

because they not looking for work, or not being willing to admit that they are claimants because they are working.

The figures for the period 1984 to 1992 are for the Spring quarter (March to May). For the period spring 1992 to the summer of 1997, the figures cover all four quarters, and have not been seasonally adjusted.

-----

## APPLICATION DE LA THEORIE DE REPONSE AUX ITEMS A L'ANALYSE D'ECHELLES D'ATTITUDE

par

Claire Durand,

(durand@ere.umontreal.ca; Département de sociologie,  
Université de Montréal, C.P. 6128, succ. Centre-Ville,  
Montréal, Québec H3C 3J7, Canada)

**Abstract.** Application of Item Response Theory to the Analysis of Attitude Scales. The article presents Item Response Theory (IRT) and illustrates its application to the analysis of two measures of work-related attitudes -- value given to social aspects of the job and satisfaction with extrinsic aspects of the job. Results of IRT analysis show that extrinsic satisfaction cannot be measured in the same way among contract professionals as opposed to regular employees. Using information provided by IRT analysis, it appeared that satisfaction with salary cannot be included in a scale measuring extrinsic satisfaction because it does not contribute to the measure in the same way for the two groups of employees; furthermore, satisfaction with social benefits may be included in the scale only if discrimination and difficulty parameters are allowed to vary across groups. Analyses of variance using theta scores estimated in this way give statistical conclusions that are notably different from those obtained using traditional additive scores and permit refinement of interpretation. **Item Response Theory, Attitude Scales, Work-Related Attitudes, Contract Professionals, Regular Employees.**

**Résumé.** Cet article présente la théorie de réponse aux items (TRI) et illustre ce que son application à l'analyse de deux mesures d'attitudes face au travail -- valorisation des aspects sociaux et satisfaction extrinsèque -- peut apporter. Les résultats présentés montrent que l'on ne peut mesurer de la même manière la satisfaction extrinsèque selon qu'il s'agit d'Employés professionnels à statut régulier ou occasionnel. Les informations produites par l'analyse TRI ont permis de conclure que la satisfaction face au salaire ne peut être incluse dans la mesure de la satisfaction extrinsèque parce qu'elle ne contribue pas à la mesure de la même manière pour les deux groupes d'Employés; de plus, la satisfaction face aux avantages sociaux peut être incluse dans la mesure seulement si l'on permet que les paramètres de discrimination et de difficulté varient selon les groupes de statut. Les analyses de variance sur les scores theta ainsi produits donnent des résultats sensiblement différents de ceux obtenus avec les scores additifs traditionnels et permettent de raffiner l'interprétation. **Théorie de réponse aux items, Echelles d'attitudes, Attitudes face au travail, Contractuels professionnels, Employés réguliers.**

En sciences sociales, et particulièrement en psychosociologie, les attitudes jouent un rôle théorique important. On propose généralement que les relations entre les déterminants sociaux -- âge, sexe, origine sociale, revenu, scolarité, etc -- et les comportements passent par les attitudes. On entend par attitudes l'ensemble des

réactions affectives, cognitives et conatives -- satisfaction, perceptions, intentions de comportement -- de même que l'ensemble des valeurs -- sociales, de vie, de travail -- c'est-à-dire les attitudes proactives. Pour recueillir des données sur le positionnement des individus sur ces mesures, l'instrument de collecte de données le plus fréquemment utilisé est le questionnaire de sondage, questionnaire qui peut avoir été devisé spécifiquement pour une enquête ou être de type standardisé (plus fréquent en psychologie). On recueille ainsi des informations sur le degré de satisfaction, d'accord, d'importance accordée en rapport avec un certain nombre d'éléments au moyen d'échelles de Likert unipolaires ou bi-polaires.

Dès cette étape, on attribue des codes arbitraires aux réponses. La correspondance entre ces codes et la mesure réelle des attitudes apparaît faible. L'échelle est ordinale mais la progression des codes est métrique. Ces échelles présentent des difficultés particulières au plan de l'analyse: le moyen le plus simple d'obtenir des mesures synthétiques de ce type de données consiste à les traiter comme si elles étaient de type métrique et de leur appliquer les analyses disponibles pour ce type de mesure, analyses qui sont généralement plus variées, plus développées et plus accessibles que les analyses appropriées aux échelles ordinales. Ce type de traitement a été décrit puisqu'il ne tient pas compte de la nature des données.

Dans la tradition psychométrique, la création d'échelles par addition des réponses aux items similaires et le recours aux indices de fidélité ( $\alpha$  de Cronbach), et plus récemment aux analyses factorielles, pour valider les regroupements est la norme. Cette manière de procéder permet la plupart du temps d'en arriver à des échelles dont la distribution est normale, ce qui les rend adéquates pour des analyses postulant une telle distribution. Toutefois, une partie de l'information se perd et plusieurs postulats sont faits, entre autres quant à une contribution égale des items (questions) à l'échelle et quant à une distance similaire entre les catégories de réponse. Ces postulats sont rarement vérifiés et souvent non appropriés.

Les scores additifs habituellement créés pour constituer des échelles d'attitudes comportent en effet des problèmes:

- a) Ils donnent un poids égal à chacune des variables composant la mesure;
- b) Ils postulent une distance égale entre les différentes catégories de réponse à chaque élément et une distance similaire quels que soient les éléments mesurés;
- c) Ils ne tiennent pas compte des patrons de réponse et accordent la même valeur au patron 2,2,2,2 (score additif de 8) propres aux personnes qui seraient, par exemple, peu satisfaites de chacun des quatre éléments sondés et au patron 2,2,3,1 (score=8) propres aux

personnes peu satisfaites des deux premiers éléments, assez satisfaites du troisième et totalement insatisfaites du quatrième;

- d) Ils ne permettent pas de tenir compte ou même de vérifier qu'un item peut ne pas avoir la même signification ou la même importance dans des populations différentes (entre autres lorsqu'il y a traduction);
- e) Ils sont liés à l'échantillon d'où il sont tirés et ils ne permettent pas de comparer entre elles des populations auprès desquelles le même concept aurait été mesuré de manière différente.

Puisque nous avons en main des échantillons de mesures de concepts similaires ou même identiques et des échantillons de populations, ne serait-il pas possible de ramener le tout sur une même métrique, ce qui permettrait de comparer les échantillons et les mesures et de faire un pas vers la cumulativité de données disparates? On est encore loin de la solution mais la théorie de réponse aux items permet de faire un pas dans cette direction. Dans cet article, nous présentons une analyse portant sur des échelles mesurant des attitudes face au travail. Les données ont été recueillies au moyen d'un questionnaire de sondage auprès de professionnels salariés de la fonction publique. L'analyse utilise la Théorie de réponse aux items (TRI) et les résultats obtenus en ayant recours à cette méthode sont comparés à ceux qui seraient obtenus par des méthodes classiques.

