

## Mesure et évaluation en éducation



# *INTRODUCTION À LA PSYCHOMÉTRIE.* Anastasi, Anne (traduction de François Gagné). Guérin Universitaire, 1994, 278 pages

Louis Laurencelle

Volume 17, numéro 3, 1995

URI : <https://id.erudit.org/iderudit/1092317ar>

DOI : <https://doi.org/10.7202/1092317ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer ce compte rendu

Laurencelle, L. (1995). Compte rendu de [*INTRODUCTION À LA PSYCHOMÉTRIE.* Anastasi, Anne (traduction de François Gagné). Guérin Universitaire, 1994, 278 pages]. *Mesure et évaluation en éducation*, 17(3), 127–133. <https://doi.org/10.7202/1092317ar>

### **INTRODUCTION À LA PSYCHOMÉTRIE**

Anastasi, Anne (traduction de François Gagné).  
Guérin Universitaire, 1994, 278 pages.

Ce volume est une traduction partielle de *Psychological Testing*, le classique d'Anne Anastasi, qu'on retrouve en 6<sup>e</sup> édition chez Macmillan avec 20 chapitres et 817 pages. Le document français est disponible en couverture souple, sur laquelle a été reproduit le motif de l'édition américaine. Le traducteur nous propose seulement quelques chapitres, sept en tout, centrés sur la méthodologie des tests: de là provient le titre, «Introduction à la psychométrie», qui diffère substantiellement du titre anglais. Voici le sommaire des contenus:

Préface de l'auteure  
Préface du traducteur

**PREMIÈRE PARTIE: Le contexte du testage psychologique**

Chapitre 1 : Fonctions et origines du testage psychologique (22 p.)  
Chapitre 2 : Nature et utilisation des tests psychologiques (28 p.)

## DEUXIÈME PARTIE: Principes techniques et méthodologiques

Chapitre 4 : Les normes et l'interprétation des scores de tests (42 p.)

Chapitre 5 : La fidélité (34 p.)

Chapitre 6 : La validité: concepts fondamentaux (29 p.)

Chapitre 7 : La validité: mesure et interprétation (41 p.)

Chapitre 8 : L'analyse d'items (36 p.)

Références et index

### La traduction

La traduction, en un mot, est impeccable. Le texte, d'une langue toujours claire, constitue une traduction authentique plutôt qu'une translittération de l'anglais comme on nous propose parfois. Qui plus est, le traducteur a relevé l'ambiguïté ou l'erreur de certains passages du texte anglais et il s'est ingénié à les rectifier, au profit du lecteur et de l'auteur. Dans la préface du traducteur, il n'est pas surprenant d'apprendre que François Gagné (le traducteur) fait profession de psychométricien. Enfin, le traducteur et la maison d'édition ont produit de concert un volume de grande qualité, bien présenté, avec très peu d'erreurs typographiques.

### Les contenus

Que nous apprend Anastasi dans les sept chapitres de ce volume traduit? Il faut se rappeler que *Psychological Testing*, dont l'édition originale remonte à 1954, est un grand classique qui s'est vendu à des centaines de milliers d'exemplaires. Le *Psychometric Methods* de J. P. Guilford (McGraw-Hill) était paru en 1936, le *Theory of Mental Tests* de H. Gulliksen (Wiley) en 1950. Ni le *Psychometric Theory* (McGraw-Hill, 1967) de J. C. Nunnally ni le *Statistical Theories of Mental Test Scores* (Addison-Wesley, 1968) de F. M. Lord et M. R. Novick n'avaient vu le jour. Anastasi s'adresse à des étudiants engagés dans une formation de psychologue et elle leur explique d'où viennent, comment sont faits et ce que sont les tests psychologiques. Il ne faut pas chercher dans ce classique l'exposé rigoureux de Gulliksen, voire du récent Traub<sup>1</sup>. On n'y trouvera pas non plus le trésor méthodologique encore présent dans le *Psychometric Methods* de Guilford. L'ambition d'Anastasi, que *Psychological Testing* semble réaliser très bien au gré des consommateurs du livre, est résumée par l'auteur en quelques lignes, au chapitre 1:

[...] Une certaine connaissance [des] tests est essentielle pour pouvoir maîtriser adéquatement la plupart des champs de la psychologie contemporaine. C'est précisément dans ce but que ce volume a été rédigé. En effet, il n'a pas été conçu pour habiliter le lecteur à devenir un excellent

examineur et interprète des tests, non plus qu'un expert en élaboration de tests. Il ne s'adresse pas au spécialiste des tests, mais à l'étudiant de premier cycle en psychologie [...] (p. 5).

