



FICHES DE PSYCHO

# L'évaluation en psychologie

## Tests et questionnaires

10 fiches pour comprendre

*Outils de mesure standardisés, évaluation des capacités cognitives, personnalité, déontologie, méthodologie de l'évaluation...*

**Philippe Chartier**

• EDITIONS IN PRESS •

# **L'évaluation en psychologie**

## **Tests et questionnaires**

## **ÉDITIONS IN PRESS**

127, rue Jeanne d'Arc – 75013 Paris

Tél. : 09 70 77 11 48

E-mail : inline75@aol.com

**www.inpress.fr**

Collection dirigée par Lydia Fernandez, professeur en psychologie de la santé et du vieillissement, psychologue clinicienne, université Lyon 2, Institut de psychologie, département du PSED.

### *L'ÉVALUATION EN PSYCHOLOGIE. TESTS ET QUESTIONNAIRES.*

ISBN 978-2-84835-486-6

© 2018 ÉDITIONS IN PRESS

*Couverture : Elise Ducamp Collin*

*Mise en pages : Lorraine Desgardin*

Toute représentation ou reproduction, intégrale ou partielle, faite sans le consentement de l'auteur, ou de ses ayants droit ou ayants cause, est illicite (Loi du 11 mars 1957, alinéa 1<sup>er</sup> de l'article 40). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait une contrefaçon sanctionnée par les articles 425 et suivants du Code Pénal.

# L'évaluation en psychologie

## Tests et questionnaires

Philippe Chartier



*Une pensée pour mes enfants, Juliette et Louis ainsi que pour mes parents, Suzanne et Gabriel.*

*Un grand merci à Maryline qui m'a soutenu lors de la rédaction et pour tout ce qu'elle m'apporte.*

*Je dédie cet ouvrage à la mémoire de Pierre Vrignaud, professeur de psychologie, collègue et ami, auprès de qui j'ai beaucoup appris.*

## L'AUTEUR

**PHILIPPE CHARTIER** est maître de conférences en psychologie au CNAM (Conservatoire National des Arts et Métiers) de Paris. Il est rédacteur en chef de la revue *l'Orientation Scolaire et Professionnelle* de l'INETOP (Institut National d'Étude du travail et d'Orientation Professionnelle). Chercheur au CRTD (Centre de Recherche sur le Travail et le Développement) du CNAM, ses recherches portent principalement sur la méthodologie de l'évaluation, l'élaboration et l'utilisation de tests et questionnaires. Il est l'auteur du test de raisonnement RCC, de plusieurs ouvrages et coordonne le groupe GEP (Groupe sur l'Évaluation des Personnes) de l'INETOP.



# Sommaire

Introduction.....	9
<u>Fiche 1</u>	
Qu'est-ce qu'un outil de mesure standardisé (tests et questionnaires) ? Éléments de psychométrie .....	11
<u>Fiche 2</u>	
L'élaboration et la validation d'un dispositif d'évaluation (test ou questionnaire).....	23
<u>Fiche 3</u>	
L'évaluation des capacités cognitives (1) : les tests de type Q.I. ....	33
<u>Fiche 4</u>	
L'évaluation des capacités cognitives (2) : les tests factoriels .....	41
<u>Fiche 5</u>	
L'évaluation de la personnalité .....	53
<u>Fiche 6</u>	
L'évaluation des intérêts professionnels.....	65
<u>Fiche 7</u>	
L'approche de l'évaluation en neuropsychologie.....	75



## Fiche 8

Illustration d'une procédure d'évaluation dans le cadre d'une activité de conseil en orientation : de l'analyse de la demande à la restitution des résultats .....	85
---	----

## Fiche 9

Évaluer un outil d'évaluation ? .....	93
---------------------------------------	----

## Fiche 10

L'utilisation des tests : aspects pratiques et déontologiques .....	103
Conclusion .....	111

# INTRODUCTION

Qu'est-ce qu'un test, un questionnaire, une évaluation standardisée ? À quoi peuvent-ils servir ? Sont-ils toujours fiables ? Comment sont-ils élaborés ? Autant de questions auxquelles cet ouvrage souhaite apporter des éléments de réponse. Issue de la psychologie différentielle, la méthode des « tests » consiste à « *développer des dispositifs d'observation qui permettent de mettre en évidence des différences entre les individus* »<sup>1</sup> (Huteau, 2010, p. 323).

