

# L'évolution de l'outil informatique modifiera-t-elle la fonction du psychologue conseiller ?

Francis Gendre et Roland Capel

---

*D'ici l'an 2000, l'humanité aura à faire face à trois explosions : l'explosion atomique, l'explosion démographique et l'explosion de l'information.*

*A. Einstein*

*L'utilisation croissante de systèmes informatiques dans le domaine de l'évaluation psychologique dérive directement du développement des théories relatives à la construction des tests. L'application pratique (construction, passation, correction et interprétation) de nouveaux instruments suivra donc nécessairement une évolution dans le sens d'une complexité plus grande dans tous les aspects de leur utilisation. Cet article esquisse une description des liens étroits existant entre la conception et l'opérationnalisation de ces théories dans la pratique quotidienne des psychologues, notamment les psychologues scolaires et les conseillers en orientation.*

## 1. INTRODUCTION

L'utilisation croissante de l'outil informatique dans de très nombreuses professions déborde largement de ses fonctions originelles de gestion ou de recherche. Son usage tend à se répandre dans pratiquement tous les aspects de la vie professionnelle, suscitant à la fois crainte et enthousiasme (Gendre, 1982 ; Capel, 1991). Nous allons esquisser les conséquences probables de cette évolution sur la pratique des conseillers : en particulier l'évaluation psychologique à l'aide de tests ou de questionnaires.

Afin de mieux faire comprendre notre point de vue, il nous semble utile d'évoquer tout d'abord l'histoire des théories sous-jacentes à la construction des épreuves psychologiques. Nous montrerons notamment comment l'ordinateur, tout d'abord utilisé comme un moyen d'effectuer rapidement des calculs de plus en plus complexes, a pu devenir un modèle du fonctionnement de l'esprit humain. Nous examinerons ensuite les retombées pratiques de ces théories par une étude analytique de la procédure de testage, en particulier en ce qui concerne les phases de construction, de passation, de correction et d'interprétation des résultats. Pour terminer, nous esquisserons quelques conséquences probables en relation avec la fonction de psychologue - conseiller.

## 2. L'ÉVOLUTION DES THÉORIES ET DES TECHNIQUES DE CONSTRUCTION DE TESTS

### 2. 1. *La théorie classique, déterministe, dite aussi « théorie de l'échantillonnage d'un domaine »*

La théorie classique des tests s'applique essentiellement aux épreuves à choix multiples dont la correction est purement mécanique et de ce fait qualifiées d'*objectives*. L'origine de cette théorie remonte au début du siècle, elle est attribuée à Spearman (1904). Elle a été révisée par Gulliksen (1950), puis complétée par Nunnally (1978), pour subsister

---

---

aujourd'hui sous une forme que l'on appelle la théorie des formes parallèles échantillonnées au hasard ou *théorie de l'échantillonnage d'un domaine*.

La théorie classique propose de considérer le score observé comme la somme de deux composantes : le *score vrai* et l'*erreur*, cette dernière résultant de la procédure d'échantillonnage des items, des sujets, etc.<sup>1</sup>

Son objectif est de construire des mesures *fidèles*, dont la *validité* ne sera évaluée qu'ultérieurement, une fois la mesure construite. Cette théorie n'est donc pas le lieu d'hypothèses concernant la nature du domaine étudié (par exemple sa structure factorielle simple ou complexe) bien que la théorie classique s'applique d'autant mieux que le domaine investigué est homogène.

La *fidélité* est définie comme une fonction de la corrélation entre le score vrai et le score observé. Elle est estimée à partir de la moyenne des intercorrélations entre les items. Découlant de ces postulats (entre autres), la technique classique de construction d'échelles (analyse d'items), se fonde essentiellement sur deux indices : la *difficulté* (ou pourcentage de réponses correctes) et la *discrimination* (ou corrélation item-total) des items. Cette procédure aboutit à ce que l'on appelle des *tests standardisés en référence à une norme*. L'interprétation du résultat (score observé d'un individu à l'échelle considérée) s'effectue en le confrontant à un ou plusieurs étalonnages. Le score observé peut également intervenir dans des équations de régression permettant de prédire certains critères pertinents du point de vue du domaine étudié (réussite<sup>2</sup>, placement, etc.).

