

Pratique professionnelle

Tests et données probantes, un état de situation



Pierre Desjardins / Psychologue

Directeur de la qualité et du développement de la pratique

pdesjardins@ordrepsy.qc.ca

_L'ÉVALUATION PSYCHOLOGIQUE BASÉE SUR LES DONNÉES PROBANTES

Meyer *et coll.* (2001) s'appuyant sur trois recherches différentes, rapportent que les psychologues, notamment les cliniciens, accordent à la psychothérapie et à l'évaluation psychologique, dans l'ordre, la première et la deuxième place en importance dans leur pratique professionnelle. Il n'y a pas de raison de croire que ce soit différent au Québec, d'ailleurs le projet de loi n° 50 reconnaît en somme l'engagement des psychologues dans ces activités professionnelles en proposant de les leur réserver.

Il y a quelques années déjà qu'on se consacre à identifier les traitements psychothérapeutiques efficaces (*evidence-based practice*). Or, la nécessité reconnue par tous d'offrir au public des services de santé valides et efficaces implique que l'on se penche également sur les processus d'évaluation, ce qui a donné lieu plus récemment à la recherche de données probantes sur cette importante activité professionnelle (*evidence-based assessment*)¹.

L'évaluation psychologique est une activité complexe qui s'appuie sur plusieurs compétences, dont les compétences en psychométrie². Bien que l'évaluation psychologique soit un processus de prise de décisions qui aille au-delà de la seule utilisation d'outils, il n'en demeure pas moins que le recours aux tests psychologiques constitue un apport significatif en la matière³. Toutefois, les tests ne sont certainement pas tous valides. Certains pourraient ne pas avoir de valeur psychométrique et d'autres, bien qu'excellents, pourraient être utilisés à tort.

_LA POLÉMIQUE ENTOURANT LES TESTS PROJECTIFS

D'entrée de jeu, il importe de rappeler que l'utilité générale des tests psychologiques est largement démontrée. C'est l'une des conclusions que tirent Meyer *et coll.* (2001), s'appuyant pour ce faire sur plus de 125 méta-analyses traitant de la validité des tests. Ils insistent également sur l'importance de recourir à plus d'une méthode d'évaluation, ceci permettant d'obtenir des informations différentes voire uniques, alors que de ne s'en tenir qu'à une seule source d'information ne pourrait permettre qu'une compréhension partielle de ce que l'on tente d'évaluer. Les tableaux comparatifs que l'on trouve dans leur publication sont éloquentes, notamment celui qui rapporte la validité prédictive de différents tests, de toute nature (auto-questionnaires, tests projectifs, tests d'intelligence, tests de personnalité, tests médicaux et autres)⁴.

En ce qui a trait aux tests projectifs, Lilienfeld *et coll.* (2003), insistant sur l'importance de disposer d'outils ayant une valeur scientifique, ont alimenté une polémique qui dure maintenant depuis plus d'une décennie⁵ en soutenant que ce type de tests ne seraient pas valides et que l'on ne devrait plus en faire usage.

De l'autre côté, Hilsenroth et Stricker (2004) ont fait une revue d'études et de recherches portant sur le Rorschach. L'une d'elles établissait qu'entre 1986 et 1996, le Rorschach a fait l'objet, en moyenne annuelle, de 95 articles scientifiques, ce qui le place au deuxième rang des tests de personnalité sur lesquels sont menées des recherches. Ils concluent par ailleurs que la fidélité et la validité du Rorschach sont au moins équivalentes à celles des autres tests. Le TAT n'est pas en reste non plus et, entre autres publications, il y a celle de Peters *et coll.* (2006) qui établit la validité du système de cotation « SCORS ».

