Document généré le 2 déc. 2024 16:35

Revue québécoise de psychologie

VALIDITÉ ET CRÉDIBILITÉ DANS LE CONTEXTE DE L'EXPERTISE EN NEUROPSYCHOLOGIE CLINIQUE

ASSESSING RESPONSE VAILIDITY AND CREDIBILITY WITHIN THE CONTEXT OF THE INDEPENDENT NEUROPSYCHOLOGICAL EVALUATION

Stephan Kennepohl

Volume 39, numéro 3, 2018

PSYCHOLOGIE LÉGALE : ENJEUX ÉPISTÉMIQUES, MÉTHODOLOGIQUES ET DÉONTOLOGIQUES
LEGAL PSYCHOLOGY: EPISTEMIC, METHODOLOGICAL AND ETHICAL CHALLENGES

URI : https://id.erudit.org/iderudit/1058184ar
DOI : https://doi.org/10.7202/1058184ar

Aller au sommaire du numéro

Éditeur(s)

Revue québécoise de psychologie

ISSN

2560-6530 (numérique)

Découvrir la revue

Citer cet article

Kennepohl, S. (2018). VALIDITÉ ET CRÉDIBILITÉ DANS LE CONTEXTE DE L'EXPERTISE EN NEUROPSYCHOLOGIE CLINIQUE. *Revue québécoise de psychologie*, *39*(3), 51–74. https://doi.org/10.7202/1058184ar

Résumé de l'article

Au cours des dernières décennies, une multitude de tests et d'indices ont été développés pour tenter de détecter la non-validité des résultats neuropsychologiques. Cependant, la prolifération et la popularité de ces mesures ont contribué à la perception qu'il est dorénavant possible de les utiliser pour identifier la simulation des troubles cognitifs (malingering). Dans cet article, un cadre conceptuel est présenté pour tenter de favoriser une meilleure compréhension des notions sous-jacentes et permettre une meilleure interprétation de la non-validité en expertise. Selon la présente conceptualisation, il existe deux étapes distinctes : la première consiste à discerner la non-validité des données à l'aide de méthodes psychométriques, notamment les tests de validité de performance (TVP). La deuxième requiert une décision quant à la crédibilité du tableau clinique, incluant un jugement quant à la cohérence de l'ensemble des informations cliniques. C'est seulement dans le contexte de certaines incohérences indéniables qu'il serait possible d'avancer une opinion évoquant l'amplification des problèmes cognitifs.

Tous droits réservés © Revue québécoise de psychologie, 2018

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

VALIDITÉ ET CRÉDIBILITÉ DANS LE CONTEXTE DE L'EXPERTISE EN NEUROPSYCHOLOGIE CLINIQUE

ASSESSING RESPONSE VAILIDITY AND CREDIBILITY WITHIN THE CONTEXT OF THE INDEPENDENT NEUROPSYCHOLOGICAL EVALUATION

Stephan Kennepohl[1]
Expertise Neurosciences

LE CONTEXTE D'EXPERTISE

Toute évaluation en neuropsychologie nécessite la formulation d'une opinion fondée sur des données objectives et fiables. Dans un contexte de pratique clinique, la responsabilité première du neuropsychologue est envers son client, l'objectif premier étant surtout de fournir, à lui et à son entourage, une vision plus juste de son fonctionnement. Tout en maintenant une vision plus objective, le clinicien demeure néanmoins « partial » et engagé envers le bien-être de son client. En revanche, l'expert n'exerce pas dans le même contexte et n'a pas de lien thérapeutique avec la personne évaluée. Il est plutôt mandaté pour découvrir la « vérité objective, factuelle ou historique, établir la réalité "vraie" » (Desjardins, 2009). Bref, son opinion doit être à la fois objective et *neutre*. Il peut même arriver que les conclusions d'une expertise nuisent à la cause défendue par la personne évaluée.

Dans le contexte d'une expertise, les partis impliqués ont souvent intérêt à défendre un point de vue, une certaine « réalité ». Il est tout à fait prévisible, donc, de se buter à des visions assez différentes, même contradictoires, de la réalité clinique. Assez fréquemment en expertise, les enjeux ne se limitent pas à la seule condition de l'individu (diagnostic) ou de ses conséquences (limitations fonctionnelles), mais englobent aussi des questions de « causalité » (p. ex., dans les questions de responsabilité d'un accident ou les raisons d'une invalidité).

Plus que dans un milieu clinique, l'expert doit donc maintenir un regard critique envers les résultats à sa disposition. Dans la mesure du possible, il devra faire tous les efforts pour obtenir la *totalité* des informations jugées pertinentes et nécessaires, le tout dans le but d'avoir une vision plus globale et équilibrée. Ceci comprend l'intégration de plusieurs sources, incluant non seulement les données obtenues lors de la passation des tests (scores aux tests, observations comportementales, questionnaires autorapportés) et les symptômes et plaintes subjectives rapportées par la personne évaluée

1. Adresse de correspondance : 101, avenue Laurier Ouest, Montréal (QC), H2T 2N6 Téléphone : 514-526-1648. Courriel : s.kennepohl@neurosciences.ca

(p. ex., lors de l'anamnèse) ou par les gens de son entourage, mais également les divers documents pertinents (p. ex., rapports médicaux et cliniques). Cependant, la présence de données potentiellement contradictoires, voire irréconciliables, pose un problème de taille : quelles informations sont vraiment représentatives de la réalité « vraie »? Bref, est-ce que notre évaluation est valide?

L'IMPORTANCE DE L'ÉVALUATION DE LA VALIDITÉ DES SCORES OBTENUS

En psychométrie, la *validité* peut se définir comme le degré de correspondance entre ce qui est mesuré et ce que l'on prétend mesurer. Dit autrement : mesurons-nous vraiment ce que nous croyons mesurer? On peut élargir le concept de validité d'un test pour inclure le contexte de l'évaluation individuelle, en se posant les mêmes types de questions : est-ce que les mesures qui supposent l'utilisation de certaines fonctions cognitives (p. ex., les tests « attentionnels ») évaluent vraiment ce que nous croyons évaluer, pour ce client en particulier? Peu importe les raisons sous-jacentes, si les scores obtenus lors d'une évaluation neuropsychologique reflètent surtout l'influence de facteurs *non cognitifs*, l'évaluation pourrait être jugée non valide. D'emblée, il serait important de souligner que la conclusion de non-validité ne permet pas *nécessairement* d'établir s'il s'agit de la production volontaire de faux symptômes (simulation). En fait, la non-validité reflète uniquement que les résultats ne sont pas indicatifs du fonctionnement cognitif réel de la personne évaluée. Cette question sera abordée de manière plus détaillée plus tard.

C'est André Rey qui aura été l'un des premiers à utiliser des tests spécifiques pour tenter d'évaluer si les personnes évaluées produisaient des résultats valides, en commençant par le Test de Comptage de Points (Carone et Bush, 2012; Rey, 1941). Dans le contexte de la présente discussion, il est intéressant de noter que Rey aurait développé ces tests en réponse aux inquiétudes de certains assureurs, préoccupés que certains de leurs clients puissent être tentés d'exagérer leurs difficultés cognitives (Fredericks, 2002). Comme tous les tests de validité, les tests conçus par Rey présentent certaines caractéristiques. À première vue, ces épreuves peuvent paraitre assez difficiles, mais nécessitent en réalité très peu de ressources cognitives pour leur réalisation. Par exemple, une tâche qui semble très exigeante sur le plan de la mémoire pourrait en fait être facilement réussie par la vaste majorité des gens en utilisant un sentiment de familiarité (mémoire implicite). Le Test de Comptage de Points, pour sa part, utilise deux conditions - une facile, une plus difficile - pour évaluer si la personne évaluée réussit mieux à la condition facile. Règle générale, ces tâches sont si faciles que les patients avec des troubles neurocognitifs connus (p. ex., déficience intellectuelle, traumatisme craniocérébral [TCC]

sévère) obtiennent le plus souvent au-delà de 90 % de bonnes réponses (Heilbronner *et al.*, 2009). L'obtention d'un score plus faible à un tel test, surtout dans le cas de gens qui présentent ou qui allèguent une atteinte de nature plus légère ou invérifiable, a donc pour effet de mettre en doute la validité de tous les autres scores aux tâches neurocognitives.

La nécessité d'utiliser des tests pour s'assurer de la validité de nos résultats devient d'autant plus évidente si l'on considère à quel point nous, comme cliniciens, sommes incapables de le faire sur la base du seul jugement clinique (Guillemette, 2012; Heaton, Smith, Lehman et Vogt, 1978). À titre d'exemple, une étude néerlandaise plus récente (Dandachi-Fitzgerald, Merckelbach et Ponds, 2017) suggère que des neuropsychologues à qui l'on demande d'identifier les profils non valides obtiennent un taux de faux positifs de 20 %. Si l'on se fie aux résultats de cette étude, le clinicien qui déclare que les résultats sont non valides se tromperait une fois sur cinq.