Si ce livre était publié aujourd'hui, pour la clientèle spécialisée et semi-spécialisée qui s'intéresse à la psychométrie et au *testing*<sup>2</sup>, que pourrait-on en dire?

La première partie, aux chapitres 1 et 2, brosse un portrait général des tests, leurs propriétés et leur utilisation. L'auteure nous trace une histoire schématique de la mesure appliquée en sciences humaines pour arriver, en passant bien sûr par Alfred Binet (1857-1911), à la panoplie de tests qu'on connaît aujourd'hui. L'apport de Francis Galton (1822-1911) est justement signalé, tout comme celui de K. Pearson (1857-1936) que l'auteure présente faussement comme «l'étudiant» de Galton. Cependant le travail de pionnier du belge Alphonse Quetelet<sup>3</sup> (1796-1874), sans lequel les apports ultérieurs eussent été problématiques, est totalement omis.

C'est la seconde partie, avec ses cinq chapitres, qui constitue la raison d'être du volume proposé par le traducteur puisque c'est ici qu'on trouve l'introduction à la méthodologie des tests. L'ordre des chapitres nous a semblé incongru. Dans notre perspective sans doute cartésienne, nous aurions commencé par la fabrication du test (chapitre 8), nous nous serions préoccupé de sa fidélité (chapitre 5) puis de sa validité (chapitres 6 et 7); enfin, une fois sa «validation» faite, la standardisation du test (chapitre 4) serait venue le préparer pour son usage général. Ceci dit, le texte a de grands mérites, qui justifient sa bonne réputation, ainsi que des faiblesses que j'ai trouvées surprenantes.

La matière exposée est riche et intéressante, à la fois pour le spécialiste en mesure et pour le futur praticien du *testing*. L'auteure n'hésite pas à aborder toutes les facettes d'une question, et l'on imagine bien une situation de classe dans laquelle cette question serait débattue entre étudiants et professeur. Elle paie un tribut à l'éducation en mentionnant entre autres «l'évaluation critériée» et les «tableaux de spécifications». Elle discute de l'application du concept de fidélité aux tests scolaires. La section du chapitre 4 portant sur les «normes développementales» m'a plu, de même qu'au chapitre 8, le traitement raffiné qu'elle donne du concept de «difficulté d'item». Enfin, l'exposé et la discussion qu'elle nous livre sur la validité des tests psychologiques et sur les méthodes de validation émanent d'une longue et profonde réflexion sur le sujet. Ces pages, dont l'intérêt réel est tout autre que technique, sont magistrales et valent à elles seules le prix du livre.

D'un autre côté, le flou que nous avons perçu dans le canevas historique du chapitre 1, nous le retrouvons un peu partout dans la matière même de la psychométrie présentée, voire dans les tableaux numériques qu'elle donne en référence. L'histogramme qu'elle nous présente en p. 57 comme typiquement normal ne l'est (significativement) pas, les centiles sont définis incorrectement (p. 67), le  $r^2$  n'est pas la proportion de variance commune à X et Y comme elle le prétend (p. 115), le coefficient de corrélation ne repose pas sur un postulat d'homoscédasticité (p. 164), la formule de l'erreur-type de prédiction n'est pas correcte (p. 165), *et caetera*. En marge des formules et du traitement quantitatif, les explications données à propos des concepts de «valeur vraie» et de «variance vraie» sont parfois du pur galimatias.

