

Le TOP 12 : comment interpréter les réponses comme des mesures de la capacité de la mémoire collective ?

TOP 12: how to interpret the responses as a measure of collective memory?

Émilie Lacot¹, Emmanuel J. Barbeau², Catherine Thomas-Anterion³, Sandrine Basaglia-Pappas³, Jérémie Pariente⁴, Michèle Puel⁴, Stéphane Vautier¹

¹ Unité de recherche OCTOGONE/CERPP, Pavillon de la Recherche, 5 allées Antonio Machado, 31058 Toulouse Cedex 9 <emilie@lacot.net>

² Unité de recherche CERCO, Toulouse III

³ Unité de neuropsychologie, CM2R, CHU de Saint-Étienne, Saint-Étienne

⁴ Unité de neuropsychologie, CM2R-CHU Purpan, Toulouse

Pour citer cet article : Lacot É, Barbeau EJ, Thomas-Anterion C, Basaglia-Pappas S, Pariente J, Puel M, Vautier S. Le TOP 12 : comment interpréter les réponses comme des mesures de la capacité de la mémoire collective?. *Rev Neuropsychol* 2011 ; 3 (4) : 284-9 doi:10.1684/nrp.2011.0189

Résumé

Introduction. Le TOP 12 évalue la mémoire collective au travers d'une série de huit types de questions portant sur le souvenir de la vie de 12 célébrités nommément désignées. La validation de tels tests est souvent envisagée dans le seul but de prédire un critère externe au test (validation externe) ; la validation interne n'est quant à elle que très rarement étudiée. **Objectifs.** Montrer comment les réponses obtenues peuvent mesurer une seule grandeur hypothétique (appelée aussi construit) : la capacité de la mémoire collective. **Méthodes.** L'échantillon est composé de 145 sujets (91 témoins, 32 patients présentant une maladie d'Alzheimer, 21 patients ayant un trouble cognitif léger de type amnésique, 1 patient ayant une démence sémantique). Deux étapes sont nécessaires : modéliser les réponses aux items à l'aide d'un modèle de réponse à l'item à trois paramètres et tester l'unidimensionnalité des scores estimés. **Résultats.** Les huit modèles s'ajustent étroitement aux données. L'analyse factorielle confirmatoire ne permet pas de rejeter l'idée selon laquelle les huit types de questions mesurent bien une seule et unique grandeur hypothétique. **Conclusion.** La modélisation psychométrique des données observées avec le TOP 12 indique qu'elles mesurent la capacité de la mémoire collective.

Mots clés : mémoire collective • TOP 12 • validation interne • modélisation psychométrique • grandeur hypothétique

Abstract

Introduction. *The TOP 12 is a test designed to assess collective (semantic) memory quickly and easily. It relies on eight questions about the life of 12 celebrities that are named (verbal presentation). Test validation is often based on the definition of a criterion external to the test (i.e. a pathological threshold, external validation); internal validation is itself rarely studied. In the present study, our aim is to assess the internal validation of the TOP 12.* **Aims.** *To show how the responses provided by the subjects measure a single hypothetical dimension (also called a construct): collective memory ability.* **Methods.** *The sample was made of 145 subjects (91 control subjects, 32 patients with Alzheimer's disease, 21 with amnesic mild cognitive impairment (aMCI) and 1 with semantic dementia). Two stages of analysis were required: modelling the responses for each item using a dedicated model with three parameters and testing for the unidimensionality of the estimated scores.* **Results.** *The eight models fit tightly to the data. A combinatorial factorial analysis does not allow rejecting the idea that the eight questions assess one and only one hypothetical dimension.* **Conclusion.** *Psychometric modelling of the TOP 12 data shows that they assess collective (semantic) memory.*

Key words: collective memory • TOP 12 • internal validation • data processing • hypothetical variable

Correspondance :
É. Lacot

Introduction

Le problème général : comment relier le référentiel des réponses au test à la notion de capacité de la mémoire collective ?

Dans l'article précédent : « Le TOP 12 : comment s'en servir pour repérer une pathologie du vieillissement cognitif ? », il n'était pas nécessaire de disposer d'une interprétation de ce que pourraient mesurer les réponses observées afin de répondre à la question posée. Lorsqu'on utilise un test afin de repérer si la personne examinée présente ou non des troubles du vieillissement cognitif, on cherche à optimiser la capacité diagnostique du score de manière empirique, ou indépendamment d'une signification neuropsychologique préalablement établie ; à la limite, n'importe quel « bricolage psychotechnique » convient à condition de fournir une caractérisation ordinale (*i.e.*, une position sur échelle ordonnée) qui optimise la valeur diagnostique du test. La question à laquelle cet article a pour objectif de répondre est de nature interprétative : il s'agit de montrer comment les réponses observables peuvent être conçues comme des mesures d'une grandeur hypothétique (appelée aussi construit). L'objectif est donc de montrer comment les réponses obtenues au TOP 12 (GRECO, 2008) peuvent mesurer la capacité de la mémoire collective.