L'approche classique laisse une place appréciable à l'usage de l'ordinateur qui peut se révéler très utile lorsqu'il s'agit de calculer les indices de corrélation qui sont à la base de tous les paramètres déterminant la

- 
1. Introduite par Cronbach et ses collaborateurs (1972), puis complétée par Cardinet (1985), cette théorie propose de décomposer l'erreur de mesure en plusieurs éléments, en fonction du plan d'échantillonnage utilisé, dans ce que l'on appelle *la théorie de la généralisabilité*.
  2. L'utilisation de tables d'espérances (empiriques et théoriques) comme celles de Lawshe (1958) apporte une aide appréciable dans ce type d'application.
-

fiabilité d'une mesure. Toutefois, l'intervention de l'informatique n'est, dans ce cas, pas à proprement parler *indispensable*, une calculette pourrait, dans la plupart des cas rencontrés dans la pratique, s'y substituer<sup>3</sup>.

## 2. 2. *La théorie stochastique des courbes caractéristiques des items, ou « théorie de l'aptitude latente »*

Si la théorie classique a toujours été considérée comme bien adaptée à la prédiction des critères courants (réussite, adaptation, etc.), son application rencontra toutefois trois difficultés :

- Le premier problème tient au risque lié à la possibilité de généraliser des résultats. Il est clair que dans la procédure décrite ci-dessus, le résultat d'un individu dépend à la fois de l'échantillon d'items inclus dans le test, *et* de l'échantillon de sujets qui ont prêté leur concours à sa mise au point. Ainsi, par exemple, l'« intelligence générale » qui a fait l'objet de nombreuses mesures depuis 1905 (par Binet et Simon), doit être décrite non seulement à l'aide d'un score, mais aussi en relation avec le type particulier de test utilisé, ainsi qu'avec les caractéristiques du groupe de référence (barèmes).
- Le deuxième problème découle du fait que la théorie classique aboutit à déterminer un écart-type d'erreur *commun* à tous les sujets, quel que soit leur score. Mais il semble toutefois difficile d'assurer que tous les scores ont le même degré de précision, quelle que soit leur position sur l'échelle de mesure.
- Le troisième problème concerne le fait que la théorie classique ignore le pattern de réponse aux items. Un même score peut donc correspondre à différentes manières de réagir à l'ensemble des items.

En réponse à ces objections, de nouvelles idées ont pu être concrétisées grâce aux progrès de l'informatique. Leur objectif est d'aboutir à des mesures aussi objectives que possible, c'est à dire à la fois indépen-

---

3. Car – ne l'oublions pas – il n'existait pas encore d'ordinateurs à l'époque des principaux développements de ces idées.

---

dantes de la distribution des mesures observées sur un échantillon de personnes, et indépendantes de l'instrument de mesure lui-même (de l'échantillon d'items).

Cette théorie (IRT : Item Response Theory) est issue des travaux de Rasch (1960), Lord & Novick (1968) et Birnbaum (1968)<sup>4</sup>. Elle repose, comme la précédente, sur la décomposition d'un score observé en score vrai et erreur, mais présuppose en outre que les réponses obtenues sont la manifestation d'une aptitude unidimensionnelle inobservable ou latente. Dans cette théorie, l'unité d'analyse n'est plus seulement le test entier, mais aussi la réponse à chacun des items.

Les réponses individuelles aux items sont modélisées par une fonction qui dépend à la fois de l'aptitude du sujet et des paramètres de l'item. La courbe caractéristique d'un item représente la probabilité d'observer une réponse correcte, en fonction du niveau du trait mesuré. Cette fonction est non linéaire et présente la forme d'une ogive normale. Elle peut être décrite mathématiquement – dans les modèles les plus élaborés – par trois paramètres : la difficulté (valeur de l'abscisse au point d'inflexion de la courbe), la discrimination (pente de la courbe au point d'inflexion), et le pseudo-devinement, correspondant à la probabilité de réponse correcte pour un sujet ayant un niveau d'aptitude infiniment bas. Ces paramètres sont dits invariants, si bien que la probabilité d'observer une réponse correcte à un item ne dépend pas de la distribution de l'attribut mesuré.