Cette réalité a contribué à la prolifération et à la popularité de différentes mesures qui tentent d'évaluer la validité des réponses (*response validity*). Il n'est pas exagéré de décrire ce phénomène comme un changement de paradigme en neuropsychologie, du moins si l'on en croit l'explosion de la littérature (Lippa, 2017). Deux grandes associations de neuropsychologues aux États-Unis ont d'ailleurs publié des énoncés (*position papers)* affirmant l'importance d'évaluer la validité des réponses, surtout dans un contexte psycholégal (Bush *et al.*, 2005; Heilbronner *et al.*, 2009). Bien que plus présente dans un contexte d'expertise, la question de la non-validité demeure une considération universelle, se manifestant dans les contextes où des avantages pourraient être tirés de l'évaluation (p. ex., évaluation d'une dyslexie dans un contexte universitaire, voir p. ex., Harrison, Edwards, Armstrong et Parker, 2010), mais également dans tout milieu clinique où l'on traite de problèmes plus complexes (p. ex., Williamson, Holsman, Chaytor, Miller et Drane, 2012), incluant la clientèle infantile (p. ex., Kirkwood, 2015).

LES TYPES DE VALIDITÉ

Pour l'expert, la source principale des informations pertinentes - et de leur validité - demeure la personne évaluée. Évidemment, ceci n'exclut pas la présence de biais de la part des autres partis, incluant la tendance de certains mandants de mettre en valeur leur point de vue. Demeure le fait que dans la majorité des situations d'expertise, c'est la personne évaluée qui détient le « fardeau de la preuve ». Pour obtenir gain de cause, c'est elle qui doit démontrer, de manière probante, la présence d'un trouble neurocognitif.

La validité des réponses peut être classifiée en deux grandes catégories, soit 1) le rendement à certains tests objectifs (p. ex., scores aux épreuves cognitives) ou 2) la description subjective de la réalité clinique (p. ex., symptômes et plaintes rapportées, réponses aux questionnaires autorapportés). Cette distinction nous permet de différencier entre les *tests de validité de la performance* (TVP; *Performance validity tests ou PVT*) et les *tests de validité des symptômes* (TVS; *Symptom Validity Tests ou SVT*) (Larrabee, 2012). En général, les TVP sont des tests qui ressemblent aux tests neuropsychologiques classiques, mais sont peu sensibles aux atteintes neurologiques les plus communes. Un score plus faible semble alors atypique ou peu probable, même dans le contexte d'un trouble cognitif plus sévère. Les TVS, quant à eux, sont généralement des questionnaires ou des entrevues structurées qui tentent d'évaluer si la description des symptômes rapportés est particulièrement exagérée ou inconstante.

Il est particulièrement important de considérer ces deux types de validité séparément, car il existe de plus en plus d'études démontrant que la même personne peut présenter des profils de réponses très différents aux TVP et aux TVS (p. ex., Nelson, Sweet, Berry, Bryant et Granacher, 2007; Van Dyke, Millis, Axelrod et Hanks, 2013). Bref, la même personne peut présenter un profil cognitif non valide (sur la base des TVP) tout en décrivant une symptomatologie cohérente (sur la base des TVS). L'inverse peut également être observé : une personne peut rapporter des symptômes cognitifs incohérents ou improbables (TVS), mais obtenir des résultats aux tests cognitifs jugés valides (TVP).

On peut aussi séparer les TVP en deux sous-types : soit les tests spécifiquement conçus pour détecter la non-validité du profil cognitif (ou TVP *spécifiques*; *stand-alone PVTs*) et les TVP « imbriqués » au sein des tests traditionnels (*embedded PVTs)*. Les TVP spécifiques n'ont qu'une seule raison d'être – évaluer la non-validité – et doivent être ajoutés à la batterie neuropsychologique standard. Pour être utiles, ces TVP doivent paraitre comme de vraies tâches cognitives, mais être réussis malgré la présence d'une neuropathologie plus sévère. Bien que la plupart des TVP spécifiques s'apparentent à des tests d'attention et de mémoire, il existe maintenant un choix d'épreuves associées à toutes sortes d'habiletés cognitives, incluant la vitesse de traitement (p. ex., Boone *et al.*, 2002), et la lecture (p. ex., Roberson *et al.*, 2007).

D'autre part, les TVP « imbriqués » (*embedded PVTs)* proviennent de l'analyse plus pointue de scores très faibles aux tests classiques, ou des « patterns » au sein du profil cognitif qui semblent assez inusités ou inhabituels dans le contexte d'un trouble neurocognitif (p. ex., attention sélective très fluctuante en l'absence d'un délirium). Ces TVP imbriqués ont plusieurs avantages. Il n'y a pas de coût surajouté ou de temps additionnel

et il est plus difficile pour la personne évaluée de s'éduquer au préalable quant à la « bonne » façon de répondre, que ce soit par le biais d'une évaluation antérieure, de l'Internet, ou d'articles ou livres publiés sur le sujet (voir Suhr et Gunstad, 2007). Bien que la notion de « coaching » des TVP pourrait paraitre comme un phénomène isolé, certains ont rapporté que des avocats jugent qu'il relève de leur devoir d'éduquer et de préparer leurs clients à la passation des TVP (p. ex., Youngjohn, 1995). Pour le moment, l'utilisation de TVP imbriqués ne semble pas près de remplacer les TVP spécifiques (notamment en raison d'une sensibilité moindre). Toutefois, ils demeurent une option moins coûteuse (en temps et en argent) et potentiellement plus adaptée à certaines nouvelles réalités.

Avant de procéder à une expertise, il est essentiel de bien se renseigner sur les différents tests de validité, leurs forces et leurs limitations. Il existe plusieurs écrits et articles qui décrivent les différents choix de TVP et TVS (p. ex., Boone, 2007; Carone et Bush, 2012; Lippa, 2017; Millis, 2009). Ce qui suit sera axé sur l'évaluation du trouble neurocognitif et les TVP, ceci étant la pierre angulaire de l'expertise en neuropsychologie.

ON MESURE QUOI, AU JUSTE?

Bien que l'utilité pratique des TVP ne soit plus vraiment à démontrer, il est intéressant de souligner la confusion des termes qui les décrivent. On pourrait même se demander ce que ces tests mesurent – en termes psychométriques, quelle est la *validité de construit* des TVP? En scrutant la littérature, on voit rapidement que ces mesures ont été associées à différents concepts, selon l'idéologie et la perspective des auteurs. Initialement décrits comme des tests de simulation (*malingering*), les TVP ont également été décrits comme des tests qui évaluent l'effort ou la motivation, le biais de réponse (*response bias*), la performance sous-optimale (*suboptimal performance)* ou le niveau de participation (*engagement*). À l'heure actuelle, du moins dans la littérature anglophone, il semblerait y avoir un certain consensus pour l'utilisation du terme *validity*, que ce soit pour décrire le rendement aux tests neurocognitifs ou les symptômes subjectifs (voir Lippa, 2017 ou Erdodi et Lichtenstein, 2017). Dans ce contexte, il est essentiel de réitérer que la notion de validité (validity) doit être comprise dans son sens le plus strict : elle représente ni plus ni moins le degré auquel nos tests mesurent réellement ce qu'ils ont été conçus pour mesurer. En généralisant cette notion à l'évaluation individuelle, la question fondamentale devient : est-ce que les résultats obtenus lors de *cette* évaluation reflètent vraiment le fonctionnement neurocognitif de la personne évaluée?

Pourtant, il est assez évident que ces tests ne sont pas une mesure directe de la validité. Il pourrait être tentant de conclure que la « réussite »

d'un TVP suggère un profil neurocognitif valide. Dans ce cas, la logique sous-jacente serait la suivante : Si la personne évaluée « passe » un TVP reconnu, alors on peut présumer qu'elle a fourni un effort constant et optimal tout au long de l'évaluation. Cependant, il y a plusieurs problèmes avec cette logique. Dans un premier temps, soulignons que ce sont les *données* qui sont évaluées pour leur validité, pas la *personne* elle-même. D'autre part, la *réussite* ou, plus précisément, le *non-échec* d'un TVP n'est pas une garantie de la validité de l'évaluation au complet. Un bon résultat à un TVP en matinée ne peut certifier que les données obtenues en après-midi seront valides. Et quelqu'un qui allègue un problème d'attention pourrait bien « choisir » d'obtenir un bon résultat à un TVP qui ressemble à un test de mémoire. Pour ces raisons, plusieurs attestent que la méthode optimale serait d'évaluer la validité *en continu* tout au long de l'évaluation, ou du moins inclure plusieurs TVP à différents moments de l'évaluation (Boone, 2009; Heilbronner *et al.*, 2009). Idéalement, ces différentes mesures ne devraient pas être trop corrélées entre elles, question de permettre une évaluation plus globale de la validité associée à différentes fonctions cognitives (Berthelson, Mulchan, Odland, Miller et Mittenberg, 2013). Même dans une situation optimale, la passation des TVP ne semble qu'assurer un *minimum* de validité, et ce, à des moments bien précis.