À travers 10 fiches, ce livre présente déjà les éléments méthodologiques relatifs à l'évaluation des personnes puis aborde les différents domaines d'application possible (l'intelligence, les intérêts, la personnalité...) avec des exemples d'épreuves. Après une illustration d'une évaluation menée dans un contexte de conseil en orientation, l'ouvrage se termine par la présentation des aspects pratiques et déontologiques.

Ce livre s'adresse aux étudiant-e-s en psychologie et en sciences humaines, ainsi qu'aux psychologues et praticien-ne-s de ces domaines mais également à toute personne (enseignant-e, parent...) souhaitant découvrir les bases de l'évaluation par les tests.

Les différentes fiches ont été conçues comme une première introduction aux thématiques complexes, abordées ici de façon synthétique. Nous invitons le-la lecteur-ric-e intéressé-e à compléter sa lecture à partir des références bibliographiques disponibles à la fin de chaque fiche.

---

1. Huteau, M. (2010). Origines et évolution de la psychologie différentielle. In J. Lautrey (ed.), *Psychologie du développement et psychologie différentielle* (pp. 289-322). Paris : PUF.



# QU'EST-CE QU'UN OUTIL DE MESURE STANDARDISÉ (TESTS ET QUESTIONNAIRES) ? ÉLÉMENTS DE PSYCHOMÉTRIE

Nous présentons dans cette fiche les principaux éléments méthodologiques qui permettent de définir ce que doit être une épreuve d'évaluation psychologique (test ou questionnaire) : un outil de mesure qui doit présenter des qualités métriques (sensibilité, fidélité, validité) qui doivent avoir été validées par des études expérimentales (notion de psychométrie). Une épreuve ne présentant pas des preuves de ces qualités ne peut donc pas être qualifiée de test car aucune garantie n'est donnée concernant la fiabilité de cet outil d'évaluation.

---

• validité • psychométrie • mesure • fiabilité •

Les dispositifs d'évaluation en psychologie, tests et questionnaires reposent sur une discipline nommée psychométrie (Hogan, 2017). Le principal objectif de cette discipline est de fournir un cadre théorique et méthodologique pour élaborer un dispositif d'évaluation qui soit le plus objectif possible. Ainsi « *les méthodes psychométriques offrent un éclairage complémentaire, souvent plus précis, présentant un caractère plus systématique et plus étayé que ce que ne sont en mesure de fournir un entretien ou une observation spontanée* » (Bernaud, 2014, p. 5).

Comme nous allons le voir ensuite, la psychométrie propose des critères précis pour déterminer si une épreuve d'évaluation présente bien les qualités métriques attendues : sensibilité, fidélité

et validité<sup>1</sup>. C'est à ces conditions que l'on pourra alors parler réellement de test au sens psychométrique.

Historiquement, la méthode des « tests » relève de la psychologie différentielle, discipline centrée sur l'analyse des différences intra- et interindividuelles. Sans pouvoir reprendre ici tous les éléments historiques de cette approche (voir Reuchlin, 1997), signalons juste que l'élaboration de tests et questionnaires a été pendant plusieurs décennies l'application principale de cette sous-discipline de la psychologie. En premier lieu, les tests dit d'intelligence, à la suite des travaux de Binet et Simon, puis des tests utilisés dans la sélection de personnel et/ou à des fins d'orientation, mais également par la suite des épreuves dans les domaines de l'éducation, de la santé, etc.

Quel que soit le format de présentation de l'épreuve, test ou questionnaire, une épreuve psychométrique doit donc présenter certaines caractéristiques que nous allons maintenant présenter.

## 1. Une situation standardisée

Qu'est-ce qu'une évaluation standardisée ? C'est une évaluation qui va respecter certaines règles ou conditions, tout au long de la procédure d'évaluation : de la passation de l'épreuve à la cotation des résultats. Ainsi la passation, comme la cotation, doivent être effectuées selon les indications du manuel du test (indiquons à ce propos qu'un test doit toujours être accompagné d'un manuel qui doit fournir un certain nombre d'informations sur l'épreuve<sup>2</sup>) afin que chaque personne soit placée dans les mêmes conditions et garantir alors la comparabilité des résultats.