#### LA THEORIE DE REPONSE AUX ITEMS

La Théorie de réponse aux items (TRI) s'intéresse à la relation probabiliste entre la manière dont un individu répond à une question mesurant une compétence, une aptitude, une attitude d'une part, et une variable latente continue mesurant cette compétence (Thissen, 1991). Il s'agit d'une collection de modèles mathématiques simples décrivant les processus utilisés par les gens lorsqu'ils répondent à des questions (items) et les procédures statistiques utilisées pour appliquer ces modèles à l'analyse de données constituées de leurs réponses. On cherche à expliquer l'association observée dans les tableaux de contingence obtenus par le tri croisé des réponses aux items (Thissen et Mooney, 1989: 299-300).

Tout comme la théorie classique des tests, la TRI postule l'existence d'un concept non mesurable directement, appelé le trait latent, dont on tente d'estimer la valeur. Elle permet à la fois de convertir les réponses aux items en un estimé du trait latent ( $\theta$ ) et d'examiner les propriétés mêmes des items (Thissen, 1991:4-2). Elle permet une plus grande richesse d'interprétation au niveau de l'item ainsi

que l'identification de la non-équivalence de contribution des Items à la mesure de certains concepts pour des populations différentes.

On estime donc que la probabilité d'une réponse donnée à une question dépend de deux éléments, la compétence de l'individu et les caractéristiques de la question. Non seulement les estimés des paramètres des Items doivent-ils être indépendants de l'échantillon d'où ils sont tirés mais l'estimé de la compétence - du positionnement sur le trait latent -- de chaque personne devrait idéalement être indépendant de l'échantillon de questions posées.

### Les paramètres et les modèles

La relation entre la compétence et la réponse à une question peut être prédite par une fonction décrite par la courbe caractéristique de l'Item (ICC) définie par:

$$P_i(\theta) = c + (1-c) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad (1)$$

où  $\theta$  est la valeur réelle de la compétence d'un individu. La probabilité d'une réponse déterminée à la question  $i$  chez un individu donné possédant une attitude générale  $\theta$  est donc définie par une fonction logistique dépendant de trois paramètres:

$a$  est l'indice de discrimination et constitue la pente de la courbe caractéristique de l'Item (ICC) au point  $b$ . Plus le paramètre est élevé, plus la pente est abrupte, et plus il est facile de différencier entre eux les individus à des niveaux différents du concept que l'on veut mesurer. L'indice de discrimination est affecté par la variance de l'Item.

$b$  est l'indice de difficulté. Il définit l'endroit sur le continuum de  $\theta$  où la probabilité d'une réponse déterminée est de 50% ou plus. Cet indice relie l'Item au score individuel sur l'échelle standardisée. Il est affecté par la moyenne de l'Item. Plus la valeur du paramètre  $b$  est élevée, plus l'individu doit être "compétent", c'est-à-dire qu'il est satisfait d'une dimension, qu'il la valorise, pour avoir une probabilité supérieure à 50% de répondre ainsi à la question posée.

$c$  est le paramètre de chance (guessing). Il définit la probabilité d'une bonne réponse obtenue au hasard. Il est généralement fixé à zéro sauf pour les tests à choix multiples. L'équivalent pour la mesure

des attitudes serait de considérer ce paramètre comme une mesure de l'acquiescement ou de la désirabilité sociale, c'est-à-dire de la tendance à répondre positivement ou à donner une bonne image de soi (Ellis et Kimmel, 1992).

L'équation (1) définit le modèle de base de la TRI; plusieurs modèles ont été élaborés selon que certains postulats sont faits. Les modèles à deux paramètres (2PL, Samejima, partial credit) postulent que le paramètre  $c$  est égal à zéro, c'est-à-dire qu'il n'y a pas de réponse au hasard. Les modèles à un paramètre (1PL, Rasch) ne retiennent que le paramètre  $b$  et postulent donc que la discrimination associée à chacun des Items est égale. (voir détails dans De Ayala, 1993; Hambleton et coll., 1991).

Le modèle utilisé ici est celui proposé par Samejima (1969 in Hambleton et coll., 1991) pour les réponses graduées. Il s'agit d'un modèle à deux paramètres où on estime la probabilité qu'un sujet réponde à un certain niveau de l'échelle de Likert ou plus pour une question donnée selon son positionnement sur l'attitude, le concept général mesuré. Plus concrètement, pour une question de satisfaction ayant quatre catégories de réponse -- codées 1 = pas du tout, 2 = un peu, 3 = assez et 4 = très -- on retrouve un paramètre de discrimination "a" et trois paramètres de "difficulté":  $b_1$  définit le point sur l'échelle  $\theta$  où la probabilité qu'un individu réponde 2 ou plus passe au-delà de 50% et donc la probabilité qu'un individu soit un peu satisfait ou plus d'un Item étant donné son niveau général de satisfaction sur le concept mesuré. De même,  $b_2$  définit le point sur l'échelle  $\theta$  où la probabilité qu'un individu réponde 3 -- soit "assez satisfait" -- ou plus passe au-delà de 50%, et ainsi de suite jusqu'à concurrence du nombre de catégories moins un.

La probabilité que la réponse d'un individu se retrouve dans une catégorie déterminée est alors facile à déduire puisqu'elle est égale à la probabilité qu'il réponde dans une catégorie ou plus à laquelle on soustrait la probabilité de répondre à la catégorie suivante. Dans le modèle de Samejima, il y a donc autant de courbes d'Item qu'il y a de catégories moins une. Ces courbes ont toutes la même pente mais ne sont pas situées au même endroit sur l'échelle  $\theta$ .

### Les postulats

La Théorie de réponse aux Items postule l'unidimensionnalité du concept mesuré et l'indépendance locale, ce qui suppose que les échelles constituées mesurent une seule dimension et qu'à un niveau constant du concept que l'on veut mesurer, les réponses aux différents Items sont indépendantes. Cette propriété permet de

postuler que la probabilité d'une combinaison, d'un "patron" de réponses, est égale au produit des probabilités de chacune des réponses. Ainsi la probabilité d'être très satisfait de trois éléments composant la satisfaction intrinsèque, par exemple, sera égale au produit de la probabilité d'être très satisfait de chacun des éléments.

L'adéquation du modèle suppose que les données, la proportion de réponses dans une catégorie particulière d'une question, ne s'écartent pas de façon significative des prédictions découlant de la fonction de réponse à l'item décrite par la courbe caractéristique de l'item (ICC). La Théorie de réponse aux items est relativement rigide et ne s'applique pas facilement aux types de données présentes en sciences sociales et plus particulièrement à la mesure des attitudes. L'adéquation du modèle et des postulats est vérifiée par une statistique, le  $G^2$ , qui se distribue comme le  $\chi^2$  avec comme degrés de liberté le nombre de patrons de réponse différents moins le nombre de paramètres à estimer.