Nous concluons ce tour des contenus par trois cas d'espèce, le premier concernant une procédure de grande utilité pratique en clinique: le test d'une différence. Sommairement, Anastasi (en p. 126 et suivantes) propose d'interpréter la différence (X-Y) en la divisant par une erreur-type ( $s_{X-Y}$ ) afin de déterminer s'il y a «une véritable différence [...] entre [les] deux types d'habiletés (p. 128)»: la différence serait «véritable» si elle déborde environ deux unités d'erreur-type, *i.e.* si  $|X-Y| \geq 1,96s_{X-Y}$ . (L'auteure ne mentionne pas que les moyennes de population doivent être égales.) L'exemple donné concerne la différence entre le QI verbal et le QI non verbal d'une personne évaluée. Or, l'auteure propose d'utiliser la formule ÉT  $\sqrt{(2-r_{XX}-r_{YY})}$ ,<sup>4</sup> qui mesure précisément jusqu'à quel point la différence (X-Y) sera stable d'une évaluation à l'autre ou, si l'on veut, jusqu'à quel point elle est solidement éloignée de zéro. La question que se pose le clinicien n'est pas celle à laquelle répond cette formule: le clinicien se demande si la différence (X-Y), observée chez une personne, est remarquable ou exceptionnelle par rapport aux autres personnes, si elle dénote un écart anormal par rapport à la population. L'univers de généralisation n'est plus constitué seulement des *occasions de mesure*, mais plutôt des *personnes*, et l'erreur-type doit alors tenir compte de la corrélation (dans la population) entre X et Y (par exemple, entre le QI verbal et le QI non verbal). La différence critique proposée par Anastasi est d'environ 10 points. L'erreur-type correcte (à notre avis), qui incorpore une corrélation optimiste de 0,7 entre QI verbal et QI non verbal, engendre une différence critique d'environ 23 points<sup>5</sup>: avis aux cliniciens!

Les deux autres points que je veux relever concernent la validité prédictive, qu'Anastasi traite d'ailleurs avec une ampleur et, disons-le, un intérêt inaccoutumés. Le tableau 7-2 (en p. 174) présente les «différentiels de sélection»<sup>6</sup> simples (*i.e.* basés sur un seul test) attendus pour des

«rapports de sélection» (RS) allant de 0,05 à 0,95 et des coefficients de validité prédictive (VP, ou corrélations test/critère) allant de 0,0 à 1,0. Toute la première ligne du tableau, correspondant à RS = 0,05, est légèrement fautive. La quantité relative à VP = 1,0 est 2,08, et elle devrait être 2,063 (Guilford la donne déjà dans son tableau G, dans *Psychometric Methods*), et les autres inscriptions sur la ligne, obtenues en multipliant par VP, sont ainsi des proportions incorrectes de cette quantité. L'erreur provient simplement de la référence utilisée par Anastasi (Brown et Ghiselli, 1953), où l'on avait mal recopié les données de l'article original de Jarret, paru en 1948<sup>7</sup>.

Les «différentiels de sélection» doubles (*i.e.* basés sur l'emploi simultané de deux tests), qu'Anastasi présente ensuite à des fins de sélection et de classification, sont à la fois un atout original de ce texte et l'occasion d'une dernière critique spécifique. Au tableau 7-5 (en p. 191), l'auteure reproduit (correctement) des données issues de Brogden, dont voici la référence: H. E. Brogden, Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors, *Educational and Psychological Measurement*, vol. 11, 1951, p. 173-195. Si, comme moi, ils ne le connaissent pas, les psychométriciens et spécialistes en mesure découvriront là un confrère éminent et l'exposé brillant d'une procédure psychométrique de premier plan. Pour ces données, il s'agit donc du gain escompté par l'utilisation de deux tests pour sélectionner et placer un groupe de candidats dans deux postes: les paramètres sont: 1) le taux de sélection, 2) le coefficient de validité prédictive de chaque test pour le poste correspondant, 3) la corrélation entre les tests. Le tout est exprimé en unités standardisées, c'est-à-dire en écarts-réduits. J'ai potassé sur la théorie et refait les calculs de Brogden; le tableau qui suit présente mes résultats (nombres en italiques, à 3 décimales), à comparer à ceux de Brogden-Anastasi. Les quantités supposent un coefficient de validité prédictive de 0,5 et sont donc la demie du différentiel de sélection propre: Anastasi ni son traducteur ne donnent cette information.

Une inspection même furtive du tableau montre qu'on est loin du compte, sauf pour le cas plus simple d'une corrélation  $r_{XY}$  nulle, que les données publiées sont dramatiquement imprécises et que les efforts méritoires de calcul de Brogden *auraient dû être vérifiés et repris*. En fait, et cela résume notre appréciation détaillée des contenus, l'auteure (Anastasi) a mis entre les mains de centaines de milliers de lecteurs, étudiants et praticiens du *testing*, un document de référence émaillé de concepts flous, d'erreurs élémentaires, de données incorrectes, qu'on n'a pas pris la peine de vérifier ou de corriger depuis la première édition de *Psychological*

*Testing* en 1954 jusqu'à la sixième, en 1988 et à celle-ci, en 1994! Il est heureux qu'il ne s'agisse pas là d'un traité de pharmacologie, voire de mécanique automobile: la sécurité de tous en pâtirait!