A chaque courbe caractéristique est associée une courbe d'information qui permet d'estimer la précision avec laquelle l'item mesure chaque niveau du trait. Si bien que lorsqu'un ensemble d'items a été convenablement calibré, de telle sorte que les paramètres de tous les items soient connus, le niveau du trait peut être calculé à partir de ces paramètres, quel que soit le sous-ensemble d'items auquel le sujet a répondu. Cela permet de comparer deux sujets même s'ils ont été testés avec des ensembles d'items différents, non parallèles.

---

4. Son approche est trop complexe pour être résumée en quelques lignes. Le lecteur intéressé par une introduction simple peut lire le chapitre 8 de Lavault & Grégoire (1997), ou consulter Wright & Panchapakesan, 1969.

L'opérationnalisation de cette théorie nécessite de très larges échantillons, (environ 1000 sujets par paramètre), et de nombreux items (au moins 200). Ces conditions étant jugées indispensables pour estimer avec une précision suffisante les différents paramètres nécessaires. Les calculs sont complexes, itératifs, et nécessitent l'utilisation d'un ordinateur ainsi que l'achat de logiciels parfois coûteux. Les applications de ces modèles sont extrêmement séduisantes : il devient possible de construire des tests mieux adaptés au niveau des populations visées, et de créer ainsi des tests adaptatifs sur mesure, en sélectionnant les meilleurs items pour un individu donné, selon ses réponses antérieures.

Cette théorie permet donc de construire des tests tout aussi précis que les tests classiques, mieux adaptés aux sujets et offrant de surcroît la possibilité d'effectuer des comparaisons entre des sujets ayant répondu à des ensembles d'items différents. Mais elles nécessitent toutefois, comme on l'a vu, l'usage de l'outil informatique qui devient indispensable pour effectuer les calculs requis par la méthode.

### **2. 3. Les théories du traitement de l'information**

Malgré la rupture radicale que représente la construction et la correction d'un test du type IRT, par rapport à un test classique, le fondement théorique des deux modes de construction est le même : toutes deux cherchent à mesurer des *traits* ou des *facteurs*. Cette approche a subi ces dernières années des critiques de plus en plus sérieuses. Car en effet, quel que soit le mode de construction du test, les instruments traditionnels présentent toujours aux sujets des problèmes simplifiés, restreints, et ne tiennent compte que du résultat final de l'activité mentale, sans se préoccuper réellement des processus cognitifs impliqués.

Certains chercheurs (*cf. infra*) se fixèrent des objectifs beaucoup plus ambitieux que ceux des constructeurs de tests traditionnels. Le but visé étant rien moins que d'expliquer et de comprendre les processus mentaux déterminant la réponse donnée, en s'appuyant sur les théories cognitives de l'apprentissage et du traitement de l'information (voir Snow, 1990).

Ces théories sont fortement influencées par le modèle du fonctionnement des ordinateurs (intelligence artificielle). Du point de vue technique, les variables continues, les relations linéaires et l'analyse factorielle ne constituent plus les éléments essentiels de ces modèles ; ceux-ci s'appuient plutôt sur des variables nominales, des relations non linéaires, l'analyse de variance et de covariance multivariée, voire les séries temporelles (Shye, 1978). Ces nouveaux modèles utilisent également les développements récents des théories stochastiques en les généralisant aux catégories, qu'elles soient ordonnées ou non. D'autre part, ils s'inspirent également des résultats de travaux effectués dans le domaine de l'intelligence artificielle, de la construction de systèmes experts, de l'instruction et de l'évaluation pilotées par ordinateur (Plake & Witt, 1986).