Le TOP 12 permet d'observer 96 réponses, leur cotation dichotomique (0 ou 1) les situant dans un référentiel de description (Vautier, 2011)¹ constitué de 2^{96} 96-uplets possibles [1]. Le problème qui se pose est donc de savoir comment on peut relier logiquement le référentiel empirique qualitatif au référentiel neuropsychologique quantitatif – et réciproquement. Autrement dit, il s'agit de savoir en quel sens un 96-uplet quelconque peut mesurer la capacité de la mémoire collective, c'est-à-dire une grandeur variant par hypothèse dans l'ensemble des nombres réels. Cette question n'a rien à voir avec les propriétés diagnostiques de la somme des 1 qui figurent dans les 96-uplets (*cf.* article précédent, « Le TOP 12 : comment s'en servir pour repérer une pathologie du vieillissement cognitif ? »).

S'il n'existait pas de solution à ce problème, il serait alors logiquement absurde d'affirmer que le TOP 12 puisse mesurer la capacité de la mémoire collective en tant que grandeur continue. En effet, le test n'opérationnalise pas le principe de continuité impliqué par la notion de capacité de la mémoire collective. C'est en ce sens que Cliff et Keats (2003) soulignent la nécessité d'utiliser un modèle

de mesure : « L'idée fondamentale lorsqu'on assigne de manière intelligible des nombres à des événements, en particulier des scores à des réponses à un test, est qu'il doit exister un modèle quantitatif ou une mini-théorie qui les articule. Le modèle peut être simple ou complexe, testable empiriquement ou non, mais ce modèle doit exister » (p. 16) [2].

Spécification du problème : comment plusieurs items dichotomiques peuvent-ils mesurer une grandeur hypothétique ?

Le problème est résolu à l'aide d'un modèle de réponse à l'item (MRI). On suppose qu'un individu u considéré à une date t occupe une position θ (thêta) sur la grandeur hypothétique. On suppose encore qu'il existe une probabilité p que, occupant cette position, il « réussisse » l'item (par exemple, obtienne le score 1). On suppose enfin que cette probabilité est une fonction numérique de l'item et de l'unité d'observation : $\omega = (u, t)$. Cette fonction est telle que pour un item donné, si on compare la probabilité de réussite de $\omega_1 = (u_1, t_1)$ à celle de $\omega_2 = (u_2, t_2)$, alors $\theta_1 > \theta_2$ implique $p(\omega_1) > p(\omega_2)$ – autrement dit, la fonction est strictement croissante. On notera $p_i(\theta)$, la probabilité de réussir l'item i lorsque la grandeur vaut θ . La solution est généralisée à tout item j à l'aide de la notion de difficulté de l'item. Ainsi, si l'item i est plus difficile que l'item j , on pose que $p_i(\theta) < p_j(\theta)$. On introduit enfin le postulat d'indépendance des événements qui se réalisent lors de la réponse aux items (postulat dit d'indépendance locale), de telle sorte que la probabilité d'un k -uplet de réponses s'obtient par le produit de la probabilité de chaque réponse [3-6].

Application du problème au TOP 12

Les items du test ne sont pas « homogènes ». En effet, les réponses obtenues dépendent de huit types de questions (familiarité, mort/vivant, nationalité, profession, question morphologique, question spécifique 1, question spécifique 2, datation) à propos de 12 personnalités évoquées par leur nom (Édith Piaf, Simone Veil, Coluche, Marilyn Monroe, Alfred Hitchcock, Ronald Reagan, Amélie Mauresmo, Martine Aubry, Tony Blair, José Bové, Georges W. Bush, Vladimir Poutine). Chaque type de question définit un champ sémantique spécifique. Par exemple, la question « profession » renvoie à l'activité professionnelle des 12 personnalités tandis que la question « mort/vivant » renvoie à l'idée de leur décès possible. Par conséquent, on ne peut pas affirmer à l'avance que la capacité de la mémoire collective telle qu'elle est mesurée avec des questions liées à la profession soit identique à la capacité de la mémoire collective telle qu'elle est mesurée avec l'indicateur de décès. *A priori*, on peut supposer que chaque champ sémantique est associé à une facette spécifique de la mémoire collective, représentée par une grandeur spécifique. C'est pourquoi il faudra dans un deuxième temps se demander si les huit grandeurs respectives peuvent être considérées comme unidimensionnelles. Ce problème est abordé plus