Plutôt, les TVP semblent surtout servir d'indicateurs de la *non-validité* des résultats cognitifs. En d'autres termes, c'est la non-validité qui devient la « condition d'intérêt ». Les raisons pour ceci sont importantes à considérer. La conclusion de non-validité peut avoir de lourdes conséquences pour la personne évaluée. Étant donné les risques, il semble nettement préférable de diminuer la probabilité d'obtenir un faux positif, soit de conclure de manière erronée qu'il s'agit d'un résultat non valide. Dans ce contexte, il semble nécessaire de maximiser le taux de *spécificité* des TVP*,* même au risque de diminuer leur *sensibilité*. [Petit rappel : la *spécificité*, c'est la proportion de ceux qui n'ont pas la condition et qui sont correctement exclus sur la base du même test; la *sensibilité*, pour sa part, représente la proportion de gens avec une condition d'intérêt qui sont correctement détectés sur la base d'un test.]

En misant sur la spécificité, on accentue l'importance d'obtenir un score « significatif » pour détecter la présence d'une condition d'intérêt (non-validité). Ceci ressemble au principe de SpIN, initialement proposé comme une des méthodes d'analyses pour le diagnostic différentiel (p. ex., Sackett, Strauss, Richardson, Rosenberg et Haynes, 2000; Smith Ivnik et Lucas, 2008). Ce principe propose que l'obtention d'un score significatif à un test spécifique (*Sp*) à une condition d'intérêt nous permet de conclure à la présence de la condition (*rule In*).

La logique sous-jacente serait la suivante : Si l'on ne note pas d'indicateurs francs de non-validité, alors les résultats obtenus peuvent être jugés minimalement valides. La question fondamentale serait donc : Est-ce que les scores obtenus sont non valides? Si la réponse est non, on peut procéder – du moins conditionnellement - à l'interprétation. Mais quel devrait être le point de bascule à partir duquel il est possible de se prononcer sur la non-validité?

LE CHOIX DU SEUIL CRITIQUE

Comme nous l'avons vu, il semble important de maximiser la *spécificité* des TVP pour tenter d'éviter de détecter de faux-positifs (rappelons ici que la condition d'intérêt demeure la *non-validité*). Le plus souvent, le seuil de spécificité est choisi au-delà de la valeur de 0,90 et même au-delà de 0,95 (Lippa, 2017). À ce stade-ci, il est important de souligner qu'une spécificité de 0,95 ne signifie pas qu'il y a une probabilité de 95 % que le profil est non valide si le score dépasse le seuil critique. En fait, une telle conclusion serait seulement applicable dans une situation purement théorique où la prévalence de la condition d'intérêt serait établie à 50 % très précisément (voir Chartrand et Kennepohl, 2016). Bien que plus élevée dans un contexte d'expertise (voir Mittenberg, Patton, Canyock et Condit, 2002), la prévalence demeure le plus souvent nettement en deçà du 50 %. Dans l'estimation de la probabilité de non-validité, il est donc primordial de considérer non seulement l'applicabilité de la mesure à la personne évaluée (soit les propriétés psychométriques du test), mais également la prévalence de la condition d'intérêt dans ce contexte précis. Nous discuterons de manière plus détaillée de ce phénomène en parlant du calcul des probabilités.

Le choix des TVP et des seuils critiques – et il est important de comprendre qu'il s'agit d'un choix - dépendent en grande partie de la personne évaluée et l'appariement avec les participants des études sur lesquelles les propriétés des tests sont établies. Entre autres, il serait déconseillé d'utiliser un TVP validé auprès d'un échantillon comportant des participants ayant subi un TCC léger si le motif de l'expertise est d'évaluer la possibilité d'une démence. À cet égard, soulignons que plusieurs des données psychométriques des TVP (p. ex., sensibilité, spécificité) ont été validées avec des groupes de « pseudo-simulateurs », soit des gens à qui l'on donne les consignes de « simuler » un trouble neurocognitif ou une condition spécifique (p. ex., un TCC). À d'autres moments, comme bien des tests psychométriques, la validité des TVP est basée sur leurs corrélations avec d'autres TVP auprès de groupes connus à risque de non-validité. Idéalement, il faut tenter de s'assurer que le TVP choisi ait une validité plus large, avec différents groupes (p. ex., « pseudo-simulateurs » et clientèles « à risque »).

La question du choix du seuil critique est d'autant plus importante dans le cas des TVP imbriqués, où les scores jouent un « double rôle », servant à la fois à la détection d'un déficit cognitif, mais également comme indicateur de non-validité. Dans certaines circonstances, il peut même arriver que le même score puisse être jugé à la fois trop faible pour être valide (« non-valide »), mais trop bon pour conclure à la présence d'un déficit (« normal »), un paradoxe nommé « invalid before impaired » par Erdodi et Lichtenstein (2017). Ces auteurs présentent l'exemple où le même score (au Trails B) pourrait être jugé non valide pour un individu plus jeune, mais dans la haute moyenne pour quelqu'un de plus âgé. Cette situation met en relief certaines incongruités plus générales liées à l'interprétation en neuropsychologie, incluant notamment la question du choix du seuil utilisé pour qu'un score soit jugé « déficitaire » (Erdodi et Lichtenstein, 2017).

Une piste de solution pourrait être de poser la question d'une manière plus probabiliste. Au lieu de se poser la question en fonction d'un seul score seuil, on pourrait se poser la question de la façon suivante : sachant que la personne évaluée possède certaines caractéristiques (p. ex., variables sociodémographiques, condition alléguée) et qu'il échoue un TVP, quelle est la *probabilité* que ses résultats cognitifs soient non valides?

L'UTILITÉ DES PROBABILITÉS

L'évaluation neuropsychologique peut être définie comme l'utilisation des tests psychométriques pour préciser nos connaissances concernant la présence ou l'absence d'une condition (Meehl et Rosen, 1955; Swets, Dawes et Monahan, 2000). Ces probabilités « conditionnelles » peuvent être décrites en fonction de la présence ou de l'absence d'une *condition d'intérêt*, que ce soit un trouble neuropathologique (p. ex., démence de type Alzheimer), un déficit neurocognitif (p. ex., amnésie), ou, comme dans le cas présent, une caractéristique plus générale (la non-validité). L'application la plus simple des probabilités conditionnelles consiste à : 1) considérer la probabilité initiale, 2) faire une mise à jour du contexte (en utilisant les résultats d'un test) et 3) calculer ou estimer la nouvelle probabilité (postérieure) qui s'applique (voir Équation 1).

Équation 1

Probabilité initiale (Prévalence)	→	Mise à jour (Test psychométrique)	= Probabilité postérieure

En consultant l'Équation 1, on note qu'une probabilité postérieure est « conditionnelle » au contexte (probabilité initiale). Cette probabilité « pré-évaluation » se définit plus simplement comme la prévalence de base de la condition dans ce contexte. En expertise, le « risque » d'obtenir des

données non valides est généralement plus élevé que dans un milieu clinique ou hospitalier. Certaines études à l'échelle nord-américaine ont suggéré que la prévalence (*base rate)* de non-validité dans un contexte médico-légal peut atteindre jusqu'à 30 % ou 40 %, selon la clientèle (voir Mittenberg *et al.*, 2002). Peu importe, sans le chiffrer précisément, l'expert devrait se faire une idée plus globale de la probabilité de non-validité dans son milieu, avant même de procéder à l'évaluation.

L'expert doit aussi prendre le soin de bien choisir le TVP. Dans l'évaluation de la non-validité, l'outil utilisé doit tenter de maximiser la *spécificité* tout en maintenant un niveau de *sensibilité* adéquat. Le choix doit se faire en consultant l'échantillon des études initiales, mais également en fonction du motif de consultation et de certaines caractéristiques de la personne évaluée, comme l'âge et le niveau de scolarité (Berry, Harp, Koehl et Coombs, 2017). L'utilisation d'un test peu spécifique avec un client avec une condition plus sévère (p. ex., début de démence, TCC sévère) pourrait mener à un faux positif. En revanche, l'utilisation d'une mesure trop peu sensible dans un autre contexte (p. ex., motif de TCC léger ou de TDAH) peut mener à trop de faux négatifs et à la surinterprétation de données non valides (p. ex., Armistead-Jehle et Hansen, 2011; Lippa, Lange, French et Iverson, 2017).