#### L'invariance des estimés

Une des propriétés les plus intéressantes de la TRI réside dans l'invariance des estimés des paramètres lorsque les échantillons sont tirés de la même population. Ainsi, les mêmes fonctions de réponse à l'item sont obtenues, quelle que soit la distribution du trait ( $\theta$ ) des répondants à partir des réponses desquels les paramètres des items sont estimés (Hambleton et coll., 1991). Ceci entraîne que l'on peut vérifier cette prétention, c'est-à-dire examiner si les estimés des paramètres pour deux échantillons sont significativement différents, auquel cas on parlera de biais de l'item. Ce biais peut survenir pour l'estimé du paramètre de discrimination -- un item discrimine mieux les répondants d'un groupe que ceux d'un autre groupe -- ou pour l'estimé des paramètres de difficulté. On dira d'un item qu'il est biaisé si des individus ayant la même compétence, le même degré de satisfaction sur une dimension, n'ont pas la même probabilité de répondre d'une certaine manière selon qu'ils soient membres d'un groupe social plutôt qu'un autre (Hambleton et coll., 1991).

Les études sur cette question sont fréquentes dans la recherche interculturelle portant sur les valeurs sociales où l'on tente de vérifier si les formulations n'entraînent pas des biais selon les langues et les cultures. La non équivalence dans les paramètres de discrimination "a" est alors interprétée comme relevant de différences ethniques alors que la non équivalence dans les estimés de paramètres de difficulté "b<sub>j</sub>" reflèterait des problèmes de traduction (Bontempo, 1993). Les biais hommes-femmes dans la

mesure de la dépression sont aussi étudiés (voir entre autres Ramsay, 1991, 1995; Schaeffer, 1988). On se demande alors si le fait de pleurer est un comportement aussi symptomatique de la dépression dans les deux groupes. Ne pas tenir compte du biais d'item amènerait soit à déclarer les femmes dépressives trop facilement, soit à ne pas détecter les hommes dépressifs. On voit donc aisément les conséquences du biais d'item dans les approches cliniques. Elles sont aussi importantes dans les épreuves de sélection scolaires ou au travail où le fait d'appliquer les mêmes critères à tous les groupes peut entraîner un préjugé favorable aux membres du groupe social majoritaire.

#### L'estimation des paramètres

Plusieurs méthodes d'estimation des paramètres existent. On parle surtout du maximum de vraisemblance combiné, du maximum de vraisemblance marginal, de méthode bayésienne et de méthode d'approximation. La méthode de maximum de vraisemblance marginale (MML) apparaît comme la meilleure dans les simulations d'estimation effectuées (Reise et Yu, 1990; Stone, 1992; Ellis et Kimmel, 1992). Toutefois, quelle que soit la méthode, plus les valeurs réelles sont extrêmes, plus les biais dans les estimés apparaissent importants. Ces biais varient en fonction de la distribution réelle de  $\theta$ , du nombre d'items et de la taille de l'échantillon.

#### L'information, la fidélité

La notion d'information est à la TRI ce que la notion de fidélité est à la théorie classique des tests. En effet, la fonction d'information donne un indice du niveau de précision de l'estimé selon le niveau  $\theta$  de compétence de l'individu.

Contrairement au concept de fidélité, propre à l'analyse classique, où l'on suppose que la marge d'erreur est la même, peu importe le niveau du concept mesuré, en TRI, la quantité d'information donnée varie en fonction des paramètres d'items et du paramètre  $\theta$  et varie donc tout au long de l'échelle  $\theta$ . Plus la valeur du paramètre a est grande, plus l'information fournie par l'item est importante; plus la valeur du paramètre b<sub>j</sub> se rapproche de la valeur  $\theta$  d'un individu, plus l'item est informatif pour l'individu. L'information est additive, c'est-à-dire que la fonction d'information d'un test est la somme des fonctions d'information des items composant le test. Elle permet d'estimer la fidélité marginale, c'est-à-dire la moyenne de la fidélité

sur l'ensemble de la distribution de  $\theta$  lorsque l'information du test est relativement uniforme tout au long de l'échelle.

## RESULTATS D'ANALYSE POUR DEUX VARIABLES D'ATTITUDE

Au-delà de la simple illustration, la présentation qui suit tente de répondre aux questions suivantes: Quelles informations ce type d'analyse fournit-il qui ne seraient pas fournies par une analyse plus traditionnelle? Est-ce que les conclusions statistiques quant aux différences entre des groupes varient selon le type de scores d'échelles utilisés? L'analyse porte sur un domaine peu étudié, l'équivalence dans la mesure des attitudes au travail chez les Employés à statut précaire comparé à ceux à statut permanent. Dans ce cas, il s'agit de professionnels du secteur public québécois.

### Méthodologie

*Population.* La population est composée de professionnels salariés travaillant pour la fonction publique québécoise (N=12,000). L'échantillon a été stratifié pour représenter les deux catégories de statut d'emploi, les Employés réguliers qui bénéficient d'une sécurité d'emploi totale et les Employés occasionnels qui travaillent selon des contrats de six mois à trois ans, renouvelables. Les Employés réguliers ont accès à des avantages sociaux, du perfectionnement, des vacances annuelles, etc., auxquels n'ont pas droit les Employés occasionnels. Par contre, ces derniers bénéficient d'avantages salariaux pour compenser cette situation. Les Employés occasionnels sont généralement plus jeunes: 74% avaient moins de 35 ans au moment de l'enquête (1988) alors que chez les professionnels réguliers, seuls 39% avaient moins de 35 ans (voir Durand, 1993, 1996a, 1996b).

*Analyse.* L'analyse des mesures d'attitude face au travail se fait en fonction de deux variables susceptibles d'influencer la manière dont les concepts sont mesurés, le statut d'emploi - catégorisé par l'emploi régulier ou occasionnel - et le groupe d'âge -- moins de 35 ans, 35 à 64 ans. Le groupe d'âge est utilisé comme variable contrôlée de façon à départager les effets de statut de ceux qui peuvent être attribués à l'âge.

La première étape de l'analyse a consisté à utiliser l'analyse factorielle comme outil d'exploration de façon à obtenir des regroupements qui seraient constants théoriquement et statistiquement. Cette étape a permis de retenir 32 variables

d'attitudes se regroupant en 10 dimensions, relatives à la satisfaction et aux perceptions face aux aspects extrinsèques (2), intrinsèques (3) et sociaux (1) de l'emploi ainsi qu'aux valeurs de travail, c'est-à-dire à l'importance attribuée aux aspects extrinsèques (2) de même qu'intrinsèque (1) et social (1). Les analyses de fidélité ( $\alpha$  de Cronbach) montrent des indices qui varient de .52 à .82; les indices les plus faibles ont trait aux mesures des valeurs, qui souffrent traditionnellement d'une variance restreinte.