¹ Un référentiel de description est un ensemble d'états possibles. Ces états caractérisent le mode de manifestation des objets étudiés selon un certain parti pris théorique. Dans le cas du TOP 12, une personne à un moment donné est caractérisée par la donnée de 96 valeurs descriptives. Ces 96 données constituent un 96-uplet de réponses. Ainsi, le TOP 12 permet de caractériser une personne à un moment donné dans un espace qualitatif et multivarié. Formellement, un référentiel de description est un produit cartésien de modalités descriptives qualitatives.

loin (L'unidimensionnalité des huit grandeurs modélisées par le TOP 12).

La question qui se pose est donc la spécification d'un modèle psychométrique. Considérons les 12 items d'un type de question donné. Chaque type de question définit un référentiel spécifique constitué de 2^{12} 12-uplets de réponses. La solution la plus simple est de faire appel au modèle de Rasch. Dans ce modèle, la probabilité de réussir l'item i étant donnée la position θ est définie comme suit :

$$p_i(\theta) = \frac{\exp(\theta - \beta_i)}{1 + \exp(\theta - \beta_i)} \quad (1)$$

L'équation 1 spécifie la probabilité $p_i(\theta)$ comme une fonction logistique des paramètres θ (capacité mnésique) et β_i (paramètre de difficulté de l'item i). Lorsque θ tend vers $-\infty$, la probabilité $p_i(\theta)$ tend vers 0. Or, le TOP 12 met les personnes examinées dans la situation de répondre à chaque question en choisissant obligatoirement parmi un nombre limité de réponses possibles (réponse à choix forcé). En supposant que la personne ait une capacité de mémoire infiniment petite, sa réponse s'effectuant au hasard, la probabilité $p_i(\theta)$ n'est pas nulle : elle est égale à $1/n$, n étant le nombre de réponses possibles pour un item donné (par exemple, pour un item à trois modalités de réponses, $p_i(\theta) = 1/3$). Ainsi, la modélisation logistique doit prendre en compte la probabilité de réussir l'item i lorsque θ tend vers ∞ . L'équation 1 doit donc être corrigée, ce qui conduit à l'équation 2 :

$$p_i(\theta) = \gamma_i + (1 - \gamma_i) \frac{\exp(\theta - \beta_i)}{1 + \exp(\theta - \beta_i)} \quad (2)$$

Où γ_i (paramètre de pseudo-chance) fixe la valeur asymptotique en $-\infty$. Ainsi, la modélisation permet d'interpréter un 12-uplet de réponses comme le résultat d'une expérience aléatoire dans des conditions théoriques spécifiques fixées par la valeur que prennent les paramètres θ , β_i et γ_i . La modélisation permet d'estimer ces conditions théoriques, c'est-à-dire de déterminer la valeur des paramètres du modèle à l'aide d'une technique itérative de minimisation d'une fonction d'ajustement du modèle aux données [4, 5]. Lorsque cette estimation est effectuée, il est possible d'évaluer *a posteriori* la qualité avec laquelle le modèle s'ajuste aux données en utilisant, par exemple, un Chi-carré d'ajustement (χ_i^2) [7-9]. Le χ_i^2 est donné par l'équation Soit : l'équation 3 qui suit, où $O_i(k)$ est la fréquence observée de la réponse k ($k=0, 1$), et $E_i(k)$ est la fréquence théorique de la réponse k obtenue à l'item i :

$$\chi_i^2 = \sum_{k=1}^2 \frac{[O_i(k) - E_i(k)]^2}{E_i(k)} \quad (3)$$

■ L'unidimensionnalité des huit grandeurs modélisées par le TOP 12

L'analyse factorielle confirmatoire (AFC) fournit un cadre analytique adapté à la question de savoir si on peut considérer que les huit grandeurs invoquées pour interpréter les réponses aux huit types de questions sont unidimensionnelles [10, 11]. Deux variables sont unidimensionnelles s'il est possible de les considérer comme dépendantes d'un seul facteur commun. Cette idée est formalisée à l'aide d'un modèle unifactoriel, qu'on appelle aussi un modèle congénérique [12].