Pour la passation par exemple, il est fréquemment indiqué qu'elle doit se dérouler dans une pièce calme (sans bruits ni

---

1. Nous détaillerons ces qualités plus loin dans cette fiche.

2. Voir la présentation du manuel dans la fiche 9 de cet ouvrage.

nuisances...) afin que le·la participant·e soit concentré·e et impliqué·e au maximum.

Matériel (la personne peut-elle utiliser un brouillon, un objet...), conditions d'utilisation (temps libre ou limité...) et comportement de l'évaluateur·rice sont aussi soigneusement décrits et ces éléments doivent être respectés.

L'objectif principal de la standardisation de la situation d'évaluation est bien de permettre la comparabilité des résultats, de s'assurer que les différences observées relèvent bien de caractéristiques de la personne et non pas de différences dans la passation ou de la subjectivité de l'évaluateur·rice.

## 2. Une épreuve structurée à partir d'items et de scores

1

L'épreuve d'évaluation, test ou questionnaire, est composée d'une série de différentes situations : petits problèmes à résoudre dans le cas de tests cognitifs, questions ou propositions sur lesquelles la personne doit se positionner dans le cas de questionnaires... Chacune de ces situations est appelée *item*. À la suite de la passation chaque item sera coté à partir de règles précises. L'ensemble des items permettra le calcul de scores. Le choix des items (contenu, présentation, modalités de réponse...) va dépendre des objectifs du constructeur de l'épreuve. Un test va donc comporter un certain nombre d'items, organisés autour d'un ou de plusieurs scores (ou échelles).

## 3. Une épreuve fiable (fidèle)

« Une mesure est dite fiable, ou fidèle, lorsqu'elle n'est pas entachée d'erreurs trop importantes » (Huteau, 2010, p. 340). L'un

des cadres conceptuels de la psychométrie est la théorie dite du « score vrai<sup>3</sup> » qui distingue le score observé (le score obtenu au test) et le score vrai (la véritable caractéristique de la personne, non observable). Le score au test n'étant qu'une estimation du score vrai, il peut exister un écart entre ces deux scores.

L'erreur de mesure est la différence entre score observé et score vrai. L'application des conditions de standardisation vise à réduire cette erreur au minimum. Une mesure sera fidèle si elle présente un minimum d'erreurs. On considère habituellement plusieurs sources d'erreur :

- celle *attribuable à la personne* : elle peut, par exemple, être plus ou moins impliquée dans l'évaluation, plus ou moins stressée... et ces éléments peuvent influencer ses réponses ;
- celle *attribuable à la situation*, comme un défaut de standardisation ;
- celle *attribuable à l'outil de mesure*, comme un test qui comporte un nombre trop faible d'items, ce qui peut nuire à la qualité du score.

On distingue classiquement deux grandes formes de fidélité : la fidélité temporelle et la fidélité interne.

La *fidélité temporelle* concerne la stabilité de la mesure. Elle peut être vérifiée par un test-retest auprès des mêmes personnes. Comme le suggère son nom, il s'agit de faire passer la même épreuve à deux occasions, par exemple à une semaine d'intervalle afin de vérifier qu'on obtient, globalement, des résultats proches. On utilise pour cela un coefficient de corrélation. Si la mesure était parfaite, on obtiendrait exactement les mêmes résultats, soit un coefficient de 1. Ce n'est pas le cas en raison, justement, de l'erreur de mesure. Bien qu'il n'existe pas de seuil officiel, on peut considérer que la fidélité est correcte à partir de

---

3. Il existe différentes théories de la mesure, celle-ci est la plus classique et la plus fréquente

0,70, satisfaisante à partir de 0,80 et élevée au-dessus de 0,90 (Bernaud, 2014, p. 96).

La *fidélité interne*, ou cohérence interne, concerne le niveau d'homogénéité des items d'une échelle. L'indice utilisé ici est le plus souvent le coefficient alpha de Cronbach. Pour son interprétation on peut reprendre les seuils indiqués plus haut.

Une possibilité de tenir compte de cette erreur dans les pratiques d'évaluation est d'entourer le score observé d'un intervalle de confiance, pratique recommandée par exemple dans l'utilisation des tests de Wechsler<sup>4</sup> WISC et WAIS. Par exemple pour un score observé de Q.I. de 96, le·la praticien·ne est incité·e à ne pas donner cette valeur unique mais à indiquer que le niveau intellectuel de la personne se situe dans l'intervalle de confiance [91-101] (Grégoire, 2004, p. 200).