En principe, un TVP utile nous permet d'établir une probabilité postérieure qui s'éloigne le plus possible du 50 %, peu importe la direction. [NB : 50 % représente le point où l'on parle essentiellement de décider au hasard.] Encore une fois, réitérons que l'utilité pratique du TVP dépendra en grande partie du contexte. À titre d'exemple, si l'on présume que la prévalence de non-validité est de 30 %, un score qui ne dépasse pas le seuil critique à un TVP raisonnablement sensible pourrait bien servir à diminuer quelque peu la probabilité de non-validité. Toutefois, si l'on obtient le même résultat dans un contexte où la prévalence est plus faible (disons 5 %), la probabilité postérieure de non-validité deviendrait essentiellement nulle.

La probabilité postérieure peut être calculée en utilisant une équation relativement simple, utilisant des valeurs psychométriques connues ou assez faciles à estimer (voir Équation 2; cf. Lippa, 2017). En fait, on doit comprendre que cette équation n'est qu'une représentation plus formelle de l'Équation 1, où les différentes probabilités sont décrites en fonction de la prévalence de la condition d'intérêt ainsi que de la sensibilité et de la spécificité du TVP. Si le TVP est utile, la probabilité postérieure résultante nous permettra de prendre une meilleure décision quant à la non-validité, en incluant un indice du niveau de certitude (soit la distance du 50 %).

Équation 2
Rapport de cote prétest X Rapport de vraisemblance positif du test = Rapport de cote posttest
(*Pre-test Odds)* *(Positive Likelihood Ratio)* *(Post-Test Odds)*

Rapport de cote Prétest = Prévalence/(1 - Prévalence)
Rapport de vraisemblance positif du test = Sensibilité/(1 – Spécificité)
Rapport de cote Posttest = Probabilité postérieure/(1 – Probabilité postérieure)

Prenons un exemple. Présumons que la prévalence de non-validité dans notre contexte d'expertise est de 20 % (0.2). Si on utilise un TVP spécifique qui ressemble à un test de mémoire et qui aurait, en utilisant un seuil critique préétabli, un taux de sensibilité de 75 % (0.75) et un taux de spécificité de 95 % (0.95), quelle est la probabilité qu'un résultat qui dépasse le seuil critique représente un profil non valide? La réponse, en utilisant l'Équation 2, nous permet d'ajuster la probabilité initiale (prévalence) à une probabilité postérieure nous permettant de prendre une décision clinique. Dans notre exemple, la prévalence de 20 % devient une probabilité de près de 80 % (voir l'Exemple du calcul d'une probabilité postérieure, en Annexe). Toutefois, cet exemple démontre également comment l'échec à un seul TVP, même avec un taux de spécificité très élevé (0,95), peut mener à un résultat qui demeure insuffisant pour conclure à la non-validité (taux de faux positifs ≈ 20 %). En pratique clinique, des études confirment qu'avec certaines clientèles, il n'est pas rare d'observer l'échec d'un TVP en l'absence de toute notion évidente de non-validité. Par exemple, Meyers et Volbrecht (2003) suggèrent que près du quart (23,5 %) des gens ayant subi un TCC échoue au moins un TVP, même s'ils ne sont pas impliqués dans un litige. D'autres résultats suggèrent que les étudiants universitaires qui participent à des études de recherche peuvent également échouer un TVP à des niveaux s'élevant jusqu'à 5-18 % (An, Kaploun, Erdodi et Abeare, 2017).

Cependant, la situation semble bien différente si la même personne échoue plus qu'un TVP. Théoriquement, si l'on utilise des TVP indépendants (soit des tests qui ne sont pas corrélés), on peut alors « enchainer » les résultats des TVP (*chaining*), un après l'autre, pour établir une nouvelle probabilité postérieure (Larrabee, 2008; Meyers ***et al.***, 2014). En utilisant l'Équation 2, la probabilité postérieure obtenue à la suite du premier TVP échoué (TVP1) deviendrait la probabilité initiale pour un deuxième calcul à la suite d'un deuxième TVP échoué (TVP2). Par exemple, présumons que la même personne qui « échoue » le premier TVP qui ressemble à un test de mémoire obtient un score tout aussi faible lors d'un deuxième TVP qui ressemble à un test de lecture. Si ces deux TVP ont des propriétés psychométriques semblables, la probabilité de non-validité passerait d'un peu moins de 80 % (après TVP1) à plus de 98 % (après TVP2).

De tels résultats, en conjonction avec les résultats d'études empiriques, ont mené à l'établissement d'une « règle » plus générale suggérant que l'échec de ≥ 2 TVP reconnus peut servir comme seuil décisionnel pour établir la non-validité dans la vaste majorité des cas, surtout s'il s'agit de mesures raisonnablement indépendantes (Larrabee, 2008). Selon plusieurs auteurs, l'échec de ≥ 3 TVP constitue en quelque sorte une preuve plus ou moins incontestable de non-validité (Chafetz, 2011; Larrabee 2008; 2014; Victor, Boone, Serpa, Buehler, et Ziegler, 2009). Il persiste un débat quant à savoir si l'augmentation du nombre de TVP passés (spécifiques et imbriqués) au sein d'une batterie hausse de manière significative la probabilité d'un faux positif, comme ce serait nécessairement le cas des tests cognitifs plus traditionnels (voir Binder, Iverson et Brooks, 2009). Alors que certains suggèrent que la passation d'un grand nombre de TVP pourrait effectivement majorer le risque de faux positifs (Bilder, Sugar et Hellemann, 2014), d'autres affirment que les TVP n'ont pas cette propriété, du fait qu'ils ne sont pas de « vrais » tests cognitifs (Davis et Millis, 2014; Larrabee, 2014). Peu importe, il semble raisonnable de croire que l'échec ≥ 2 TVP peut suffire pour mettre en doute la validité des résultats cognitifs, dans la plupart des cas.

LA NON-INTERPRÉTATION DES RÉSULTATS NON VALIDES

Ceci nous mène à un constat particulièrement important : en situation de non-validité, les résultats cognitifs deviennent essentiellement non interprétables. Cette conclusion est basée sur des résultats empiriques, mais également des éléments de logique. Plusieurs études ont démontré que les résultats aux TVP sont fortement corrélés avec les résultats des autres tests neuropsychologiques (Fox, 2011; Green, Rohling, Lees-Haley et Allen, 2001; Iverson, 2005; Lange, Pancholi, Bhagwat, Anderson-Barnes et French; Meyers, Volbrecht, Axelrod et Reinsch-Boothby, 2011). La taille de cet effet peut même s'apparenter aux séquelles d'un TCC sévère ou d'une démence (Green *et al.*, 2001; Iverson, 2005). D'autre part, sur le plan de la logique de l'interprétation, il devra être établi que la personne évaluée possède toujours certaines capacités cognitives, ne serait-ce que pour s'assurer qu'elle peut suivre les consignes ou répondre aux questions de base. Ceci constitue en quelque sorte le « plancher » pour pouvoir différencier et interpréter un profil cognitif de manière plus précise. Il existe nécessairement une certaine « hiérarchie » au sein des épreuves cognitives – dit autrement, on doit généralement présumer que la personne obtiendra de meilleurs résultats aux tâches faciles qu'aux tâches complexes. La seule présence de scores trop faibles à des épreuves très faciles (les TVP) nous oblige à remettre en cause la possibilité d'effectuer une interprétation clinique. La fondation sur laquelle on doit bâtir notre conclusion n'est pas assez solide. Devant une telle situation, il devient essentiellement impossible d'émettre une opinion plus précise quant à la présence d'un

déficit neurocognitif, de commenter les séquelles ou la causalité en lien avec un évènement traumatique (p. ex., un accident de la route) ou de faire des recommandations quant à la nécessité ou à la pertinence d'interventions cognitives (Heilbronner *et al.*, 2009).

Dans plusieurs cas, la conclusion de non-validité peut être suffisante pour répondre à certaines des questions posées au sein d'un mandat d'expertise. Ceci est surtout vrai dans la situation où le « fardeau de la preuve » relève de la personne évaluée. Prenons, par exemple, la question suivante : « Sur la base de votre évaluation, êtes-vous en mesure d'objectiver la présence de déficits neurocognitifs directement attribuables au TCC léger? ». La réponse courte : non. La non-validité des données observées ne permet pas une interprétation raisonnablement précise du profil neurocognitif.