### L'appréciation globale

Le jugement sévère que nous portons concernant le détail des contenus du livre ne doit pas nous faire perdre de vue ses qualités globales, en particulier les qualités globales du *Psychological Testing* de 817 pages, dont la version française nous communique seulement les premières parties. Il ne faut pas compter pour rien les centaines de pages suivantes et les douze chapitres qu'Anastasi consacre aux tests eux-mêmes et aux modalités psychométriques applicables dans les nombreuses rubriques du *testing* psychologique. Comme le révèle la citation en début de ce texte, Anastasi se soucie avant tout de former des psychologues, et ses chapitres psychométriques, que nous retrouvons dans le livre traduit, sont comme un préambule pour le corps du livre original. Il était donc audacieux de fragmenter ce livre, de constituer le «préambule», avec son âge réel et ses défauts, en un livre autonome et de le présenter aujourd'hui comme une *Introduction à la psychométrie*. Non pas, comme je l'ai dit, que la traduction soit mal faite, bien au contraire.

«Différentiels de sélection» doubles (multipliés par 0,5)  
tels qu'apparaissant dans le volume d'Anastasi (p. 191)  
et recalculés (en dessous et en italiques)

Rapport de sélection dans chaque emploi	$r_{XY} = 0$	$r_{XY} = 0,40$	$r_{XY} = 0,80$
0,05	1,03 <i>1,028</i>	1,01 <i>1,027</i>	0,96 <i>1,021</i>
0,10	0,87 <i>0,871</i>	0,84 <i>0,868</i>	0,79 <i>0,857</i>
0,20	0,68 <i>0,682</i>	0,65 <i>0,680</i>	0,59 <i>0,658</i>
0,30	0,55 <i>0,546</i>	0,50 <i>0,544</i>	0,43 <i>0,507</i>
0,40	0,42 <i>0,424</i>	0,37 <i>0,420</i>	0,29 <i>0,362</i>
0,50	0,31 <i>0,282</i>	0,25 <i>0,219</i>	0,17 <i>0,126</i>

NOTES

1. R. E. Traub: *Reliability for the social sciences*, SAGE Publications, 1994. Voir la recension dans cette revue, volume 17, n° 2.
2. Le traducteur (si ce n'est la reviseuse linguistique, Marie-Josée Guy) a pris, pour deux termes anglais consacrés: «testing» et «coaching», une option de francisation que nous trouvons discutable. D'une part, les vocables anglais sont passés dans l'usage, et la traduction «coaching» = «entraînement» est consacrée. D'autre part, le mot «testage» (pour «testing») n'existe pas dans ce sens et constitue un néologisme, une substantivation qui, par exemple, transformerait «parking» en «parkage» plutôt qu'en «stationnement». Quant à «monitorat» (pour «coaching»), le mot existe et signifie tout autre chose.
3. On doit notamment à Quetelet le concept quantitatif de «l'homme moyen», qui constitue la base axiomatique du *testing* à interprétation normative.
4. Cette formule suppose que les deux scores, X et Y, ont un même écart-type, symbolisé ÉT. Les fidélités de X et Y sont respectivement notées  $r_{XX}$  et  $r_{YY}$ .
5. Noter que, si les deux scores corrélaien parfaitement en valeurs vraies, *i.e.*  $r(V_x, V_y) = 1$  ou  $r_{XY} \approx \sqrt{(r_{XX}r_{YY})}$ , alors la formule d'erreur type:
 
$$2-2r(V_x, V_y) \sqrt{(r_{XX}r_{YY})}$$
 devient approximativement égale à celle donnée par Anastasi.
6. L'étiquette de «selection differential» semble attribuable à P. M. Burrows (Expected selection differential for directional selection, *Biometrics*, vol. 28, 1972, p. 1091-1100), quoique le concept remonte au moins à 1944 (Richardson), 1946 (Brogden) ou 1948 (Jarrett). Ces auteurs, tout comme Anastasi, s'y réfèrent au moyen d'une périphrase.
7. R. F. Jarrett, Per cent increase in output of selected personnel as an index of test efficiency, *Journal of Applied Psychology*, vol. 32, 1948, p. 135-145, et C. W. Brown et E. E. Ghiselli, Per cent increase in proficiency resulting from the use of selective devices, *Journal of Applied Psychology*, vol. 37, 1953, p. 341-344.

Louis Laurencelle  
Université du Québec à Trois-Rivières