## 4. Une épreuve valide

1

Un test est valide s'il mesure bien ce qu'il est censé mesurer. C'est la principale question concernant l'analyse de la qualité d'un test : qu'est-ce qu'il mesure ? Et quels éléments objectifs sont apportés sur ce plan ? Effectivement, un test peut présenter de bonnes qualités métriques au niveau de sa fidélité mais s'il ne mesure pas réellement ce qu'il affirme évaluer quelle serait alors son utilité ? Le·la praticien·ne doit trouver dans le manuel du test des informations détaillées sur cet aspect de la mesure, ainsi que des indications lui permettant d'interpréter les scores observés et d'élaborer des hypothèses sur les caractéristiques de la personne évaluée.

Plusieurs formes de validité sont différenciées (pour une présentation plus détaillée voir Chartier & Vrignaud, 2018 ; Chartier, 2013) :

---

4. Voir la présentation de ces tests dans la fiche 3 de cet ouvrage.



- la *validité de contenu* : il s'agit ici de vérifier que les items du test reflètent bien ce qui doit faire l'objet de l'évaluation (raisonnement logique, traits de personnalité...);
- la *validité critérielle* : elle concerne les liaisons entre ce qui est évalué par l'épreuve et un critère externe, comme la réussite à une formation;
- la *validité prédictive* est sans doute la forme de validité la plus utile pour le-la praticien-ne car elle permet, à partir de ses résultats à l'épreuve, de prédire, avec une certaine marge d'erreur, la conduite de la personne (niveau de performance, niveau de satisfaction, mode privilégié de relation...) lorsqu'elle est placée dans une autre situation;
- la *validité conceptuelle* (validité théorique) : elle repose sur les liaisons attendues entre une épreuve et un modèle théorique de référence. C'est cette forme de validité qui va permettre de donner du sens aux scores observés, de poser des hypothèses et de généraliser, toujours avec prudence, les caractéristiques de la personne à un ensemble de situations définies. Par exemple, dans le cas le cas d'un questionnaire d'intérêt reposant sur le modèle de Holland<sup>5</sup> l'épreuve doit fournir les six dimensions (six scores) postulées par le modèle RIASEC;
- la *validité de conséquence* : l'accent est mis ici sur les implications souhaitées et non souhaitées de l'usage des tests, en particulier l'analyse des possibles effets indésirables (Laveault et Grégoire, 2005, p. 166). Cet aspect du test, qui concerne à la fois des considérations théoriques et pratiques, doit être étudié par le-la praticien-ne qui doit toujours évaluer la pertinence d'utiliser un test particulier dans un contexte spécifique;
- la *validité synthétique* (ou écologique) : « À quoi bon un questionnaire aux qualités psychométriques sans failles si le psychologue a du mal à l'interpréter et si le sujet ne

---

5. Voir la présentation de ce modèle dans la fiche 6 de cet ouvrage.

*comprend pas les résultats qui lui sont communiqués!* » (Vrignaud & Bernaud, 2005, p. 207). Cette forme de validité intègre les aspects de l'utilisation de l'outil d'évaluation dans les phases d'interprétation, de communication des résultats et de leur intégration par la personne évaluée. Le fait de disposer d'études de cas dans le manuel, de guide et/ou de livret de restitution des résultats à proposer à la personne permet d'assurer un bon niveau de validité synthétique ou écologique en fournissant des informations et/ou documents permettant d'éclairer et de faciliter l'utilisation de l'épreuve par le-la praticien-ne et l'intégration des résultats par la personne évaluée ;

- la *validité apparente* : elle n'a pas de valeur au sens psychométrique et ne repose sur aucun argument. Il s'agit uniquement ici d'une analyse superficielle de l'épreuve qui ne garantit aucunement sa fiabilité.