La Society for Personality Assessment, consciente de la controverse sur les tests projectifs, s'est penchée sur le test de Rorschach et a pris en 2005 la position officielle suivante :

« (...) le test de Rorschach a, en matière de fidélité et de validité, des qualités similaires à celles des autres instruments reconnus pour l'évaluation de la personnalité, et l'utilisation responsable de cet outil à cette même fin est appropriée et justifiable⁶. »

_L'EFFET D'ALLÉGEANCE

On peut constater à la lecture de différents auteurs que chacun, dans sa position, en appelle à la rigueur scientifique et parfois même reproche à l'autre sa subjectivité, en alléguant que les convictions de l'autre ne reposeraient que sur un choix de recherches et d'études sinon partial, du moins partiel.

À cet égard, bien que notre code de déontologie nous enjoigne de ne pas discréditer sans fondements des méthodes psychologiques différentes de celles que nous utilisons, il demeure que la subjectivité joue de fait pour une part dans les prises de position de chacun. En effet, Norcross *et coll.* (2006) ont mené une recherche perceptuelle auprès de professionnels en santé mentale de diverses allégeances et ils ont relevé l'influence significative de l'orientation théorique dans la détermination du degré de discrédit accordé aux traitements ou aux tests psychologiques. Ils constatent également la tendance à se faire une opinion sur la valeur de certains traitements ou tests en se basant davantage sur ce qu'on en dit que sur l'état des connaissances actuelles sur le sujet, notamment quant à la validité réelle d'un test.

_LE CONCEPT DE VALIDITÉ

Le consensus est donc difficile à établir sur la validité des tests et ce qui permet de la déterminer. Le défi se complique d'autant plus que l'efficacité des tests ne se mesure pas qu'à partir d'un seul critère, qu'il y a une multitude de tests disponibles et qu'il s'en développe constamment des nouveaux⁷. Pour tenter d'y voir plus clair, il faut revenir sur le concept de validité et en référer aux publications scientifiques pertinentes.

De façon générale, la validité « (...) correspond au degré d'appui que les faits accumulés fournissent à l'interprétation recherchée des scores de tests pour le but proposé »⁸ Il n'y a pas qu'un type de validité à laquelle il est possible de se référer. Outre la validité apparente, il y a la validité de contenu (les différents items du test sont représentatifs de ce qu'on cherche à mesurer), la validité de construit (le test est cohérent avec la théorie sous-jacente), la validité de convergence (la capacité du test à mesurer un objet se compare à celle d'autres tests ou méthodes diverses reconnus valides pour mesurer ce même objet), la validité de discrimination (le test mesure bien ce qu'il est censé mesurer, sans mesurer autre chose) et la validité prédictive (le test permet de prévoir, par exemple, la réussite scolaire d'un enfant ou l'impact d'un traitement donné sur un client)⁹.

De plus, la validité scientifique d'un outil demeure conditionnelle à l'usage spécifique qu'on compte en faire et on ne peut faire abstraction, notamment, des objectifs poursuivis. Ces objectifs varient autant que les multiples mandats qui peuvent être confiés aux psychologues, dont, par exemple :

- :: confirmer un diagnostic psychologique (ce peut être en référence au DSM-IV TR ou à un autre manuel du même ordre);
- :: relever la présence ou l'intensité de symptômes;
- :: déterminer quel serait le meilleur traitement pour un client donné, dans un contexte donné¹⁰;
- :: mesurer l'impact du traitement offert à un client donné et surveiller son évolution¹¹;
- :: faire une sélection de clients ou de candidats à l'admission à un programme donné ou contribuer au processus d'embauche d'employés;
- :: mesurer ce qui ne peut être évalué directement en entrevue ou à l'aide de questionnaires que complète la clientèle (par exemple des processus physiologiques ou encore inconscients);
- :: mesurer rapidement ce que d'autres méthodes ne permettraient d'apprécier qu'au prix d'un plus grand investissement de temps et d'énergie;
- :: évaluer la valeur d'un traitement dans le but de recueillir des données probantes.