Suite à ces analyses, la deuxième étape a consisté à effectuer des analyses d'items utilisant la TRI pour chacune des dimensions<sup>1</sup>. Ces analyses visaient d'abord à examiner si les mesures étaient équivalentes selon les groupes, c'est-à-dire si, par exemple, la satisfaction face au salaire est une composante de la satisfaction extrinsèque de la même manière chez les plus jeunes et chez les plus âgés, chez les Employés réguliers et occasionnels.

La troisième étape consistait, après avoir obtenu des mesures équivalentes en retirant des items si nécessaire, à récupérer les scores d'échelles pour des analyses subséquentes. Des analyses de variance comparatives des scores  $\theta$  et des scores additifs traditionnels ont ensuite été effectuées.

Cette article présente les résultats des analyses pour deux dimensions. La première est la valorisation sociale, incluant l'importance accordée aux relations interpersonnelles et à la reconnaissance du travail. Une deuxième dimension est celle de la satisfaction face aux aspects extrinsèques -- salaire, avantages sociaux, horaire de travail.

### Résultats

Les tableaux 1 et 2 montrent les estimés des divers paramètres d'items et d'adéquation du modèle selon que l'estimation est faite pour chaque groupe séparément ou pour les groupes conjointement. Deux séries d'estimés sont produits, selon les groupes de statut et selon les groupes d'âge.

Les deux premières colonnes du tableau 1 montrent d'abord les estimés des paramètres d'items pour chacun des groupes de statut séparément alors que la troisième colonne donne les mêmes estimés lorsque l'on postule l'équivalence des estimés pour les deux groupes. Ces estimés sont assez similaires ou à tout le moins, ne diffèrent pas significativement comme l'indiquent les faibles différences dans le  $G^2$ , différences qui se distribuent de façon similaire au  $X^2$ . La valeur du  $G^2$  pour le modèle sans contrainte d'équivalence est de

133,8 pour 61 degrés de liberté (dl) alors qu'il est de 162,5 (73 dl) pour le modèle d'invariance, c'est-à-dire le modèle où les estimés des paramètres sont contraints à être égaux pour les deux groupes. La différence de 28,7 (12 dl) est considérée acceptable, sa valeur étant moins de quatre fois le nombre de degrés de liberté. La même situation est constatée pour l'analyse selon l'âge, la différence entre les modèles étant encore plus faible (21,7 pour 12 dl). Il est donc possible de ne pas rejeter l'hypothèse nulle relative à l'équivalence des mesures selon le groupe d'âge et selon le statut.

Les estimés des paramètres de discrimination "a" varient de 1,18 à 1,42 ce qui signifie que les items discriminent à peu près également et apportent donc une contribution similaire à la mesure de l'échelle. Cette contribution est acceptable puisque, à partir d'une valeur de 0,75, on considère que l'item discrimine suffisamment. Les paramètres  $b_j$  sont aussi similaires pour les divers items, ce qui confirme la contribution équivalente des items à la mesure du concept. Le paramètre  $b_1$  indique à quelle valeur de  $\theta$ , la probabilité d'accorder assez d'importance à l'item -- plutôt que peu ou pas du tout -- passe au delà de 50%. Une valeur de -3,87 ( $b_1$  pour support du supérieur) indique qu'il faut accorder très peu d'importance (près de quatre écart-types sous la moyenne) à la dimension de valorisation sociale pour déjà considérer que le support du supérieur est assez important. De telles valeurs des paramètres de difficulté illustrent les problèmes liés à la variance habituellement très restreinte des mesures des valeurs en psychosociologie, encore plus des valeurs de travail intrinsèques ou sociales.

Les deux dernières colonnes du tableau 1 montrent que les estimés sont relativement stables selon qu'ils sont produits à partir des patrons de réponse des groupes composés par le croisement de l'âge et du statut (quatre groupes) ou à partir de l'ensemble des patrons (tous). Les valeurs de  $G^2$  sont tout à fait acceptables, à peine le double des degrés de liberté.

Le tableau 1 présente aussi la valeur de la moyenne des groupes en fixant la distribution des valeurs du dernier groupe (groupe de référence) comme de type N (0,1). Ces valeurs sont pertinentes uniquement lorsque le modèle d'invariance est, au moins partiellement, supporté. Les estimés présentés tendent à montrer qu'il n'y aurait pas de différence significative estimée entre les groupes selon le statut ( $\mu_1 = -.09$ , erreur-type = .05) mais une différence significative selon l'âge ( $\mu_1 = .39$ , erreur-type = .05), les jeunes valorisant plus les aspects sociaux de l'emploi que les plus âgés. Ces estimés varient sensiblement pour l'analyse à partir des quatre groupes, ce qui illustre l'instabilité des estimés lorsque la taille des groupes est réduite. Par ailleurs, l'information est

relativement constante tout le long de l'échelle  $\theta$ , ce qui permet de considérer que l'estimé de fidélité moyen donne une information valide. Cet estimé varie de 0,51 à 0,55 selon les quatre sous-groupes composés par le croisement de l'âge et du statut. La fidélité est donc relativement faible ce qui permettrait de conclure, étant donnée la valeur du  $G^2$ , qu'il s'agit bien d'une échelle unidimensionnelle mais que les items ne sont pas de très bons indicateurs de la dimension.

Le tableau 2 montre un portrait très différent quant à l'équivalence des mesures. D'une part, la valeur du  $G^2$  passe de 262,7 à 436,2 (une différence de 173,5 pour 12 degrés de liberté) lorsque l'on postule l'équivalence des paramètres d'item selon les groupes de statut. La différence est moins importante lorsque l'on postule l'équivalence selon les groupes d'âge (77,5) mais tout de même significative. On doit donc rejeter l'hypothèse d'une équivalence selon le statut et selon l'âge. A première vue, les différences entre les paramètres semblent plus importantes pour le paramètre de discrimination quant à la satisfaction face aux avantages sociaux: la valeur de  $a_2$  est de 4,05 pour les réguliers et 2,48 pour les occasionnels, de 2,29 pour les jeunes et 4,14 pour les vieux. Ces différences semblent importantes mais elles peuvent être dues à des artefacts de la procédure d'estimation, les valeurs de plus de 4,0 pour le paramètre de discrimination étant exceptionnelles. Toutefois, les paramètres de difficulté  $b_{2j}$  pour ce même item sont similaires selon les groupes alors qu'ils sont passablement différents pour l'item relatif à la satisfaction quant au salaire. De façon à déterminer statistiquement le modèle le plus approprié, la méthode décrite par Reise, Widaman et Pugh (1993) est utilisée. Il s'agit de fixer les paramètres de chaque item comme non équivalents à tour de rôle et de voir par la différence du  $G^2$  quel modèle est le plus approprié, le cas échéant.

