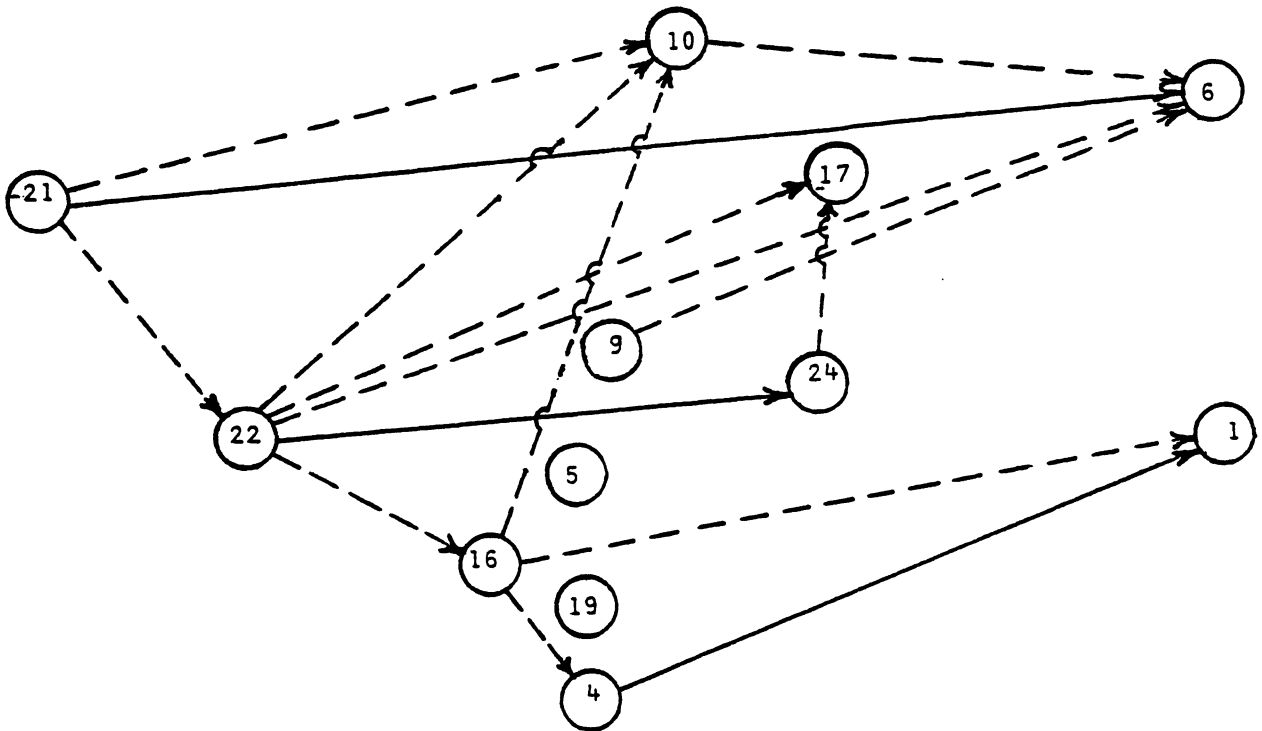


Figure 4. Graphe de surclassement des items pour la dimension technocratique

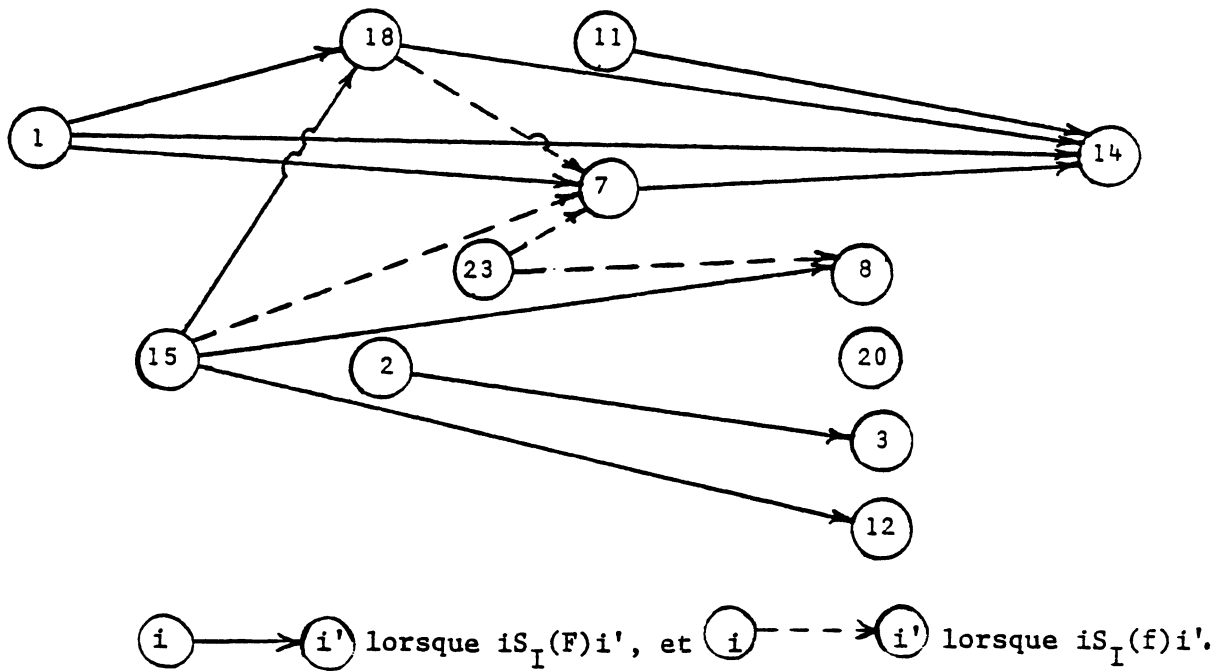


Figure 5. Graphe de surclassement des items pour la dimension politique

valeurs des paramètres (les seuils et les poids) qui interviennent dans la construction de la relation de surclassement. Toute modification des valeurs des paramètres est susceptible de changer l'ordre de classement. Les poids et les seuils n'étant que des ordres de grandeur, il importe de tester la sensibilité des résultats à des variations, dans une fourchette raisonnable, des valeurs des paramètres. Dans la mesure où ces variations perturbent peu ou pas du tout l'ordre de classement final, on sera enclin à accorder une grande confiance dans les résultats. Dans le cas opposé et limite où de légères variations viennent bouleverser l'ordre de classement final, il faudra conclure que l'information contenue dans les données est insuffisante pour permettre de départager valablement les items.