_L'OBJET DE LA MESURE

D'autre part, la validité d'un test ne peut être établie qu'à la lumière d'une définition claire et explicite de ce qu'il prétend mesurer. La difficulté est accrue lorsqu'il s'agit de mesurer des construits théoriques tels l'intelligence, la dépression, le leadership, la créativité. Dans tous les cas, la représentativité du test ne peut se juger que par rapport à la définition retenue ou convenue du construit mesuré, définition qui doit être la plus claire et la plus opératoire possible.

Joiner *et coll.* (2005) illustrent cette importance de bien comprendre et définir l'objet de mesure. Désireux d'établir une démarche probante d'évaluation de la dépression chez les adultes, ils dégagent à partir du concept de dépression ce qui devrait être mesuré : les symptômes associés à la dépression, les manifestations de l'humeur dépressive, les idéations suicidaires et autres.

_CONCLUSIONS

Les controverses en sciences sont monnaie courante et l'expression de points de vue différents, voire diamétralement opposés, est à la source même du développement scientifique. La polémique liée à l'utilisation des tests projectifs permet de dégager un certain nombre de constats et de recommandations soutenant un exercice professionnel prudent et rigoureux.

Les critères de validité des tests doivent, en premier lieu, être établis par des praticiens de tous les horizons. Ensuite, les tests doivent être évalués rigoureusement, en tenant compte de leurs qualités psychométriques propres et de leurs qualités relatives, en lien avec les protocoles d'évaluation dans lesquels ils s'insèrent. Par ailleurs, il faut distinguer la valeur du test en soi et l'usage qu'on peut en faire; ainsi le recours aux tests doit être approprié au mandat confié afin d'éviter que ne soit affectée sa validité. Enfin, il faut tenir compte de la valeur ajoutée¹² du test que l'on compte utiliser dans la mesure où ce test a sa valeur propre et en considérant l'importance de diversifier les sources d'information et les méthodes de cueillette de cette information pour bien étayer les conclusions.

Enfin, un test ne peut être discrédité dans l'absolu puisque l'usage qu'on en fait demeure relatif à la clientèle visée, au contexte de son utilisation, aux objectifs poursuivis et aux objets à mesurer. C'est au psychologue qu'incombe la responsabilité de choisir les bons tests et le devoir d'en démontrer la validité et la pertinence, non pas à partir de ses habitudes ou de ses croyances mais bien en se basant sur ce qu'en dit la recherche clinique et empirique, en tenant compte de l'état actuel des connaissances et en exerçant, cas par cas, son jugement professionnel.

Références

American Psychological Association (2003). *Normes de pratique du testing en psychologie et en éducation*. Montréal : Institut de recherches psychologiques.

American Psychological Association (2005). *Policy Statement on evidence-based practice in psychology*. Adresse URL : <http://www.apa.org/practice/ebpstatement.pdf>.

Antony, M. M., Rowa, K. (2005). Evidence-Based Assessment of Anxiety Disorders in Adults. *Psychological Assessment*, 17(3), 256-266.

Bartram, D. (2001). Guidelines for Tests Users : a Review of National and International Initiatives. *European Journal of Psychological Assessment*, 17(3), 173-186.

Desjardins, P. (2008). L'utilisation des tests. *Psychologie Québec*, 24(5), 14-15.

Desjardins, P. (2008). La psychothérapie : recherche et pratique. *Psychologie Québec*, 25(8), 14-15.

Hilsenroth, M. J., Stricker, G. (2004). A Consideration of Challenges to Psychological Assessments Instruments in Forensic Settings : Rorschach as Exemplar. *Journal of Personality Assessment*, 83(2), 141-152.

Hunsley, J., Mash, E. J. (2005). Introduction to the Special Section on Developing Guidelines for the Evidence-Based Assessment (EBA) of Adult Disorders. *Psychological Assessment*, 17(3), 251-255.

Hunsley, J., Meyer, G. J. (2003). The Incremental Validity of Psychological Testing and Assessment : Conceptual, Methodological, and Statistical Issues. *Psychological Assessment*, 15(4), 446-455.