Dans le modèle congénérique, les variables indicatrices ne sont pas dichotomiques mais continues et non bornées. Elles sont ici approchées à l'aide des estimations des paramètres θ associées à chacun des huit modèles de réponse à l'item. La fonction qui les relie au facteur commun est linéaire et l'erreur de mesure est conçue comme une variable aléatoire centrée. On introduit la notion d'indépendance locale en spécifiant que les huit variables d'erreur ne sont pas corrélées.

Comme son nom l'indique, l'AFC permet de tester l'hypothèse nulle. Dans l'AFC, l'hypothèse nulle stipule que les moments (espérances, variances, covariances) d'un ensemble de variables indicatrices dépendent d'un ensemble de paramètres dits structurels [10]. Comme dans la modélisation des réponses à l'item, la modélisation des moments estimés repose sur l'estimation des paramètres selon une méthode itérative d'ajustement du modèle aux données. Si l'hypothèse nulle doit être rejetée, cela signifie que le modèle congénérique, compatible avec l'idée que les huit grandeurs estimées mesurent la même capacité, ne permet pas d'interpréter les données. La conséquence théorique du rejet de l'hypothèse nulle est la remise en question de l'idée selon laquelle le TOP 12 mesurerait une seule capacité mnésique.

Comme son nom l'indique, l'AFC permet de tester l'hypothèse nulle. Dans l'AFC, l'hypothèse nulle stipule que les moments (espérances, variances, covariances) d'un ensemble de variables indicatrices dépendent d'un ensemble de paramètres dits structurels [10]. Comme dans la modélisation des réponses à l'item, la modélisation des moments estimés repose sur l'estimation des paramètres selon une méthode itérative d'ajustement du modèle aux données. Si l'hypothèse nulle doit être rejetée, cela signifie que le modèle congénérique, compatible avec l'idée que les huit grandeurs estimées mesurent la même capacité, ne permet pas d'interpréter les données. La conséquence théorique du rejet de l'hypothèse nulle est la remise en question de l'idée selon laquelle le TOP 12 mesurerait une seule capacité mnésique.

■ Méthode

■ Échantillon et matériel

L'échantillon de l'étude se compose des 145 sujets décrits dans l'article précédent (« Le TOP 12 : comment s'en servir pour repérer une pathologie du vieillissement cognitif ? »). Le matériel, lui aussi décrit dans ce premier article, est le TOP 12 (GRECO, 2008).

■ Analyses

Les analyses sont effectuées en deux étapes. La première consiste à modéliser les réponses aux items, la seconde à tester l'unidimensionnalité des scores estimés lors de la première étape. La modélisation des données selon un MRI à trois paramètres (équation 2) a été effectuée à l'aide du logiciel Multilog [13]. Le modèle est spécifié avec deux contraintes structurelles : la pente de chaque fonction logistique est invariante afin que la difficulté relative des items ne varie pas en fonction des capacités de chaque sujet. Le paramètre de pseudo-chance est fixé en fonction du nombre de modalités de réponses de l'item. La qualité d'ajustement des modèles aux données a été mesurée à l'aide du χ^2 d'ajustement (équation 3) [14]. L'AFC a été effectuée à l'aide du logiciel Mplus (Muthén et al., 1998).

Résultats

Modèles de réponse à l'item

Les huit modèles s'ajustent étroitement aux données. Les 96 χ^2 d'ajustement sont compris entre 0 et 0,09. Par conséquent, les paramètres θ_{ij} estimés sont utilisables pour l'AFC.

Modèle congénérique

La matrice de variances, de covariances et de corrélations est présentée dans le *tableau 1*. Le modèle congénérique s'ajuste étroitement aux données $\chi^2 (N = 145, dl = 20) = 26,05, p = 0,16$, ce qui permet de ne pas rejeter l'idée selon laquelle les huit facettes de la mémoire collective indiquent une seule et même grandeur hypothétique. La puissance statistique du test peut être estimée à l'aide du *root means square error of approximation* (RMSEA) si on spécifie sa valeur dans la population (ici 0,08) et sa valeur pour l'hypothèse nulle (ici 0). La puissance statistique estimée pour une erreur de type I de 0,05 est 0,73 [15].