Dans cette perspective, les tests tendent à devenir plutôt éduométriques que psychométriques, dans la mesure où ils sont conçus comme des plans d'observation, permettant de construire des modèles inférables aux mécanismes de production des réponses.

Pour illustrer cette approche, on peut évoquer le *modèle componentiel* de Sternberg (1977, complété en 1981), qui propose un modèle du fonctionnement mental qui serait le produit du travail de différentes unités de traitement appelées *composants*. Ceux-ci auraient des fonctions spécifiques, telles l'encodage des stimuli, la comparaison des caractéristiques, l'induction et l'application de règles, ainsi que la justification de la réponse. Une telle approche nécessite une indexation des performances sur deux dimensions principales : la vitesse et la précision, tout en tenant compte des patterns de réponse. On peut également relever les tentatives de génération automatique d'items dans le domaine des aptitudes spatiales (Bejar, 1990) ou verbales (Bejar & al, 1991), de même que dans les problèmes de sélection de personnel (Hesketh & Robertson, 1993), ainsi que pour le diagnostic médical (Parker et Miller, 1988).

Comme on peut s'en douter, ces nouveaux instruments génèrent, eux aussi, de nombreux problèmes (non encore résolus) relatifs à leur fidélité, à leur validité et à la possibilité de généraliser leur application. Ils s'orientent cependant vers une direction souhaitable : comprendre. Mais leur inconvénient majeur réside dans leur complexité technique, leur développement et leur application nécessitent des connaissances appro-

fondées des disciplines mathématiques, statistiques et informatiques, dépassant largement le niveau moyen de compétence des praticiens, voire de la plupart des chercheurs non spécialisés dans ce domaine.

En résumé, si la théorie classique est surtout associée aux tests papier-crayon et aux passations collectives, les « nouvelles théories » (que ce soient celles des courbes caractéristiques ou celles du traitement de l'information) impliquent des passations individuelles, administrées *exclusivement* par un ordinateur. L'outil informatique devient donc, outre sa fonction d'auxiliaire indispensable au calcul, un modèle théorique de l'activité mentale dans la mesure où il permet un véritable « pilotage » du *testing* en adaptant et créant des items en fonction des réactions du répondant.

### **3. L'ÉVOLUTION PROBABLE DES INSTRUMENTS PSYCHOMÉTRIQUES**

Nous allons maintenant examiner l'impact de l'informatisation sur l'évolution de la théorie des tests, en relation avec les trois phases principales de la pratique du testage.

#### ***3. 1. Informatisation et construction des tests***

Sans prétendre à la clairvoyance, notre opinion est que dans les années à venir, la construction des tests restera pour une grande part de type traditionnel, que ce soit dans le domaine des tests d'aptitudes ou dans celui des questionnaires d'intérêts et de personnalité. Ce secteur restera vivace non seulement à cause de sa propre inertie (les habitudes ne changent pas facilement), mais aussi parce que les épreuves les plus anciennes sont aussi celles sur lesquelles on dispose du maximum d'informations. De leur côté, les variantes « modernes » n'ont pas encore prouvé leur validité incrémentale, c'est-à-dire leur supériorité à prédire des critères globaux, socialement importants. Les épreuves construites sur la base des théories des courbes caractéristiques, et plus spécialement celles utilisant le testage adaptatif, ne remplaceront donc, à notre avis, que très lentement les épreuves classiques. De plus, cette mutation n'interviendra

---

qu'à la condition de stocker efficacement les informations disponibles sur un grand nombre d'items, de manière à être assemblées sous forme de banque de données calibrées (*cf.* Wood & Skurnik, 1969).

Nous imaginons plus tardive l'introduction de formes commercialisées d'épreuves utilisant l'ordinateur à la fois comme générateur d'items et de système expert – couplant l'instruction pilotée par ordinateur avec l'évaluation automatique.