NON-VALIDITÉ ≠ NON-CREDIBILITÉ

Toutefois, rappelons que la conclusion de non-validité ne permet pas de conclure à la non-crédibilité du tableau clinique, et encore moins d'établir qu'il s'agit de la simulation. L'échec aux TVP n'explique pas le *pourquoi.* De nombreuses études dans le domaine ont documenté certaines conditions qui peuvent augmenter le risque de faux positifs aux TVP, notamment la démence (p. ex., Dean, Victor, Boone, Philpott et Hess, 2009) et la déficience intellectuelle (p. ex., Dean, Victor, Boone et Arnold, 2008) (voir également Lippa, 2017 pour une discussion plus détaillée). D'autres études suggèrent qu'une altération plus transitoire du fonctionnement cognitif peut aussi affecter les scores aux TVP, incluant une crise épileptique récente (Drane *et al.*, 2016) ou la prescription de certaines médications comme le lorazépam (Loring, Marino, Drane, Parfitt, Finney et Meador, 2011).

Même si l'échec des TVP semble attribuable à des facteurs non cognitifs, ceci ne mène pas nécessairement à la conclusion que ceci représente une amplification volontaire des difficultés cognitives. On peut penser aux fluctuations du niveau d'effort chez l'adolescent oppositionnel, certaines difficultés à comprendre les consignes (p. ex., quelqu'un évalué dans sa langue seconde) ou la présence d'une psychopathologie plus sévère (p. ex., épisode psychotique).

LES ÉTAPES DU PROCESSUS DÉCISIONNEL

Dans la majorité des cas, la détection de résultats non valides n'est que la première étape du processus. La deuxième nécessite l'interprétation des résultats des TVP dans le contexte de tous les autres éléments qui forment le tableau clinique, incluant les observations comportementales, les symptômes rapportés (incluant les TVS) et les autres éléments au dossier.

La question de base lors de cette deuxième étape devient : est-ce que le tableau clinique est *crédible*?

Pourquoi parler de *crédibilité*? Plusieurs descripteurs comme la simulation laissent entendre une certaine intentionnalité, une volonté, qu'elle soit consciente ou inconsciente, de fausser les données. En revanche, la notion de *crédibilité* nous renvoie à une interrogation plus générale, mais surtout plus neutre, nous incitant à nous questionner non seulement sur la véracité des données, mais également et surtout sur leur fiabilité et leur cohérence.

Selon cette conceptualisation, il serait possible d'observer des résultats manifestement non valides dans le contexte d'un tableau clinique qui retient une certaine crédibilité. Prenons l'exemple de quelqu'un chez qui l'on soupçonne une démence plus avancée. Le profil cognitif pourrait bien être largement non interprétable (du moins pour ce qui serait d'un profil détaillé), mais le tableau pourrait malgré tout être cohérent. Dans un contexte d'expertise, la personne qui présente une dépression très sévère pourrait fluctuer dans son niveau de participation à divers moments de l'évaluation, produisant des résultats qui sont non représentatifs de ses capacités cognitives (non-validité). Sur le plan comportemental, toutefois, un tel manque d'engagement pourrait retenir une certaine congruence, surtout si la validité des symptômes psychologiques demeure adéquate (p. ex., en utilisant des TVS).

En passant, cette distinction tient également pour la validité des symptômes. Prenons la personne qui présente un trouble neurocognitif sévère et une autocritique limitée (anosognosie). Le contenu de ses propos pourrait être jugé non valide, au sens où sa description de ses propres symptômes ne reflète pas sa réalité fonctionnelle. Mais le tableau clinique demeure crédible, dans la mesure où cette présentation demeure tout à fait cohérente avec l'ensemble des données.

LA NOTION DE COHÉRENCE

Alors que la première étape demeure une décision de nature plus psychométrique, la deuxième est nécessairement plus clinique. En fait, il n'existe pas d'indicateurs plus objectifs ou de référents quantitatifs pour établir la non-crédibilité des données (Heilbronner *et al.*, 2009). L'outil de base pour établir la crédibilité d'un tableau clinique demeure l'analyse de la *cohérence :* est-ce que les données cognitives sont compatibles avec les autres éléments du tableau? Est-ce que le tout se tient? La logique d'interprétation est la même que pour la validité : Si on ne détecte pas d'indicateurs francs de non-crédibilité, alors on présume que le tableau est minimalement représentatif de la réalité clinique.

Il existe différentes catégories d'incohérences nous permettant d'apprécier la non-crédibilité de la présentation clinique (Heilbronner *et al.*, 2009; Slick, Sherman et Iverson, 1999). Parmi ceux-ci, on peut nommer : i) les disparités évidentes entre le fonctionnement au quotidien et les résultats aux tests de validité (p. ex., ne rapporte aucune difficulté pour la conduite automobile, se rend au rendez-vous seul en voiture, mais fait preuve d'une lenteur cognitive excessive sur un TVP); ii) des incohérences marquées entre le rendement aux tests et la condition alléguée (p. ex., personne scolarisée qui obtient des résultats très faibles à des tâches langagières simples à la suite d'un TCC mineur); iii) des différences marquées entre les scores et les descriptions des gens de son entourage (p. ex., échecs aux TVP, mais rendement adéquat rapporté à domicile ou au travail); iv) des incohérences entre les scores aux tests et le comportement observé (p. ex., obtient un score particulièrement faible à un TVP de mémoire, mais décrit de manière détaillée les évènements survenus pendant la pause du diner); v) une évolution atypique de la condition ou des symptômes allégués (p. ex., à la suite d'un TCC, baisse importante du rendement cognitif ou majoration de la symptomatologie dans le temps, en l'absence de facteurs aggravants).

Il est également possible de détecter des variantes plus improbables au sein des résultats aux tests cognitifs. Entre autres, dans le contexte de non-validité, on observe parfois des scores dans la « normale » à des épreuves reconnues comme plus sensibles à la présence d'une atteinte cognitive. On peut alors utiliser le principe de SnOut (Sackett *et al.*, 2000; Smith *et al.*, 2008). [Mentionnons que le SnOut reflète le pendant du principe de SpIn mentionné plus tôt en discutant de la logique d'interprétation de la non-validité.] Ce principe propose que l'obtention d'un bon score à un test sensible (*Sn*) à la condition d'intérêt nous permet d'exclure (*rule Out*) la présence de cette condition (*SnOut*). Considérons le scénario suivant : un homme qui rapporte de graves difficultés mnésiques échoue un TVP qui s'apparente à un test de mémoire. En revanche, il produit, lors de la même évaluation, de bons résultats lors d'une épreuve mnésique nettement plus complexe (comme le CVLT ou les 15 mots de Rey). Ce type de résultat nous permet de conclure que : 1) il y a une incohérence marquée entre les données aux tests de mémoire et 2) il est très peu probable que cet homme présente un déficit mnésique plus sévère ou généralisé. En contexte d'expertise, la possibilité de détecter de telles incohérences constitue un argument fort pour ne pas discontinuer ou trop écourter l'évaluation, même si la personne échoue plusieurs TVP.

Ce lien entre la validité et la crédibilité demeure largement unidirectionnel. Bien que des résultats cognitifs non valides ne soient pas toujours associés à un tableau non crédible, à l'inverse, des résultats cognitifs au sein d'un tableau franchement non crédible demeurent

suspects. Ceci explique pourquoi, dans les cas où d'importantes incohérences sont présentes, certains proposent que la non-validité peut être établie sur la base d'un seul TVP, du moins sous certaines conditions (p. ex., Proto *et al*., 2014; Vickery *et al.*, 2004).

LA NON-CRÉDIBILITÉ ET LES INCOHÉRENCES « INCONTESTABLES »

Dans les cas où l'on peut observer de telles incohérences, il peut être possible pour le neuropsychologue-expert d'en arriver à la conclusion qu'il est probable que la personne amplifie volontairement ses difficultés cognitives. À cet égard, notons qu'il existe des descriptions de ce que constitue un « trouble neurocognitif simulé » (*malingered neurocognitive disorder)*, permettant à l'expert de prendre une décision en utilisant des critères plus formels et reconnus (Bianchini, Greve et Glynn, 2005; Slick *et al.*, 1999; Slick et Sherman, 2012). Entre autres, Bianchini et ses collaborateurs (2005) décrivent ce qu'ils appellent des incohérences « incontestables » (*compelling inconsistencies)*, soit des situations où la présentation clinique serait si incohérente ou improbable qu'il est raisonnable de conclure à la simulation ou à l'exagération volontaire. À cet effet, il importe de mentionner une situation rare, mais notable, qui permet de trancher de manière non équivoque sur la présence d'une exagération volontaire. Pour les TVP qui exigent un choix forcé, il arrive que la personne évaluée produise un score nettement en deçà du seuil qui serait obtenu au hasard (p. ex., << 50 % à un test avec deux choix de réponses). Dans ces cas, il est tout à fait sensé de conclure que la personne a *choisi* volontairement les mauvaises réponses.