Le tableau 2 présente les estimés obtenus lorsque l'on permet la non-équivalence des paramètres pour les items relatifs au salaire et aux avantages sociaux. La valeur de la différence de  $G^2$  (8 dl) comparé au modèle où tous les estimés peuvent varier - est de 35,9 lorsque les paramètres pour la satisfaction face au salaire varient et de 67,8 lorsque ceux de la satisfaction face aux avantages sociaux varient. Elle est de 164,8 lorsque seuls les paramètres pour la satisfaction face à l'horaire de travail varient. Il faut noter que l'estimé de la moyenne change de façon importante selon l'item que l'on laisse varier, changeant même de signe (de +0,10 à -0,38). Les mêmes analyses pour l'âge donnent une situation similaire, la différence de  $G^2$  la plus faible survenant lorsque l'on permet la variation des paramètres pour le salaire (22,1 pour 8 dl). Ces analyses amènent à conclure qu'il y a non équivalence entre les groupes de statut quelque soit l'item que l'on laisse varier et qu'il y a non équivalence pour l'item de satisfaction face au salaire selon les

OK ~~Je suis sûr~~ ?

groupes d'âge. La satisfaction face au salaire ne contribue donc pas de la même manière à la mesure de la satisfaction extrinsèque selon les deux caractéristiques. S'il y avait non-équivalence selon une seule des caractéristiques ou s'il était possible de tenir compte de la non-équivalence pour les quatre groupes formés par les deux variables (statut-âge), il serait possible d'estimer les scores  $\theta$ . Comme ceci n'est pas possible, il est nécessaire de refaire les analyses en supprimant l'item sur le salaire. Les résultats de ces analyses sont présentées au tableau 3.

Lorsque l'on estime l'équivalence des mesures pour une échelle de satisfaction extrinsèque comprenant seulement les items relatifs aux avantages sociaux et à l'horaire de travail, on note une équivalence des mesures selon l'âge, la différence de  $G^2$  étant de 20,0 pour 8 degrés de liberté. En ce qui concerne le statut, la situation est différente: la différence de  $G^2$  est de 40,4 pour 8 degrés de liberté. L'hypothèse d'une équivalence selon le statut doit donc être rejetée. Lorsque l'on permet la non-équivalence pour les avantages sociaux, la différence de  $G^2$  baisse à 6,1 pour 4 dl, confirmant la justesse de ce modèle. La non équivalence pour un item sur deux n'est pas une situation idéale. En pratique, on chercherait à ce qu'une majorité d'items d'une échelle soient équivalents. Nous poursuivons toutefois l'analyse à tout le moins pour fins d'illustration. Notons que l'indice de fidélité marginal est à ,69, ce qui est acceptable.

L'étape suivante consiste à estimer les scores  $\theta$  pour les deux échelles, celle portant sur la valorisation de l'intégration et celle portant sur la satisfaction extrinsèque. Dans ce dernier cas, les estimés de scores prendront en compte des paramètres de discrimination et de difficulté différents pour les deux groupes de statut. Les estimés de scores utilisés sont les EAP (expected a posteriori); c'est-à-dire, la moyenne de l'estimé.

Le tableau 4 montre les résultats des analyses de variance pour les scores additifs et pour les scores  $\theta$ . Pour ce qui est de la valorisation sociale, il n'y a pas beaucoup de différence entre les résultats de l'analyse avec les deux types de scores. La valeur des tests F est plus basse pour les scores  $\theta$  mais d'autres analyses montrent que la situation inverse peut tout aussi bien se produire.

Par contre, l'intérêt de la TRI apparaît plus évident pour ce qui est de la satisfaction extrinsèque. La première ligne donne les résultats pour le score additif tel qu'il aurait été constitué à partir des renseignements de base utilisés en s'inspirant de la théorie classique des tests. Il faut toutefois noter que l'indice de fidélité (de Cronbach) est relativement bas pour cette échelle (0,5827). Il est toutefois similaire pour les deux groupes d'âge et de statut. Suite à l'information fournie par la TRI quant à la non-équivalence de

contribution de la satisfaction face au salaire, les analyses traditionnelles ont été faites pour la satisfaction face au salaire et pour le reste de l'échelle séparément. Ces analyses montrent que le fait de grouper ensemble les trois indicateurs obscurcissait la relation entre l'âge, le statut et la satisfaction. Les effets principaux de l'âge et du statut sur la satisfaction semblent uniquement dus au salaire et non au positionnement sur le reste de l'échelle. Comme il y a présence d'interaction (marginale significative toutefois pour le score additif sans la satisfaction face au salaire -  $F=3,92$ ,  $p=.048$ ), il est nécessaire de procéder à une analyse des effets simples.

L'analyse des effets simples, présentée à la suite dans le tableau montre la situation suivante: D'une part, chez les jeunes, il apparaît qu'il y a une différence de satisfaction, en fonction du statut d'emploi, face au salaire et face aux autres aspects extrinsèques, différence qui n'apparaissait pas avec le score additif global. Cette situation s'explique par le fait que les moyennes montrent que les jeunes occasionnels sont plus satisfaits du salaire mais moins satisfaits des autres aspects extrinsèques que les jeunes ayant un emploi régulier. Ainsi, les effets s'annulent lorsque l'on utilise le score global. Par ailleurs, lorsqu'on utilise le score  $\theta$ , la relation passe d'une probabilité marginale de ,045 à une probabilité de ,011. Les moyennes sont de 6,27 pour les réguliers et 6,09 pour les occasionnels pour les scores additifs alors qu'elles sont de 0,14 comparé à -0,01 pour les scores  $\theta$ . Cette situation peut s'expliquer par le fait que l'échelle de satisfaction extrinsèque sans l'item salaire apparaît comme souffrant d'hétérogénéité des variances (Bartlett-Box F de 6,71,  $p=.000$ ) alors que les scores  $\theta$  ne souffrent pas de ce problème (Bartlett-Box F de 0,26,  $p=.857$ ). Chez les plus vieux, contrairement à ce que laissent croire les analyses sur le score global, c'est uniquement la satisfaction face au salaire qui distinguent les réguliers et les occasionnels, ces derniers étant plus satisfaits (moyenne de 2,97 versus 2,50). Il y a peu de différence entre les résultats des scores  $\theta$  et ceux des scores additifs. Les deux dernières colonnes du tableau 4 montrent les valeurs du test de Bartlett pour l'homogénéité des variances. Les scores  $\theta$  présentent généralement une variance plus homogène entre les groupes, ce qui est une des conditions de l'utilisation du test F. Les valeurs de F pour les scores  $\theta$  devraient donc être d'autant plus valides.

Il faut noter que ces différences existent en présence d'une corrélation de ,976 entre les scores  $\theta$  et additifs pour la satisfaction extrinsèque et de ,992 pour la valorisation sociale. La figure 1. montre graphiquement la relation entre les scores  $\theta$  et additifs selon le groupe de statut. Les figures montrent que la dispersion des scores  $\theta$  en fonction des scores additifs est plus substantielle pour l'échelle de satisfaction extrinsèque où les scores  $\theta$  tiennent compte

du fait qu'un item ne contribue pas à l'échelle de la même façon pour les deux groupes. Cette situation entraîne que le même patron de réponse ne correspond pas au même score  $\theta$  selon le groupe de statut d'appartenance.