### ***3. 2. L'informatisation de la passation***

Comme nous l'avons vu, seuls les tests classiques se prêtent à des passations collectives, alors que toutes les nouvelles formes impliquent des passations individuelles. Une forme intermédiaire nous semble cependant promise à un grand développement, en effet : un grand nombre de tests classiques se prêtent fort bien à une présentation sur écran et à une correction automatique. L'existence de salles informatiques dans les écoles permet des passations simultanées d'une vingtaine d'élèves, chacun d'eux disposant de sa propre machine. De plus, l'informatisation complète de la passation et de la correction permettra la construction de batteries « synthétiques » mesurant simultanément divers aspects de l'individu (intérêts, personnalité, valeurs, aptitudes, etc.), de manière à rendre un diagnostic différencié (instrument par instrument), ou global, sous la forme d'une liste de professions susceptibles de convenir au mieux au sujet, en fonction de ses réponses, par exemple<sup>5</sup>.

L'informatisation de tests classiques donne également la possibilité de proposer des items plus variés par leur contenu et leur forme, et on peut imaginer une certaine interactivité en programmant la machine pour qu'elle envoie des messages au répondant, en fonction (par exemple) de son temps de réponse ou du nombre d'erreurs commises.

---

5. Cet objectif a déjà été atteint avec la création du système O.R.U.S. (Orientation Raisonnée Unifiée et Synthétique (Gendre & *al.* 1997).

Cette modernisation inévitable – et à notre avis nécessaire – de l'instrumentation psychométrique présente sans doute certains inconvénients : on court le risque de transformer le testage en une opération « magique », et il faut également s'interroger à propos des biais dus à la maîtrise plus ou moins développée de l'outil ; de plus, seul face à l'écran, le répondant devient plus difficile à observer. De nombreuses questions restent sans réponse à propos des caractéristiques individuelles susceptibles d'influencer le style et le contenu des réponses données devant un écran.

### ***3. 3. L'informatisation de la correction***

Le simple décompte du nombre de réponses correctes exigé par les épreuves classiques peut se faire aussi bien à la main que par ordinateur. Par contre, l'utilisation de statistiques bayésiennes, de procédures d'estimation par maximum de vraisemblance ou l'analyse des résultats de plans d'expériences complexes, nécessitent une correction que seul l'ordinateur peut effectuer, et ceci de manière quasi instantanée.

Par conséquent, il nous semble probable que l'on doive bientôt distinguer deux sortes de tests :

- d'une part des épreuves extrêmement simples, robustes, et malgré tout valides, telles par exemple les épreuves d'auto-évaluation, auto-correctives et auto-interprétées, du genre SDS de Holland ;
- et d'autre part, les épreuves plus ou moins sophistiquées, impliquant dans leur correction des transformations, des catégorisations, des pondérations (voir par exemple le manuel du système O.R.U.S.), ou encore des résolutions d'équations probabilistes (tests du type IRT).

### ***3. 4. L'informatisation de l'interprétation***

L'interprétation d'un résultat de test implique une standardisation par référence à une norme. Mais cette opération n'est pas nécessaire lorsqu'il s'agit de prédire – soit la réussite globale vis-à-vis de critères

---

---

socialement importants, soit la réussite différentielle d'un éventuel placement dans différents « traitements ». Cette dernière application devrait aboutir à une évaluation en termes de probabilité de « réussite ».

L'utilisation de scores observés dans des équations de régression suppose une certaine confiance dans leur fiabilité, cette sécurité d'emploi sera favorisée par l'utilisation de plus en plus fréquente des résultats de méta-analyses<sup>6</sup>. Et finalement, l'utilisation de tables d'espérances permettra de formuler des pronostics basés sur des pourcentages de réussite attendue, en relation avec la validité du test. Toutes ces prestations peuvent être automatisées, mais pour l'instant, la plupart des systèmes de correction automatique se bornent à ne fournir que des scores standardisés. On peut toutefois supposer<sup>7</sup> que de véritables rapports, procurant une synthèse en clair des résultats du testage, feront bientôt partie intégrante de toute correction automatique.