Toutes les descriptions du *Trouble neurocognitif simulé* nécessitent d'établir la présence d'un bénéfice externe objectivable (*gain secondaire)* pour expliquer la motivation de l'individu, que ce soit un gain financier potentiel ou un autre avantage social (p. ex., éviter le retour au travail) (Bianchini *et al.*, 2005; Slick *et al.*, 1999; Slick et Sherman, 2012). Il est certainement vrai que la présence d'un tel *gain secondaire* constitue un facteur probant dans la plupart des études de groupe évaluant la non-validité (Heilbronner *et al.*, 2009). Toutefois, il demeure tout aussi vrai qu'il reste impossible d'objectiver les effets de ce gain secondaire dans une évaluation clinique (sauf dans le cas extrême d'un aveu).

Par ailleurs, certaines études suggèrent que des facteurs autres que le gain secondaire pourraient expliquer certains résultats liés à la non-validité ou à la non-crédibilité. Un exemple assez frappant serait l'étude réalisée par Williamson et ses collègues (2012) décrivant un groupe de patients présentant des crises non épileptiques d'origine psychogénique (*Psychogenic Non-Epileptic Seizures - PNES).* Ceci est une problématique présente assez souvent avec un tableau clinique qui est « médicalement non expliqué » (*medically unexplained)*, donc relativement peu crédible sur

le plan physique. Dans cette étude, la variable la plus prédictive d'un échec à un TVP était la présence d'antécédents d'abus à l'enfance (que ce soit sexuel, physique ou psychologique) et non la présence d'un incitatif financier. En fait, ni la présence d'un gain secondaire ou d'une psychopathologie plus sévère (selon les résultats du MMPI-2) n'étaient prédicteurs significatifs de l'échec du TVP.

En neuropsychologie, il semble y avoir un plus grand intérêt pour les Troubles à symptomatologie somatique comme le PNES (anciennement troubles *somatoformes)*, notamment dans la mesure où ceci représente une clientèle qui se retrouve assez souvent en expertise et chez qui on note un taux de non-validité plus élevé (Boone, 2017; Delis et Wetter, 2007; Lamberty, 2008). Selon le DSM-5, ces troubles psychologiques se définissent par la présence de symptômes somatiques particulièrement atypiques ou excessifs, incluant souvent des plaintes cognitives. Ces conditions nous rappellent qu'il existe des situations ou des conditions qui peuvent affecter la validité et la crédibilité des données, sans nécessairement signifier la présence d'une simulation franche. Sur le plan de la validité, certaines études nous suggèrent que les échecs aux TVP pourraient refléter une stratégie visant à éviter ou minimiser leurs symptômes. À titre d'exemple, Suhr et Spickard (2012*)* rapportent que chez un groupe de gens souffrant de céphalées récurrentes il y avait une corrélation entre le rendement à un TVP et les croyances que l'effort cognitif pourrait accentuer leurs douleurs (« cogniphobie »)*.* Pour ce qui concerne la crédibilité, Boone (2017) prétend que le trouble *somatoforme* représente une forme d'auto-illusion (*self-deception)* où la personne affectée demeure convaincue de la présence d'une condition sévère et de son invalidité. Selon Boone, la personne souffrant d'une telle condition peut présenter un tableau assez cohérent sur le plan fonctionnel, intégrant pleinement son rôle de malade (*sick role)*. En revanche, selon Boone, la personne qui simule aurait tendance à « choisir » certains éléments de non-fonctionnalité, manifestant une certaine incongruité selon le contexte. Toutefois, il n'existe pas, pour le moment, de critères reconnus pour établir un tel diagnostic différentiel, surtout en matière d'une amplification consciente ou inconsciente des symptômes (voir Merten et Merckelbach, 2013). Ceci n'exclut pas non plus la possibilité qu'une personne avec un trouble psychologique réel puisse choisir d'amplifier certaines difficultés (comme pour le Trouble factice).

Bref, la non-crédibilité du tableau clinique peut assez souvent mener à une forme d'impasse sur le plan diagnostique. Ceci ne devrait pas être surprenant. Si la non-validité des résultats cognitifs exclut l'interprétation plus détaillée du profil neuropsychologique, la non-crédibilité diminue notre capacité de formuler une opinion plus juste sur le plan du diagnostic neuropsychologique/psychologique. On pourrait se rassurer en se disant que ceci semble compatible avec la pratique de la plupart des

neuropsychologues. Selon des sondages, une majorité de neuropsychologues auraient tendance à éviter le terme *malingering (*simulation), misant plutôt sur les notions de non-validité et les incohérences au sein du profil clinique (Martin, Schroeder et Odland, 2015; Slick, Tan, Strauss et Hultsch, 2004).

QUELQUES QUESTIONS PERSISTANTES

Il serait impossible de décrire tous les enjeux, débats et controverses qui persistent autour d'une question aussi complexe et multifactorielle que la validité. Il existe de nombreux livres et chapitres plus récents qui permettent une appréciation plus approfondie du sujet (Boone, 2007; 2013; Carone et Bush, 2012; Larrabee, 2007; 2012; Millis, 2009). Ceci étant dit, il semble utile d'aborder quelques questions persistantes dans le contexte du présent article.

D'abord, il faut retenir que la plupart du temps la présence d'un profil non valide, même non crédible, n'est pas pour autant une *preuve* que la personne n'a pas de trouble neurocognitif ou psychologique. Dans le cadre d'une expertise pour un employeur, par exemple, il serait erroné de conclure qu'en raison de données non crédibles, la personne évaluée est apte pour effectuer un retour au travail. À moins de pouvoir démontrer de manière concluante la présence de capacités préservées, le mieux que l'on peut dire serait que les résultats obtenus reflètent une sous-estimation de ces capacités réelles. Rappelons également qu'un tableau non valide ou non crédible dans une sphère (p. ex., sur le plan cognitif) n'est pas non plus une preuve de la non-validité dans les autres sphères (p. ex., sur le plan émotionnel ou physique).

D'autre part, il n'est pas nécessaire de limiter notre interprétation d'un TVP à une simple dichotomie (réussite ou échec). Évidemment, un score à un TVP se situant tout juste en deçà d'un score seuil n'a pas la même valeur interprétative qu'un score si faible qu'il s'approche du niveau du hasard. De plus en plus d'études nous permettent d'ajuster les cotes de sensibilité/spécificité en fonction du score obtenu, pas seulement en fonction d'un score seuil (Lippa, Agbayani, Hawes, Jokic et Caroselli, 2014; Lippa *et al*., 2017; Walters, Berry, Rogers, Payne et Granacher, 2009).

Il y aura toujours des cas limites. Certaines données pourraient suggérer un niveau de cohérence plus problématique (p. ex., échec à un seul TVP, mais scores très variables), sans pouvoir statuer de manière plus définitive sur la non-validité des résultats cognitifs. De plus, il peut arriver que certains éléments du tableau nous mènent à mettre en doute la crédibilité d'une certaine portion des données, mais pas d'autres. Toutefois, si l'on comprend la cohérence comme un indice de la « variabilité » des

éléments au sein du profil clinique plus large, il peut être raisonnable de considérer certaines données « moins » valides que d'autres ou des présentations comme « moins » cohérentes ou crédibles. Dans ces cas, l'*erreur de mesure* associée à notre évaluation sera nécessairement plus élevée, augmentant du coup notre niveau d'incertitude envers nos conclusions. De tels profils nécessitent donc une interprétation plus prudente et nuancée.

Enfin, comme c'est le cas pour bien des aspects de la pratique en neuropsychologie clinique, la situation s'avère plus complexe dans le contexte québécois, dans la mesure où la vaste majorité des études répertoriées ont été réalisées dans un contexte linguistique, socioculturel et légal différent du nôtre. À quelques exceptions près (p. ex., Chouinard et Rouleau, 1997), il existe relativement peu d'études ou d'écrits portant sur les questions de la validité au Québec. Pourtant, la situation québécoise nous permettrait de poser des questions intéressantes. Par exemple, est-ce que la prévalence de non-validité diffère selon l'accessibilité aux soins de santé et à la réadaptation? Est-ce qu'un système d'assurance automobile universel *(*comme celui de la SAAQ) aurait une influence sur la présentation des cas d'expertise?