## 5. Une épreuve permettant de situer la personne évaluée dans un groupe de référence (étalonnage)

Comme nous l'avons indiqué, une épreuve comporte un certain nombre de questions ou de situations (les items). Chaque item fera l'objet d'une cotation (par exemple 0 ou 1 point le plus souvent pour des tests cognitifs). Le-la praticien-ne va ensuite sommer ces scores pour calculer le score total à l'épreuve (ou plusieurs scores selon la structure de l'épreuve). Ce premier score, appelé *score brut* ne va prendre sens que lorsqu'il sera confronté à des scores de référence : c'est l'objet de l'étalonnage qui va permettre de transformer ce score brut en un score étalonné.

L'étalonnage est la distribution des scores observés à l'épreuve sur un échantillon de personnes qui va servir de référence, de

norme (d'ailleurs on parle aussi de test *normé*) pour situer la personne évaluée.

L'exemple suivant permettra d'illustrer cette démarche. Il s'agit d'un test de raisonnement logique, le test RCC qui comporte 40 items (Chartier, 2012). Chaque item est coté 0 (mauvaise réponse) ou 1 (bonne réponse). En faisant la somme des items on obtient le score total RCC qui peut donc varier de 0 à 40. Que signifie par exemple un score observé (appelé score brut) de 35 ? À l'inverse d'une évaluation scolaire classique, notée de 0 à 20, avec la « moyenne » fixée à 10 qui sert de référence, on ne dispose pas ici de telles références. L'objectif d'un étalonnage va nous fournir ces références en nous indiquant les résultats observés sur un échantillon de référence (ici des personnes ayant un certain niveau d'étude) avec la répartition des scores étalonnés selon un pourcentage défini (ici 10 % des personnes dans chaque classe étalonnée, nous avons donc 10 scores, ou classes étalonnées). La lecture de cet étalonnage nous indique que, par exemple, un score brut compris entre 29 et 34 (qui correspond aux scores étalonnés 4 à 7) peut être considéré comme un score moyen et qu'un score de 37 et plus peut être considéré comme un très bon score (score étalonné de 10, catégorie ++), atteint seulement par 10 % des personnes de cet échantillon (voir *tableau 1*).

Reprenons notre exemple : grâce à cet étalonnage le score observé (score brut) de 35 va prendre sens. Il correspond à la classe 8 en score étalonné ce qui signifie que :

- 1. Ce score est situé ici au-dessus de la moyenne (catégorie +) et peut correspondre à un bon niveau de raisonnement *comparativement* aux personnes de même niveau de qualification ;
- 2. Seulement 20 % des personnes de ce niveau de qualification obtiennent un score plus élevé (classes 9 et 10) ;
- 3. 10 % des personnes de ce niveau de qualification obtiennent un score comparable ;

scores bruts	0-22	23-26	27-28	29-30	31	32-33	34	35	36	37 et +
pourcentage théorique	10	10	10	10	10	10	10	10	10	10
classes étalonnées	1	2	3	4	5	6	7	8	9	10
interprétation	--	-	-	moyenne				+	++	

**Tableau 1. Exemple d'un étalonnage RCC, de type décalage, pour un niveau inférieur au baccalauréat (extrait de Chartier, 2012).**



Tests, questionnaires, outils de mesure... L'évaluation est aujourd'hui incontournable en psychologie. Elle vise principalement deux objectifs : fournir au psychologue des informations objectives sur les caractéristiques des personnes (capacités cognitives, personnalité, intérêts professionnels...) et permettre au consultant une réflexion sur soi (aide au développement personnel).

De l'analyse de la demande à la restitution des résultats, ce livre aborde les différentes étapes de l'évaluation et permet d'explorer les rapports que le psychologue clinicien entretient avec la norme et les données scientifiques. De nombreuses questions éthiques et déontologiques entourent donc cette pratique centrale dans l'exercice professionnel du psychologue.

En 10 fiches synthétiques et toujours accessibles, cet ouvrage permet de faire le point. Chaque fiche est composée d'un résumé et de mots-clés, d'une partie générale et d'illustrations (cliniques, exemples). Les différentes fiches mettent en lumière la variété et la richesse des travaux et des pratiques du psychologue dans ce champ. Un livre-outil destiné aux étudiants en psychologie, aux psychologues et à tous les cliniciens.

**Philippe Chartier**, maître de conférences en psychologie au CNAM/INETOP/Paris.

**Collection dirigée par Lydia Fernandez**



9 782848 354866

ISBN : 978-2-84835-486-6

10 € TTC – France

[www.inpress.fr](http://www.inpress.fr)