#### **4. CONCLUSION : L'ÉVOLUTION PROBABLE DE LA PROFESSION DE PSYCHOLOGUE - CONSEILLER**

Tout ce que nous avons dit jusqu'ici à propos de l'évolution des théories et des techniques psychométriques peut inciter à croire que loin de simplifier le travail du psychologue, l'outil informatique risque bien de lui rendre la vie de plus en plus compliquée. Cette crainte n'est pas sans fondement, car l'évolution que nous avons tenté de préfigurer pourrait bien conduire à un clivage radical : soit les praticiens psychologues abandonneront l'évaluation psychologique aux mains des fabricants de logiciels ; soit ils tenteront de maîtriser les nouveaux outils mis à leur disposition par un effort de formation dans le sens d'une spécialisation.

La première réaction entraînerait le risque de voir les critères commerciaux prendre le pas sur les critères scientifiques, et celui de renforcer l'opacité des procédures d'évaluation. Dans cette perspective, des associations de type « associations de consommateurs de tests », devront se constituer pour effectuer les tests (!) de ces produits. Elles auront pour

---

6. Méta-analyse : synthèse de diverses recherches.

7. Faut-il également l'espérer ? La question reste à débattre...

rôle d'exiger des fabricants qu'ils fournissent toutes les informations jugées indispensables pour l'évaluation comparative de l'utilité et de la fiabilité de leurs produits. La seconde option impliquerait une diversification de la profession en plusieurs spécialisations : le psychologue-conseil, le psychométricien, le « conseiller-dépanneur » en informatique, le spécialiste de l'économie et du marché du travail, celui des problèmes de motivation ou de réinsertion, le formateur-évaluateur, etc. Il se peut que l'on observe une évolution comparable à celle qui a déterminé la profession de médecin, en rapport avec le développement des techniques de diagnostic. Cette évolution nécessitera de la part des praticiens une formation plus poussée dans les domaines de l'informatique, de la psychométrie, de la connaissance du travail et de l'économie, ainsi que des procédures d'entretien, d'aide ou de soutien. On peut aussi imaginer que certaines habitudes changeront : ces spécialisations exigeront des psychologues qu'ils collaborent davantage entre eux et que la prise en charge des consultants ne soit plus exclusivement individuelle, mais plutôt intégrée dans le cadre d'un travail d'équipe visant à optimiser la complémentarité des compétences. On peut aussi imaginer qu'ils s'impliqueront davantage dans la définition et l'évaluation des programmes de recherches (par exemple le développement d'un nouvel instrument) ainsi que dans la formation des futurs conseillers.

En conclusion, nous pensons que ces prochaines années verront probablement la profession de psychologue-conseiller chercher sa nouvelle voie en répondant à trois types de tentations : 1) une certaine volonté de préserver le *status quo* (utiliser des outils éprouvés de longue date et s'économiser ainsi l'apprentissage de nouvelles procédures), 2) le transfert de compétences (faire confiance aux spécialistes), et 3) le désir de renouveler sa panoplie d'instruments tout en garantissant la maîtrise et la compréhension de leur fonctionnement.

La dernière option nous paraît être la meilleure garantie de l'intérêt du travail des praticiens et de la qualité du service qu'ils rendent à la collectivité. Or nous savons bien à quel point elle est coûteuse en temps de formation, recyclage, séminaires, etc. Ceci nous rappelle que si l'évolution de la profession de psychologue-conseiller dépend effectivement du développement des idées et des techniques, elle est aussi (hélas) large-

---

ment tributaire de la situation économique du moment : des psychologues surchargés n'ont en effet que peu de temps à consacrer à l'apprentissage de nouveaux instruments, aussi séduisants soient-ils.

Pour conclure, et en réponse à notre intitulé, nous pensons avoir montré en quoi l'évolution de l'outil informatique peut changer la fonction et les modes de faire des psychologues conseillers. Notre opinion est que cette évolution des pratiques est aussi inexorable que la mainmise progressive de l'ordinateur sur des aspects de plus en plus nombreux et variés de notre existence : chacun d'entre nous se plie ou résiste à cette évolution en fonction de ses intérêts et de ses disponibilités. Par ce texte bien trop bref en regard de l'importance des enjeux identitaires de la profession, nous espérons avoir aidé les psychologues à justifier leurs choix.