CONCLUSION

Il n'y a pas de tests de simulation. Ceci étant dit, l'évaluation de la non-validité demeure une composante complexe, mais absolument essentielle de l'expertise en neuropsychologie, nécessitant une approche à la fois structurée dans sa formulation et nuancée dans son interprétation. Idéalement, l'évaluation de la non-validité (la première étape) devrait se faire en utilisant plusieurs TVP spécifiques et imbriqués à différents moments de l'évaluation. Ces résultats contribuent alors au jugement de la non-crédibilité du tableau clinique (la deuxième étape). Devant un profil manifestement non crédible, il peut être possible – sous certaines conditions – de conclure à la probabilité d'une exagération volontaire des difficultés cognitives, notamment en présence d'incohérences « incontestables ». En d'autres circonstances, il serait préférable de s'en tenir à une description des éléments qui contribuent à la non-validité et à la non-crédibilité du tableau clinique.

Étant donné l'importance des enjeux, il faut tenir compte des biais trop souvent associés à cette démarche. D'un côté, il devrait être évident qu'il n'est plus défendable d'exclure les tests de validité dans un contexte d'expertise, surtout sous prétexte de pouvoir détecter un profil non valide sur la base du jugement clinique ou de vouloir donner le « bénéfice du doute » à la personne évaluée. C'est ce que certains auteurs nomment le biais *béta* (Erdodi et Lichtenstein, 2017*).* En fait, on pourrait avancer que

l'utilisation judicieuse des TVP peut servir à « protéger » les personnes faussement accusées d'exagérer ou de simuler leurs difficultés cognitives. En contrepartie, il est tout aussi préjudiciable de soutenir pouvoir identifier la *simulation* sur la base de ces mêmes tests, notamment avec la prétention de vouloir dire « les vraies affaires » (voilà le biais *alpha;* Erdodi et Lichtenstein, 2017). En fin de compte, il semble préférable de maintenir une approche plus équilibrée et conservatrice, limitant nos interprétations aux étapes des processus nécessaires pour répondre à l'essentiel du mandat d'expertise.

Enfin, rappelons que notre responsabilité à titre d'expert se limite à l'évaluation de la fiabilité et de la validité des *informations* qui nous sont fournies, dans le but de s'assurer de pouvoir fournir une opinion objective et neutre sur le fonctionnement de la personne. Dans la plupart des cas, ceci exclut un jugement plus global de la crédibilité de la personne évaluée ou de son témoignage plus large. Dans le contexte psycholégal, cette responsabilité appartient à la Cour.

RÉFÉRENCES

An, K. Y., Kaploun, K., Erdodi, L. A. et Abeare, C. A. (2017). Performance validity in undergraduate research participants: A comparison of failure rates across tests and cutoffs. *The Clinical Neuropsychologist, 31*(1), 193-206.

Armistead-Jehle, P. et Hansen, C. L. (2011). Comparison of the repeatable battery for the assessment of neuropsychological status effort index and stand-alone symptom validity tests in a military sample. *Archives of Clinical Neuropsychology, 26*(7), 592-601.

Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J. et Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury, 27*(7–8), 909-916.

Berry, D. T. R., Harp, J. P., Koehl, L. M. et Combs, H. L. (2017). Applying diagnostic standards to performance validity tests and the individual case. Dans S. C. Bowden (dir.), *Neuropsychological assessment in the age of evidence-based practice: Diagnostic and treatment evaluations* (p. 239-257). New York, NY : Oxford University Press.

Bianchini K. J., Greve, K. W. et Glynn, G. (2005). On the diagnosis of malingered pain-related disability: lessons from cognitive malingering research. *The Spine Journal, 5*(4), 404-417.

Bilder, R. M., Sugar, C. A. et Hellemann, G. S. (2014). Cumulative false positive rates given multiple performance validity tests: Commentary on Davis and Millis (2014) and Larrabee (2014). *The Clinical Neuropsychologist, 28*(8), 1212-1223.

Binder, L. M., Iverson, G. L. et Brooks, B. L. (2009). To Err is human: "abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology, 24*(1), 31-46.

Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective.* New York, NY : Guilford Press.

Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist, 23*(4), 729-741.

Boone, K. B. (2013). *Clinical practice of forensic neuropsychology: An evidence-based approach*. New York, NY : Guilford Press.

Boone, K. B. (2017). *Neuropsychological evaluation of somatoform and other functional somatic conditions.* New York, NY : Routledge.

Boone, K. B., Lu, P., Back, C., King, C., Lee, A., Philpott, L.,… Warner-Chacon, K. (2002). Sensitivity and specificity of the Rey Dot Counting Test in patients with suspect effort and various clinical samples. *Archives of Clinical Neuropsychology, 17*(7)*,* 625–642.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H.,… Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity (NAN Policy and Planning Committees). *Archives of Clinical Neuropsychology, 20*(4), 419–426.

Carone, D. A. et Bush, S. S. (2012). Introduction: historical perspectives on mild traumatic brain injury, symptom validity assessment, and malingering. Dans D. A. Carone et S. S. Bush (dir.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (p. 1-30). New York, NY : Springer.

Chafetz, M. (2011). Reducing the probability of false positives in malingering detection of social security disability claimants. *The Clinical Neuropsychologist, 25*(7), 1239-1252.

Chartrand, J.-P. et Kennepohl, S. (2016). Tout ce que vous n'avez jamais voulu savoir sur les probabilités… mais qui vous sera utile. *Psychologie Québec, sept. 2016.*

Chouinard, M.-J. et Rouleau, I. (1997). The 48-Pictures Test: A two-alternative forced-choice recognition test for the detection of malingering. *Journal of the International Neuropsychological Society, 3*(6)*,* 545–552.

Dandachi-Fitzgerald, B., Merckelbach, H. et Ponds, R. W. H. M. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology, 39(3),* 257-264.

Davis, J. J. et Millis, S. R. (2014). Examination of performance validity test failure in relation to number of tests administered. *The Clinical Neuropsychologist, 28*(2), 199–214.

Dean, A. C., Victor, T. L., Boone, K. B. et Arnold, G. (2008). The relationship of IQ to effort test performance. *The Clinical Neuropsychologist, 22*(4), 705–722.

Dean, A. C., Victor, T. L., Boone, K. B., Philpott, L. M. et Hess, R. A. (2009). Dementia and effort test performance. *The Clinical Neuropsychologist, 23*(1), 133–152.

Delis, D. C. et Wetter, S. R. (2007). Cogniform disorder and cogniform condition: proposed diagnoses for excessive cognitive symptoms. *Archives of Clinical Neuropsychology, 22*(5), 589-604.

Desjardins, P. (2009). L'expertise des psychologues en contexte légal. *Psychologie Québec, 23(5)*, 9-10.

Drane, D. L., Ojemann, J. G., Kim, M. S., Gross, R. E., Miller, J. W., Faught, R. E. Jr. et Loring, D. W. (2016). Interictal epileptiform discharge effects on neuropsychological assessment and epilepsy surgical planning. *Epilepsy & Behavior, 56,* 131-138.

Erdodi, L. A. et Lichtenstein, J. D. (2017). Invalid before impaired: an emerging paradox of embedded validity indicators. *The Clinical Neuropsychologist, 31(6-7),* 1029-1046.

Fox, D. D. (2011). Symptom validity test failure indicates invalidity of neuropsychological tests. *The Clinical Neuropsychologist, 25(3),* 488-495.

Fredericks, R. I. (2002). A review of Rey's strategies for detecting malingered neuropsychological impairment. *Journal of Forensic Neuropsychology, 2*(3-4)*,* 1-25.

Green, P., Rohling, M. L., Lees-Haley, P. R. et Allen, L. M. 3rd (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15(12),* 1045-1060.

Guillemette, T. J. (2012). The role of clinical judgement in symptom validity assessment. Dans D. A. Carone et S. S. Bush (dir.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (p. 31-44). New York, NY : Springer.

Harrison, A. G., Edwards, M. J., Armstrong, I. et Parker, K. C. H. (2010). An investigation of methods to detect feigned reading disabilities. *Archives of Clinical Neuropsychology, 25*(2)*,* 89-98.

Heaton, R. K., Smith, H. H. Jr., Lehman, R. A. et Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46*(5), 892-900.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R. et Conference Participants (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*(7), 1093-1129.

Iverson, G. L. (2005). Outcome from mild traumatic brain injury. *Current Opinion in Psychiatry, 18*(3), 301-317.

Kirkwood, M. W. (2015). *Validity testing in child and adolescent assessment: evaluating, exaggeration, feigning, effort and noncredible effort.* New York, NY : Guilford Press.

Lamberty, G. J. (2008). *Understanding somatization in the practice of clinical neuropsychology.* New York, NY : Oxford University Press.

Lange, R. T., Pancholi, S., Bhagwat, A., Anderson-Barnes, V. et French, L. M. (2012). Influence of poor effort on neuropsychological test performance in U.S. military personnel following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 34*(5), 453-466.

Larrabee, G. J. (2007) (Éd.). *Assessment of malingered neuropsychological deficits*. New York, NY : Oxford University Press.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22*(4), 666-679.