Les résultats de notre analyse de sensibilité sont présentés sur la figure 6. Ils permettent de constater que, d'une façon générale, le classement final des items ne s'est pas avéré très sensible à des variations raisonnables des seuils de discordance et des poids. Certes, par exemple, l'item 17 a gagné 6 rangs lorsque les seuils de discordance ont été modifiés pour le critère 1, mais en réalité le classement final suite à cette modification s'établit ainsi: 21, 22, {4, 5, 9, 16, 17, 19}, 10, 24, {1, 6}; alors que l'ordre avant modification était le suivant: 21, 22, {4, 5, 9, 16, 19}, 10, {17, 24}, {1, 6}. De plus, on observe que les rangs des items classés comme prioritaires se montrent très peu sensibles (à l'exception de l'item 18 qui enregistre quelques légers déplacements) aux variations de seuils et de poids tant pour la dimension "Technocratique" que la dimension "Politique".

IV - CONCLUSIONS

Les graphes issus de l'application de l'algorithme ELECTRE II représentent des priorités de choix parmi les items. Il devient alors possible de ré-

pondre aux deux questions, intimement reliées, à l'origine de notre étude: (a) combien d'items doit-on retenir? (b) le seuil de fiabilité de l'ensemble des items choisis s'avère-t-il satisfaisant?