## Références

- Bejar, I.I. (1990). A generative analysis of a three dimensional spatial task, *Applied Psychological Measurement*, 14, 237-245.
- Bejar, I.I., Chaffin, R. & Embretson, S. (1991). *Cognitive and psychometric analysis of analogical problem solving*. New York : Springer Verlag.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an Examinee's ability, In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, Ma : Addison Wesley Publishing Company.
- Capel, R. (1991). Quelques réflexions au sujet de l'informatisation du testage en psychologie. *Orientation et formation professionnelle*, 5, 14-17.
- Cardinet, J. & Tourneur, Y. (1985). *Assurer la mesure : Guide pour les études de généralisabilité*. Berne : Peter Lang.
- Cronbach, L., Gleser, G., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurement : theory of generalisability for scores and profiles*. New York : Wiley.
- Frederiksen, N. Mislevy, R.J. & Bejar, I.I. (Eds). (1993). *Test theory for a new generation of tests*. Hillsdale, New Jersey : Lawrence Erlbaum.
- Gendre, F. (1982). Les développements contemporains dans la construction des mesures psychologiques. *Revue internationale de psychologie appliquée*, 91-115.
- Gendre, F., Capel, R. & Salanon, A. (1997). O.R.U.S. Orientation raisonnée unifiée et synthétique. Manuel de l'utilisateur. *Actualités psychologiques édition spéciale*.
- Gulliksen, H. (1950). *Theory of Mental Tests*. New York : Wiley.

- 
- Hesketh, B. & Robertson, I. (1993). Validating personnel selection : a process model for research and practice, *International Journal of Selection and Assessment*, 1 (1), 3-17.
- Lavault, D. & Grégoire, J. (1997). *Introduction aux Théories des Tests en Sciences Humaines*. De Boeck & Larcier : Bruxelles.
- Lawshe, C.H., Bolda, R.A., Brune, R.L. & Auclair, G. (1958). Expectancy charts : their Theoretical Development. *Personnel Psychology*, 11, 545-560.
- Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading (Mass.) : Addison Wesley.
- Nunnally, J.-C. (1978). *Psychometric theory* (3rd ed.). New York, McGraw-Hill.
- Parker, R.C. & Miller, R.A. (1988). Using causal knowledge to create simulated patient cases : CPCS Project as an extension of Internit-1. In P.L. Miller (Ed.), *Selected topics in medical artificial intelligence* (pp. 99-115). New York : Springer Verlag.
- Plake, B.S. & Witt, J.C. (1986) (Ed.). *Buros Nebraska Symposium on Measurement and Testing*. Hillsdale, New Jersey : Lawrence Erlbaum.
- Rasch, G. (1960). *Probabilistic models for some intelligence and assessment tests*. Copenhagen : Danish Institute for Educational Research.
- Shye, S.(Ed.). (1978). *Theory construction and data analysis in the Behavioral Sciences : a volume in honor of Louis Guttman*. Jossey-Bass : San Francisco.
- Snow, R.E. (1990). New approaches to cognitive and conative assessment in education. *International Journal of Educational Research*, vol. 14, 5, 455-470.
- Spearman, C. (1904). The proof and measurement of association between two things. *American psychology*, 15, 72-101.
- Sternberg, R.J. (1977). *Intelligence, information processing and analogical reasoning : the componential analysis of human abilities*. New York : Wiley.
- Sternberg, R.J. (1985). *Beyond I. Q. A triarchic theory of human intelligence*. London : Cambridge University Press.
- Wood, R., & Skurnik, L.S. (1969). *Item Banking*. London : National Foundation for Educational Research,
- WRight, B. & Panchapakesan, S.(1969). A procedure for sample free item analysis. *Educational and Psychological Measurement*, 29, 23-48.