## CONCLUSION

Les analyses basées sur la théorie de réponse aux items permettent de raffiner l'interprétation des données. Bien que l'on puisse soutenir que beaucoup de l'information obtenue aurait pu être découverte en ayant recours à des analyses classiques plus sophistiquées -- analyse des structures factorielles et analyses de fidélité pour chacun des sous-groupes, il demeure que certaines informations ne seraient pas ressorties aussi aisément et avec autant de précision. Les biais d'items sont rapidement repérables avec les analyses TRI alors qu'ils n'apparaissent même pas avec les analyses de fidélité. Les scores produits, dont on peut toutefois objecter qu'ils ne sont que des estimés, sont généralement plus homogènes et respectent donc plus aisément les postulats du recours à des analyses basées sur le modèle linéaire. Enfin, le modèle logistique correspond sans doute mieux à certaines données d'attitude puisqu'il permet de respecter la nature ordinale et la distribution souvent non normale de ce type de données. Par ailleurs, les indices d'adéquation du modèle permettent d'estimer de façon relativement précise si le modèle logistique correspond aux données.

Les désavantages surviennent au plan de la taille des échantillons nécessaires, de la disponibilité et de la convivialité des logiciels disponibles ou même des ressources spécialisées disponibles pour utiliser les logiciels de façon appropriée et efficace. Le jeu en vaut-il la chandelle? L'avantage le plus important des analyses basées sur la théorie du trait latent réside dans les propriétés des modèles, dont l'invariance des estimés selon les échantillons, qui permettent éventuellement de ramener sur une même métrique des mesures partiellement différentes des mêmes concepts. Cette avancée permettra de faire un pas important vers l'accumulation des données en sciences sociales puisque les données déjà amassées pourront être comparées entre elles avec une assurance nettement plus importante quant à l'équivalence des mesures.

Si ce type d'analyse se répand, il est fort probable que nous assisterons à la production de logiciels plus accessibles et peut-être même à l'intégration de ces possibilités d'analyse dans les logiciels tout usage plus courants de type SPSS, Systat, etc. Il demeurera que la théorie elle-même et ses particularités doit être maîtrisée

avant de se lancer dans les analyses et que les recherches doivent continuer pour bien connaître les conditions nécessaires à une bonne stabilité des estimés et éventuellement à l'estimation de modèles multidimensionnels, phase entreprise pour les données dichotomiques.

## NOTES

- (1) Le logiciel utilisé est Multilog (Scientific Software) de D. Thissen, le seul qui permettait ce type d'analyse.
- (2) Pour les valeurs, les catégories "pas du tout" et "peu important" ont été combinées étant donné le faible nombre de répondants dans la catégorie "pas du tout".

## REFERENCES

- Bartholomew, D. J.; Schuessler, K. F. (1991). Reliability of Attitude Scores Based on a Latent Trait Model. *Sociological Methodology*, 21, 97-123.
- Bontempo, R. (1993). Translation Fidelity of Psychological Scales: An Item Response Theory Analysis of an Individualism-Collectivism Scale. *Journal of Cross Cultural Psychology*, 24 (2), 149-166.
- De Ayala, R. J. (1993). An Introduction to Polytomous Item Response Theory Models. *Measurement and Evaluation in Counseling and Development*, 25 (4), 172-189.
- Durand, C. (1993). *L'aspiration à la mobilité en emploi chez les professionnels, diverses formes, divers déterminants*, Thèse présentée à la Faculté des études supérieures, Université de Montréal, Montréal, Canada.
- Durand, C. (1996a). Intérêt de la théorie de réponse aux items pour élaborer des échelles de mesure des attitudes. XXVIII<sup>ème</sup> Journées de statistique, Québec, Université Laval, 27-30 mai 1996.
- Durand, C. (1996b). Looking for Equivalence in the Measurement of Attitudes across Populations and Samples. Fourth International Conference on Social Science Methodology, Essex, England, July 1-4.



Ellis, B. B., & Kimmel, H. D. (1992). Identification of Unique Cultural Patterns par Means of Item Response Theory. *Journal of Applied Psychology*, 77 (2), 177-184.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement Issues and Practice*, 12 (3), 38-47.

Hambleton, R. K.; Swaminathan, H.; Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage.

Ramsay, J. O. (1991). Kernel Smoothing Approaches to Nonparametric Item Characteristic Curve Estimation. *Psychometrika*, 56, 611-630.

Ramsay, J. O. (1995). *Testgraf, A Program for the Graphical Analysis of Multiple Choice Test and Questionnaire Data*. Montreal: McGill University.

Reise, S. P.; Widaman, K. F.; Pugh, R. H. (1993). Confirmatory Factor Analysis and Item Response Theory: Two approaches for Exploring Measurement Invariance. *Psychological Bulletin*, 114 (3), 552-566.

Reise, Steven P; Yu, Jlayuan (1990). Parameter Recovery in the Graded Response Model Using MULTILOG. *Journal of Educational Measurement*, 27 (2), 133-144.

Schaeffer, N. C. (1988). An Application of Item Response Theory to the Measurement of Depression. *Sociological Methodology*, 18, 271-307.

Stone, C. A. (1992). Recovery of Marginal Maximum Likelihood Estimates in the Two-Parameter Logistic Response Model: An Evaluation of MULTILOG. *Applied Psychological Measurement*, 16 (1), 1-16.

Thissen, D. (1991). *MULTILOG User's Guide: Multiple, Categorical Item Analysis and Test Scoring Using Item Response Theory*. Chicago: Scientific Software, Inc.

Thissen, D., & Mooney, J. A. (1989). Loglinear Item Response Models, with Applications to Data from Social Surveys. *Sociological Methodology*, 19, 299-330.