Larrabee, G. J. (2012). Assessment of malingering. Dans G. J. Larrabee (dir.), *Forensic neuropsychology: A scientific approach* (p. 116-159). New York, NY : Oxford University Press.

Larrabee, G. J. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology, 29*(4), 364-373.

Lippa, S. M. (2017). Performance validity testing in neuropsychology: a clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist, 32*(3), 391-421.

Lippa, S. M., Agbayani, K. A., Hawes, S., Jokic, E. et Caroselli, J. S. (2014). Effort in acute traumatic brain injury: Considering more than pass/fail. *Rehabilitation Psychology, 59*(3), 306-312.

Lippa, S. M., Lange, R. T., French, L. M. et Iverson, G. I. (2017). Performance validity, neurocognitive disorder, and post-concussion symptom reporting in Service members with a history of mild traumatic brain injury. *Archives of Clinical Neuropsychology*, *33*(5), 606-618.

Loring, D. W., Marino, S. E., Drane, D. L., Parfitt, D., Finney, G. R. et Meador, K. M. (2011). Lorazepam effects on word memory test performance: a randomized, double-blind, placebo-controlled, crossover trial. *The Clinical Neuropsychologist, 25*(5), 799-811.

Martin, P. K., Schroeder, R. W. et Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist, 29*(6), 741-776.

Meehl, P. E. et Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological bulletin*, *52*(3), 194-216.

Merten, T. et Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: a critical review. *Psychological Injury and Law, 6*(2)*,* 122-137.

Meyers, J. E. et Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology, 18*(3), 261-276.

Meyers, J. E., Volbrecht, M., Axelrod, B. N. et Reinsch-Boothby, L. (2011). Embedded symptom validity tests and overall neuropsychological test performance. *Archives of Clinical Neuropsychology, 26*(1), 8-15.

Meyers, J. E., Miller, R. M., Thompson, L. M., Scalese, A. M., Allred, B. C., Rupp, Z. W. et Junghyun Lee, A. (2014). Using likelihood ratios to detect invalid performance with performance validity measures. *Archives of Clinical Neuropsychology, 29*(3), 224-235.

Millis, S. R. (2009). What clinicians really need to know about symptom exaggeration, insufficient effort, and malingering: Statistical and measurement matters. Dans J. E. Morgan et J. J. Sweet (dir.), *Neuropsychology of malingering casebook* (p. 21–38). New York, NY : Psychology Press.

Mittenberg, W., Patton, C., Canyock, E. M. et Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*(8)*,* 1094-1102.

Nelson, N. W., Sweet, J. J., Berry, D. T. R., Bryant, F. B. et Granacher, R. P. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society, 13*(3), 440-449.

Proto, D. A., Pastorek, N. J., Miller, B. I., Romesser, J. M., Sim, A. H. et Linck, J. F. (2014). The dangers of failing one or more performance validity tests in individuals claiming mild traumatic brain injury related postconcussive symptoms. *Archives of Clinical Neuropsychology, 29*(7), 614-624.

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. *Archives de Psychologie, 28,* 215-285.

Roberson, C. J., Boone, K. B., Goldberg, H., Miora, D., Cottingham, M., Victor, T.,… Wright, M. (2013). Cross validation of the b Test in a large known groups sample. *The Clinical Neuropsychologist, 27*(3), 495-508.

Sackett, D. L., Straus, S. E., Richardson, W.S., Rosenberg W. et Haynes, R. B. (2000). *Evidence-based medicine. How to practice and teach EBM*. Édimbourg, Royaume-Uni : Churchill Livingstone.

Slick, D. J. et Sherman, E. M. S. (2012). Differential diagnosis of malingering. Dans D. A. Carone et S. S. Bush (dir.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (p. 57-72). New York, NY : Springer.

Slick, D. J., Sherman, E. M. S. et Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*(4), 545-561.

Slick, D. J., Tan, J. E., Strauss, E. H. et Hultsch, D. F. (2004). Detecting malingering: A survey of experts » practices. *Archives of Clinical Neuropsychology, 19*(4), 465-473. doi:10.1016/j.acn.2003.04.001

Smith, G. L., Ivnik, R. J. et Lucas, J. (2008). Assessment techniques: Tests, test batteries, norms, and methodological approaches. Dans J. E., Morgan et J. H. Ricker (dir.), *Textbook of clinical neuropsychology* (p.38-58). New York, NY : Taylor & Francis.

Suhr, J. A. et Gunstad, J. (2007). Coaching and Malingering: A review. Dans G. J. Larrabee (dir.), *Assessment of malingered neuropscyhological deficits* (p. 287–311). New York, NY : Oxford University Press.

Suhr, J. et Spickard, B. (2012) Pain-related fear is associated with cognitive task avoidance: exploration of the cogniphobia construct in a recurrent headache sample. *The Clinical Neuropsychologist, 26*(7), 1128-1141.

Swets, J. A., Dawes, R. M. et Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*(1), 1-26.

Van Dyke, S. A., Millis, S. R., Axelrod, B. N. et Hanks, R. A. (2013). Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist, 27*(8), 1234-1246.

Vickery, C. D., Berry, D. T., Dearth, C. S., Vagnini, V. L., Baser, R. E., Cragar, D. E. et Orey, S. A. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology, 19*(1), 37-48.

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J. et Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist, 23*(2), 297-313.

Walters, G. D., Berry, D. T. R., Rogers, R., Payne, J. W. et Granacher, R. P. Jr. (2009). Feigned neurocognitive deficit: Taxon or dimension? *Journal of Clinical and Experimental Neuropsychology, 31*(5), 584-593.

Williamson, D. J., Holsman, M., Chaytor, N., Miller, J. W. et Drane, D. L. (2012). Abuse, not financial incentive, predicts non-credible cognitive performance in patients with psychogenic non-epileptic seizures, the clinical neuropsychologist. *The Clinical Neuropsychologist, 26*(4), 588-598.

Youngjohn, J. R. (1995). Confirmed attorney coaching prior to neuropsychological evaluation. *Assessment, 2*(3), 279–283.

RÉSUMÉ

Au cours des dernières décennies, une multitude de tests et d'indices ont été développés pour tenter de détecter la non-validité des résultats neuropsychologiques. Cependant, la prolifération et la popularité de ces mesures ont contribué à la perception qu'il est dorénavant possible de les utiliser pour identifier la simulation des troubles cognitifs (malingering). Dans cet article, un cadre conceptuel est présenté pour tenter de favoriser une meilleure compréhension des notions sous-jacentes et permettre une meilleure interprétation de la non-validité en expertise. Selon la présente conceptualisation, il existe deux étapes distinctes : la première consiste à discerner la non-validité des données à l'aide de méthodes psychométriques, notamment les tests de validité de performance (TVP). La deuxième requiert une décision quant à la crédibilité du tableau clinique, incluant un jugement quant à la cohérence de l'ensemble des informations cliniques. C'est seulement dans le contexte de certaines incohérences indéniables qu'il serait possible d'avancer une opinion évoquant l'amplification des problèmes cognitifs.

MOTS CLÉS

expertise, validité des réponses, simulation.

ABSTRACT

Over the past few decades, a multitude of tests and indicators have been developed to attempt to detect invalid neuropsychological test findings. However, the proliferation and popularity of these measures have contributed to the impression that it is possible to use them to identify malingering. In the following, a conceptual framework is proposed to attempt to better understand the underlying issues and aid with the interpretation of response invalidity. It is proposed that this involves two distinct steps. The first involves detecting non-valid cognitive results using psychometric methods, and more specifically performance validity tests (PVTs). The second requires judgment of the non-credibility of the overall clinical picture, integrating and assessing inconsistencies among various sources. Only in the presence of compelling inconsistencies might it be possible to make an informed decision about the exaggeration of cognitive symptoms.

KEYWORDS

independent medical examination, response validity, malingering,

Annexe

Exemple de calcul d'une probabilité postérieure

Étape 1

De l'Équation 2 :

Rapport de cote Prétest	X	Rapport de vraisemblance positif du test	X	Rapport de cote posttest
Prévalence/(1 - Prévalence)	X	Sensibilité/(1 – Spécificité)		
0,2 /(1,0 – .0,2)	X	(0,75 /1,0 - 0,95)		= (0,2 /0,8) * (0,75 /0,05) = 0,25 * 15 = 3,75

Étape 2

Rapport de cote posttest	=	Probabilité post/(1 – probabilité post)
3,75	=	PP/(1 – PP)
3,75 (1 – PP)	=	PP
3,75 – 3,75 PP	=	PP
3,75	=	PP + 3,75 PP
3,75	=	4,75 PP
3,75 /4,75	=	PP
0,79	=	Probabilité postérieure