Les estimations des paramètres du modèle congénérique sont indiquées dans la *figure 1*. Pour chaque variable indicatrice (notée $Y_i = \lambda_i F + \varepsilon_i$, où λ_i est le coefficient de régression linéaire de cet item i et ε_i sa valeur résiduelle), il est possible d'estimer la fidélité définie par l'équation 4 [16] :

$$\varphi(Y_i) = \frac{\lambda_i^2 \cdot \text{Var}(F)}{\lambda_i^2 \cdot \text{Var}(F) + \text{Var}(\varepsilon_i)} \quad (4)$$

où $\text{Var}(F)$ est la variance du facteur commun appelé ici mémoire collective.

Les estimations de la fidélité varient de 0,28 (variable datation) à 0,46 (variables questions spécifiques 1 et 2).

Tableau 1. Variance des huit variables estimées avec les modèles de réponse à l'item. Les covariances et les corrélations sont indiquées respectivement au-dessous et au-dessus de la diagonale.

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
θ_1	0,31	0,50	0,51	0,58	0,50	0,50	0,56	0,36
θ_2	0,19	0,50	0,49	0,56	0,56	0,63	0,53	0,39
θ_3	0,20	0,24	0,48	0,56	0,53	0,60	0,60	0,34
θ_4	0,20	0,25	0,24	0,39	0,52	0,54	0,63	0,34
θ_5	0,21	0,30	0,28	0,25	0,57	0,64	0,66	0,48
θ_6	0,21	0,33	0,37	0,25	0,36	0,56	0,71	0,50
θ_7	0,23	0,27	0,37	0,29	0,37	0,39	0,54	0,40
θ_8	0,15	0,21	0,18	0,16	0,28	0,29	0,23	0,60

Discussion

L'objectif de l'article était de montrer que la modélisation psychométrique des données permet d'opérationnaliser l'idée selon laquelle les performances observées avec le TOP 12 dépendent de manière probabiliste d'une ou de plusieurs capacités représentées quantitativement. Ainsi, les items de chaque sous-test du TOP 12 peuvent être interprétés comme indicateurs d'une grandeur latente. De plus, les huit grandeurs latentes construites avec les modèles de réponse à l'item peuvent être interprétées de façon unidimensionnelle : l'AFC des estimations issues de la MRI montre qu'il n'est pas nécessaire de postuler un espace multidimensionnel pour interpréter les scores estimés.

Les estimations structurelles du modèle congénérique suggèrent que la variable indicatrice « datation » a une fidélité particulièrement faible par rapport à celle des autres variables indicatrices. Autrement dit, les différences interindividuelles relatives au facteur commun expliquent peu les différences interindividuelles mesurées. La variabilité résiduelle peut être interprétée de différentes manières :

- en termes d'erreur de mesure ;
- en termes de validité discriminante si l'on suppose que la variance résiduelle de la variable datation s'explique par l'existence d'un autre facteur.

Pour départager ces deux interprétations, il serait nécessaire de développer le TOP 12 en ajoutant des items spécifiquement dédiés à la mesure de ce second facteur.

Conclusion générale

La validation d'un test désigne une multitude d'approches critiques destinées à préciser la signification des scores attribués aux personnes en fonction de leurs réponses aux items [17]. L'article précédent, « Le TOP 12 : comment s'en servir pour repérer une pathologie du vieillissement cognitif ? », se centrait sur la signification diagnostique des scores du TOP 12. La contribution principale de cet article consiste à fonder la signification diagnostique des scores sur l'utilisation d'un critère de pathologie établi indépendamment des scores. D'un point de vue pratique, un seul seuil de décision est par hypothèse valide pour toute personne satisfaisant les critères d'inclusion utilisés dans cette étude.

Le présent article a abordé un autre aspect de la validation du TOP 12 : il ne s'agissait non pas de prédire un critère externe au test (validation externe), mais de fonder l'intuition selon laquelle les performances observées avec le TOP 12 indiquent une capacité mnésique (mémoire collective) représentée de façon quantitative et univariée. Ainsi, la modélisation psychométrique (validation interne) a une finalité non pas pratique mais interprétative. Les analyses effectuées ne montrent pas la nécessité