Tableau 1. Valorisation sociale  
Équivalence des mesures

Importance accordée à ...	STATUT		ÂGE		4 grp.	TOUS	
	Tous estimés variant		Tous estimés variant				Estimés invariants
	Réguliers	Occasion.	34 ans -	35 ans +			
Relations avec employés	a 1,26 b <sub>1</sub> -2,48 b <sub>2</sub> 0,19	1,32 -3,34 -0,49	1,16 -2,81 0,13	1,40 -2,95 -0,38	1,23 -3,16 -0,35	1,28 -3,27 -0,54	
Support du supérieur	a 1,08 b <sub>1</sub> -3,32 b <sub>2</sub> 0,22	1,73 -3,44 -0,57	1,21 -3,81 0,12	1,23 -3,42 -0,38	1,24 -3,64 -0,34	1,26 -3,78 -0,54	
Reconnaissance du travail	a 1,51 b <sub>1</sub> -2,97 b <sub>2</sub> -0,05	1,22 -5,11 -0,94	1,42 -3,32 -0,09	1,50 -3,73 -0,72	1,36 -3,96 -0,66	1,45 -3,95 -0,83	
Accès au perfectionnement.	a 1,18 b <sub>1</sub> -1,93 b <sub>2</sub> 0,91	1,12 -3,68 -0,33	1,10 -2,94 0,56	1,12 -2,58 0,28	1,18 -2,72 0,20	1,14 -2,99 -0,01	
μ	0,75 (0,03)		0,85 (0,04)		0,39 (0,05)		
G <sup>2</sup> (Δ)	133,8	162,5 (28,7)	129,6	151,3(21,7)	250,0	96,9	
DL (patrons- paramètres)	86-25 = 61	73 (12)	90-25 = 65	77 (12)	140-15 = 125	54-12 =42	

Tableau 2. Satisfaction extrinsèque  
Equivalence des mesures

Satisfaction relative à...		STATUT						ÂGE			
		Tous estimés variant		Estimés invariants	satisf. face salaire varie		satisf. face avant. sociaux varie		Tous estimés variant		Estimés invariants
		réguliers	occasion.		Régul.	Occas.	Régul.	Occas.	34 ans -	35 ans +	
Satisfaction salaire	a	1,52	1,81	1,60	1,74	1,43	1,75	1,70	1,41	1,66	
	b <sub>1</sub>	-1,66	-2,46	-2,13	-1,62	-2,87	-2,19	-2,02	-1,88	-1,87	
	b <sub>2</sub>	-0,30	-1,48	-0,92	-0,39	-1,70	-1,02	-0,86	-0,50	-0,68	
	b <sub>3</sub>	2,20	1,10	1,53	1,90	1,31	1,35	1,70	2,02	1,73	
Satisf. avantages sociaux	a	4,05	2,48	2,81	3,65	3,94	2,80	2,29	4,14	2,61	
	b <sub>1</sub>	-2,15	-2,10	-2,39	-2,07	-2,78	-1,89	-1,99	-2,20	-2,21	
	b <sub>2</sub>	-1,23	-1,21	-1,50	-1,26	-1,82	-1,04	-1,17	-1,29	-1,30	
	b <sub>3</sub>	1,22	0,88	0,92	0,99	0,63	0,95	1,28	1,04	1,17	
Satisf. horaire de travail	a	1,06	1,07	1,07	1,03	0,94	1,15	1,07	1,03		
	b <sub>1</sub>	-3,42	-3,63	-3,72	-3,68	-4,23	-2,87	-3,97	-3,59		
	b <sub>2</sub>	-1,92	-1,92	-1,91	-2,04	-2,47	-1,49	-2,18	-1,96		
	b <sub>3</sub>	1,39	0,98	0,98	1,20	1,00	1,43	1,08	1,28		
μ (e.t.)		0,23 (0,05)		-0,15 (0,05)	+0,10 (0,05)	-0,38 (0,05)	0,36 (0,05)		0,29 (0,05)		
G <sup>2</sup> (Δ)		262,7		436,2 (173,5)	298,6 (35,9)		330,5 (67,8)		274,6		352,1 (77,5)
DL* (Δ)		95-25=70		82 (12)	86 (8)		86 (8)		97-25=72		84 (12)

\* nb patrons différents, c'est-à-dire de combinaisons de réponses aux trois items, auquel on soustrait le nombre de paramètres à estimer (Nb patrons - nb param.)

Tableau 3. Satisfaction extrinsèque sans le salaire  
Equivalence des mesures

		STATUT					ÂGE		
		Tous estimés variant		Estimés invariants	satisf. face avant. sociaux varie		Tous estimés variant		Estimés invariants
		réguliers	occasion		Régul.	Occas.	34 ans -	35 ans +	
Satisf. avantages sociaux	a	2,03	1,43	1,76	2,28	1,45	1,55	1,90	1,84
	b <sub>1</sub>	-2,68	-2,78	-2,82	-2,79	-2,75	-2,56	-2,94	-2,75
	b <sub>2</sub>	-1,45	-1,54	-1,65	-1,64	-1,52	-1,50	-1,61	-1,63
	b <sub>3</sub>	1,55	1,12	1,22	1,22	1,15	1,46	1,30	1,19
Satisf. horaire de travail	a	1,60	1,65	1,58	1,52	1,60	1,60	1,70	1,50
	b <sub>1</sub>	-2,44	-2,74	-2,76	-2,82	-2,29	-2,29	-2,91	-2,87
	b <sub>2</sub>	-1,33	-1,47	-1,58	-1,61	-1,19	-1,19	-1,66	-1,64
	b <sub>3</sub>	1,26	0,76	0,91	0,94	1,23	1,23	0,82	0,94
μ (e.t.)		0,34 (0,03)		0,06 (0,05)	0,07 (0,05)	0,34 (0,04)		0,06 (0,05)	
G <sup>2</sup> (Δ)		70,5		110,9 (40,4)	76,6 (6,1)		77,7		97,7 (20,0)
Degrés de liberté (Δ)		32-17=15		23 (8)	19 (4)		32-17=15		23 (8)

**Tableau 4.**  
**Comparaison des analyses de scores**  
**additifs et 0**

	statut d'emploi		Âge		interaction statut- Âge		Homogénéité (Bartlett-Box)	
	F	p	F	p	F	p	F	p
<i>Importance accordée à l'intégration</i>								
SCORE ADDITIF	7,36853	,007	25,98018	,000	1,43241	,232	4,41	,004
SCORE 0	5,64777	,018	23,43783	,000	1,92295	,166	2,74	,042
<i>Satisfaction extrinsèque</i>								
- SCORE ADDITIF total	13,91382	,000	7,96923	,005	6,39501	,012	1,56	,196
- Score additif (sans satisf. salaire)	,75839	,384	,37329	,541	3,91642	,048	6,70	,000
SCORE 0 (sans satisfaction salaire)	,89955	,343	2,14113	,144	5,74487	,017	,257	,857
- Satisf. salaire seul	80,22967	,000	24,34607	,000	5,47874	,019	3,61	,013

**Effets simples du statut d'emploi sur la satisfaction extrinsèque**

	Chez les moins de 35 ans		Chez les 35 ans et plus	
	F	p	F	p
- Score additif total	,34	,557	25,13	,000
- Score additif (sans satisf. salaire)	4,04210	,045	,97790	,323
Score 0 (sans satisf. salaire)	6,42000	,011	2,17000	,141
- Satisf. salaire	16,87885	,000	78,56655	,000