Les courbes d'accroissement de fiabilité en fonction du nombre d'items démontrent qu'à partir d'une corrélation inter-item constante située entre .4 et .5, un nombre de 4 à 5 items permet d'atteindre un taux acceptable de fiabilité de .80 (coefficient de Spearman-Brown, cf.: Figure 1). Pour la dimension technocratique, les quatre items classés prioritaires (item 21, item 22, item 16, item 4) sur le graphe obtenaient un taux de fiabilité α de .79; pour la dimension politique, les cinq premiers items (item 13, item 15, item 18, item 2, item 23) affichaient un taux de fiabilité α de .77.

Il est notoire que l'ensemble choisi des items ne correspond pas nécessairement à l'ensemble pouvant maximiser le taux de fiabilité α . L'ensemble optimal pour le taux de fiabilité α , pour un nombre d'items donné, est constitué par le sous-ensemble des items candidats possédant les plus fortes corrélations inter-items. On perçoit à travers cette remarque l'intérêt d'une approche multicritère à l'analyse des items. L'approche multicritère, en classant les items par rapport à leurs résultats sur un ensemble de critères de validation permet d'extraire un sous-ensemble suffisant d'items, lesquels sont puisés au sein des diverses sous-dimensions du concept mesuré: l'approche multicritère tend ainsi à conserver la variété théorique des sous-dimensions composant le concept.

Opérer une réduction de l'ensemble des items à partir du critère unique de fiabilité revient à privilégier un sous-ensemble d'items rattachés à une des sous-dimensions ou aux sous-dimensions les plus similaires du concept, et dont les mesures enregistrent, par rapport à tout sous-ensemble d'items de même cardinal, la plus grande redondance. L'intérêt de l'approche mul-

ticritère est de réaliser un compromis, le meilleur eu égard à l'ensemble des critères utilisés pour un ensemble donné d'items candidats, entre la fiabilité, les critères de validité, la variété des sous-dimensions et le nombre d'items. En d'autres termes, l'approche multicritère permet d'isoler un ensemble réduit d'items qui, pour un niveau satisfaisant de fiabilité, offre de plus grandes chances de capter dans sa globalité théorique le concept mesuré.

BIBLIOGRAPHIE

- BERTIER, P., ROY, B., "La méthode ELECTRE II, une application au media-planning", VI Conférence internationale de recherche opérationnelle, Dublin, août 1972, M. Ross Editor, OR 72, North Holland Publishing Company.
- BORHNSTEDT, G.W., "Reliability and Validity Assessment in Attitude Measurement", in SUMMERS, G.F., ed., Attitude Measurement, Chicago, Rand and McNally, 1969, 80-99.
- CHENEY, P.H., "Organizational Characteristics and Information Systems: an Investigation", Ph.D. Thesis, University of Minnesota, 1977.
- CRONBACK, L.J., "Coefficient Alpha and the Internal Structure of Tests", Psychometrika, 1951, 297-334.
- CURETON, E.E., "Corrected Item-Test Correlations", Psychometrika, 1966, 91-96.
- EDWARDS, A.L., Techniques of Attitude Scale Construction, New York, Appleton-Century-Crofts, 1957.

- FERRAND, D.J., "Les logiques d'interprétation de la rationalisation du système d'information organisationnel", Ph.D. Thesis, Université Laval, 1982.
- HELMER, O., The Use of the Delphi Technique, Rand Corp. Santa Monica, 1966.
- JENKINS, A.M., "An Investigation of Some Management Information Systems Design Variables and Decision-Making Performance: a simulation experiment", Ph.D. Thesis, University of Minnesota, 1977.
- POITOU, J.P., La dynamique des groupes: une idéologie au travail, Conseil national de la recherche scientifique, 1978.
- ROY, B., "Critères multiples et modélisation des préférences", Revue d'économie politique, janvier-février 1974.
- SHUKLA, R.K., "A Contingency Model of Influence Dispersion and the Bases of Power in Organizational Decision-Making", Ph.D. Thesis, University of Wisconsin-Madison, 1977.
- SPEARMAN, C., "Correlation Calculated with Faulty Data". British Journal of Psychology, 1910, 271-295.