$\chi^2(N = 145, dl = 20) = 26,05, p = 0,16$

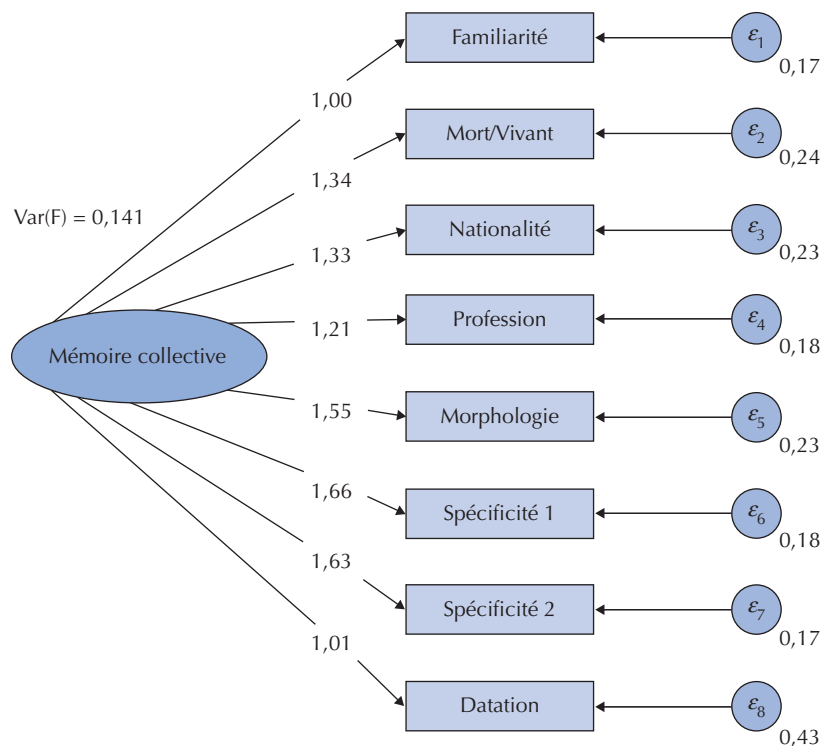


Figure 1. Estimations des paramètres du modèle congénérique.

d'invoquer plus qu'une variable latente pour interpréter les observations. Il paraît cependant intéressant que la variable indicatrice « datation » ait une fidélité faible par rapport aux autres variables, ce qui pourrait indiquer qu'une autre variable latente entre en jeu, par exemple la mémoire épisodique (item situé dans le temps), par oppo-

sition à la mémoire sémantique pour les autres variables indicatrices. ■

Conflits d'intérêts

Aucun.

Références

1. Vautier S. The operationalization of general hypotheses versus the discovery of empirical laws in psychology. *Philosophia Scientiae* 2011 ; 15 : 105-22.
 2. Cliff N, Keats JA. *Ordinal measurement in the behavioral sciences*. Mahwah : Lawrence Erlbaum Associates ed., 2003.
 3. Bertrand R, Blais JG. *Modèles de mesure : l'apport de la théorie des réponses aux items*. Montréal, Québec : Presses de l'Université du Québec, 2004.
 4. Embretson SE, Reise SP. *Item Response Theory for Psychologists*. London : Lawrence Erlbaum Associates, 2000.
 5. Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of item response theory*. Newbury Park : Sage Publications ed., 1991.
 6. Rasch C. *Probabilistic models for some intelligence and attainment tests*. Copenhagen : Danish Institute for Educational Research ed., 1960.
 7. Bolt DM, Hare RD, Vitale JE, et al. A multigroup item response theory analysis of the psychopathy checklist-revised. *Psychological Assessment* 2004 ; 16 : 155-68.

8. Stark S, Chernyshenko OS, Drasgow F. Detecting differential item functioning with confirmatory factor analysis and item response theory. *Journal of Applied Psychology* 2006 ; 91 : 1292-306.
 9. Zickar MJ, Russell SS, Smith CS, et al. Evaluating two morningness scales with item response theory. *Personality and Individual Differences* 2002 ; 33 : 11-24.
 10. Bollen KA. *Structural equations with latent variables*. New York : Wiley Series in Probability and Mathematical Statistics (Wiley ed.), 1989.
 11. Dicks P, Tournois J, Fieller A, et al. *La psychométrie. Théories et méthodes de la mesure en psychologie*. Paris : Presses Universitaires de France ed., 1994.
 12. Jöreskog KG. Statistical analysis of a set of congeneric tests. *Psychometrika* 1971 ; 36 : 109-33.
 13. Thissen D. *Multilog user's guide: multiples, categorical item analysis and test scoring using item response theory*. Chicago : IL : Scientific Software ed., 1991.

14. Drasgow F, Levine MV, Tsien S, *et al.* Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement* 1995 ; 19 : 143-65.
15. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure analysis. *Psychological Methods* 1996 ; 1 : 130-49.

16. Vautier S, Gaudron JP, Jmel S. Modèles factoriels pour l'analyse de la fidélité des variables composites. *Revue Epidémiologie Santé Publique* 2004 ; 52 : 441-53.
17. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist* 1995 ; 50 : 741-9.