**Relation entre scores theta et additifs**

**Valorisation sociale**

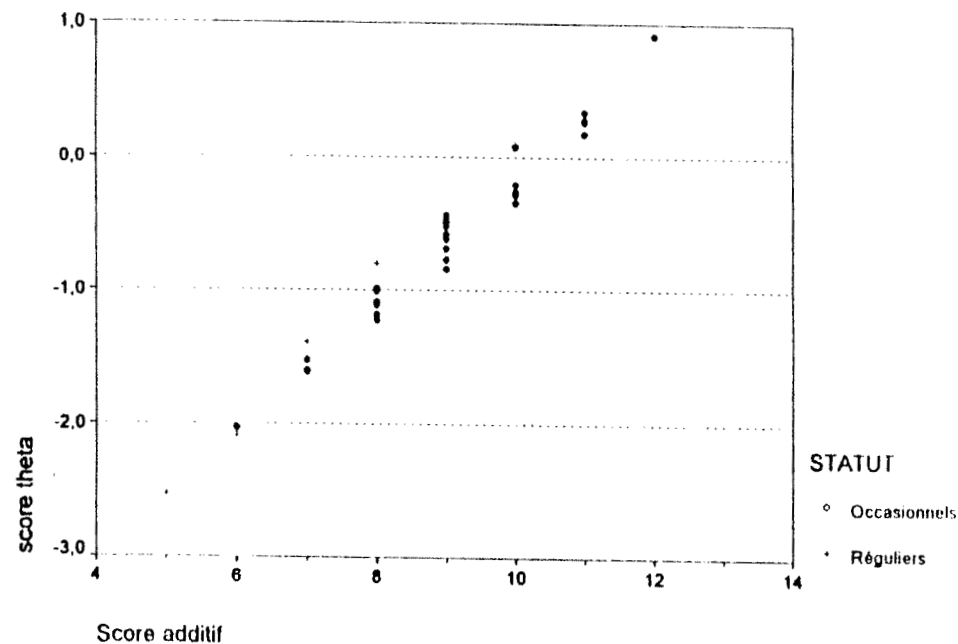


Figure 1.a

### Relation entre scores theta et additifs

Satisfaction extrinsèque (sans salaire)

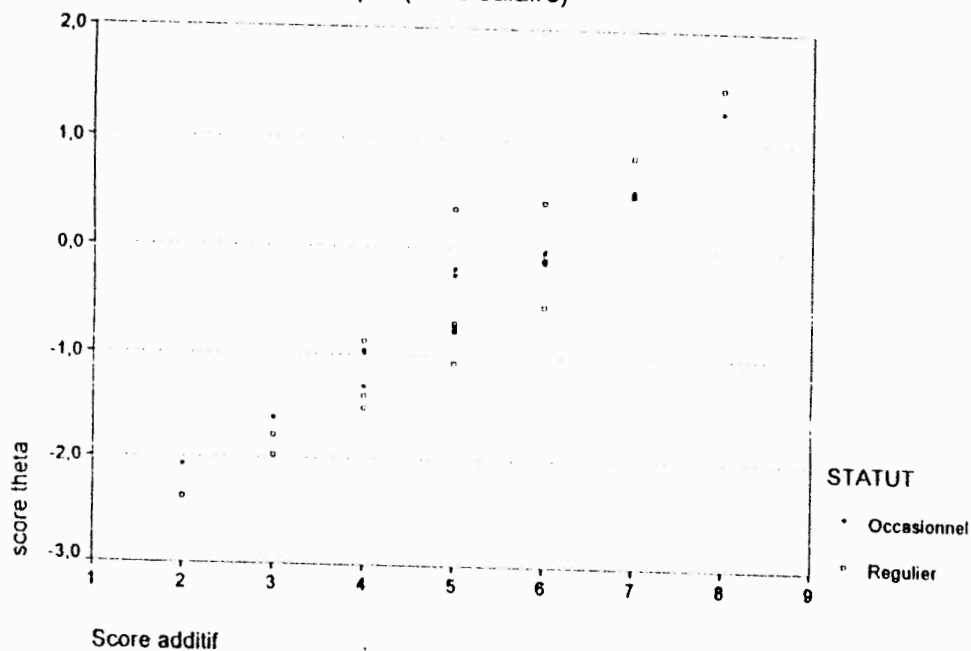


Figure 1.b

## THE USE OF LOCAL NETWORKS IN A STUDY OF HEROIN USERS: ASSESSING AVERAGE LOCAL NETWORKS

by

Ingegerd Jansson

(Department of Statistics, Stockholm University, Sweden;  
ingege.jansson@stat.su.se)

Marinus Spreen

(Department of Statistics and Measurement Theory (ICS),  
University of Groningen, Netherlands;  
m.spreen@ppsw.rug.nl)

**Résumé.** L'utilisation de réseaux locaux dans l'étude des consommateurs d'héroïne. **Évaluation de réseaux locaux moyens.** Dans l'étude d'une population de consommateur d'héroïne dans une ville néerlandaise, une approche par réseaux locaux a été utilisée dans la collecte et l'analyse de données concernant l'environnement social des consommateurs d'héroïne. Au lieu d'appliquer une procédure d'échantillonnage par boule de neige pour contacter des répondants par vagues successives, cette étude s'est concentrée sur la collecte de données relationnelles par une procédure boule de neige à une seule vague. La population cible a été divisée en deux strates: consommateurs d'héroïne participant, respectivement, ne participant pas, à un programme de soins. Dans le groupe de ceux impliqués dans un programme de soins, un tirage au hasard a été réalisé. Dans l'autre groupe mal délimité, un tirage non au hasard a été réalisé. Plusieurs aspects de la représentativité sont discutés dans la considération de l'environnement social des répondants, y compris le pourcentage des consommateurs participant régulièrement à un programme de soins, ceux participant à d'autres programmes, ceux participant aux deux types de programmes, et ceux ne participant à aucun programme. **Consommateurs d'héroïne. Données relationnelles. Échantillonnage par boule de neige. Programmes d'assistance aux consommateurs de drogues.**

**Abstract.** In a study of a population of heroin users in a Dutch city, the local network approach was used to collect and analyse information about the social environment of the heroin users. Instead of applying a snowball sampling procedure to contact respondents by successive waves, the focus in this study was on collecting relational data with a one-wave snowball sampling procedure. The population was divided into two strata: heroin users attending, respectively, not attending, regular drug assistance sessions. From the group attending regular drug assistance sessions, a random sample was drawn; for the other unknown group, a non-random sample. Several aspects of representation reliability are discussed when assessing in the social environment of the respondents, including the proportion of users visiting the regular drug assistance, of users visiting other drug assistance sessions, of users visiting both kinds of assistance sessions, and of users visiting no assistance sessions at all. **Heroin Users. Relational Data. Snowball Sampling. Drug Assistance Programs.**

### INTRODUCTION

Problems of collecting data in a drug-user population is part of the so-called hidden populations discussion. Some recent contributions are Watters and Biernacki (1989), van Meter (1990), and Spreen