Joiner, T. E. et coll. (2005). Evidence-Based Assessment of Depression in Adults. *Psychological Assessment*, 17(3), 267-277.

Lilienfeld, S. et coll. (2003). *Science and Pseudoscience in Clinical Psychology*. New-York : The Guilford Press.

Meyer, G. J. et coll. (2001). Psychological Testing and Psychological Assessment. A review of Evidence and Issues. *American Psychologist*, 56(2), 128-165.

Norcross, J. C. et coll. (2006). Discredited Psychological Treatments and Tests : A Delphi Poll. *Professional Psychology : Research and Practice*, 37(5), 515-522.

Peters, E. et coll. (2006). Reliability and Validity of the Social Cognition and Object Relations Scale in Clinical Use. *Psychotherapy Research*, 16, 617-626.

Society for Personality Assessment (2005). The Status of the Rorschach in Clinical and Forensic Practice : An Official Statement by the Board of Trustees of the Society for Personality Assessment. *Journal of Personality Assessment*, 85(2), 219-237.

Turner, S. M. (2001). APA'S Guidelines for Test User Qualifications. An Executive Summary. *American Psychologist*, 56(12), 1099-1113.

Notes

- 1 Hunsley et Mash (2005).
- 2 Bartram (2001), Turner et coll., (2001).
- 3 Voir Desjardins, P. (2007).
- 4 Ces tableaux permettent par exemple de constater que le Rorschach aurait une plus grande validité prédictive sur l'impact potentiel de la psychothérapie qu'un rayon X sur la détection de caries situées entre les dents ou que certains examens de la prostate ou mammographie pour détecter le cancer.
- 5 Society for Personality Assessment (2005).
- 6 Traduction libre, non officielle, tirée d'un article paru en 2005 dans la revue *Journal of Personality Assessment*.

- 7 Antony et Rowa (2005), entre autres, illustrent bien la complexité du défi que représente le développement de stratégies d'évaluation qui soient probantes et qui s'appuient sur des outils valides.
- 8 *Normes de pratique du testing en psychologie et en éducation*, p. 11.
- 9 Ceci n'est qu'une énumération de types de validité auxquels on fait le plus souvent référence, sans tenir compte des interrelations entre chacun de ces concepts.
- 10 Voir à cet égard l'énoncé de politique de l'American Psychological Association, énoncé que l'Ordre a fait sien et qui se trouve dans notre site Web.

- 11 Il n'est pas dit qu'un traitement, aussi validé qu'il soit sur le plan expérimental ou empirique, soit efficace pour un client donné dans un contexte particulier.
- 12 Hunsley, J. et Meyer, G. J. (2003) abordent cette question de « valeur ajoutée » d'un test (traduction libre de l'expression « incremental validity »). Ajoutons que si on utilise un test malgré le fait qu'il ne constitue pas une valeur ajoutée, on augmente la possibilité de le rendre inutile ou invalide alors que d'autres circonstances justifieraient après coup son utilisation auprès d'un même client.

PARE-CHOCS

Un programme d'intervention sur la dépression des adolescents qui a de l'impact !

- 84 %** des utilisateurs estiment que **PARE-CHOCS** contribue à réduire ou prévenir la dépression des adolescents.¹
- 100 %** des utilisateurs estiment que **PARE-CHOCS** aide à développer des habiletés de protection chez les adolescents.¹
- 87 %** des participants à une formation sur **PARE-CHOCS** disent être mieux informés des conséquences de la dépression sur le parcours scolaire.²

¹ Étude réalisée au printemps 2008 auprès d'utilisateurs de la trousse
² Synthèse d'évaluations d'ateliers offerts en 2007-2008

Quand la vie te rentre dedans!

Pour obtenir davantage d'information sur la trousse **PARE-CHOCS** et sur les formations offertes : www.ctreq.qc.ca/depression

En vente chez **Septembre éditeur**
1 800 361-7755
www.septembre.com

CTREQ
Le centre de transfert pour la réussite éducative du Québec