

Validité de performance aux tests neuropsychologiques: prise de position de l'Association québécoise des neuropsychologues

Josie-Anne Bertrand^{1*}, Jean-Pierre Chartrand^{2*}, Anne-Karine Gauthier^{3*}, Stephan Kennepohl^{4*}, Sébastien Monette^{5*}

¹ Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, Canada ; Institut Universitaire en santé mentale Douglas, Montréal, Canada

² Groupe Expertise 2120, Montréal, Canada

³ Clinique d'évaluation neuropsychologique de l'Outaouais, Gatineau, Canada

⁴ Expertise Neurosciences, Montréal, Canada

⁵ Département de psychologie, Université du Québec à Montréal, Montréal, Canada

Traditionnellement, les neuropsychologues présumaient que les résultats obtenus lors de leurs évaluations représentaient un tableau fiable du fonctionnement cognitif de leurs clients. Or, la littérature démontre qu'une proportion non négligeable des personnes évaluées produit des résultats non valides, et ce, que ce soit dans un contexte clinique¹ ou en expertise. De plus, des études suggèrent que le jugement clinique serait insuffisant pour dépister une performance non valide aux tests neuropsychologiques. Le présent article vise à définir les concepts sous-jacents concernant la validité de performance, décrire les prises de position adoptées par les associations professionnelles, faire un bref état des lieux pour les clientèles pédiatrique, adulte et gériatrique, et présenter quelques méthodes appropriées pour son évaluation. Nous recommandons que toute évaluation neuropsychologique comporte une appréciation formelle de la validité des données obtenues, minimalement avec des indices imbriqués, en utilisant des outils et méthodes reconnus.

Introduction

Mise en contexte

Généralement, les neuropsychologues sont bien au fait des facteurs secondaires qui peuvent influencer les résultats obtenus aux tests neuropsychologiques, tels que le sommeil et le stress (Arnett, 2013). L'influence de tels facteurs en neuropsychologie clinique a d'ailleurs été la thématique principale du 2^e congrès de l'Association québécoise des neuropsychologues (AQNP) en 2014. Plus particulièrement, certains de ces facteurs peuvent affecter la validité de la performance aux tests, concept autrefois abordé principalement sous l'angle de la simulation (*malingering*).

Au cours des dernières années, nous avons pu observer la prolifération des tests de validité de performance (TVP), du moins en Amérique du Nord.

La correspondance concernant cet article doit être adressée à /
Correspondence concerning this article should be addressed to:

Josie-Anne Bertrand, Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Canada.

Courriel/e-mail: info@aqnp.ca

*Tous les auteurs ont contribué équitablement à la conceptualisation et rédaction de l'article.

Certains auteurs font référence à un véritable changement de paradigme en neuropsychologie clinique (Lippa, 2018). En 2015, deux sondages ont montré la présence d'un consensus dans la pratique clinique aux États-Unis. Dans l'étude de Sweet et al. (2015), plus de 88 % des cliniciens, incluant ceux qui n'avaient pas de pratique psycholégale, donnaient leur appui à l'utilisation d'indices de validité de performance. Dans un deuxième sondage, près de 90 % des répondants étaient en accord avec un énoncé proposant que les données des tests de validité étaient plus précises que les impressions cliniques pour déterminer la crédibilité d'un client (Martin et al., 2015). Ce même sondage relevait également que l'utilisation des tests de validité était jugée « désirable » ou « obligatoire » par près de 94 % des répondants.

Cette pratique est d'ailleurs fortement encouragée par deux des grandes associations de neuropsychologie clinique américaines. En 2005, la National Academy of Neuropsychology (NAN) a pris position dans un article portant exclusivement sur l'évaluation de la validité (Bush et al., 2005). Cette position suggérait notamment aux cliniciens d'utiliser des tests qui permettent de se prononcer sur la validité des données d'évaluation lorsqu'il y a possibilité de présence de gains secondaires (p. ex., gain financier,

¹ Afin d'éviter d'alourdir le texte, le terme "contexte clinique" ou "pratique clinique" seront utilisés pour référer à l'évaluation clinique en neuropsychologie dans un contexte hors-expertise et le terme "expertise" pour référer à l'évaluation clinique en neuropsychologie dans un contexte d'expertise, bien que ces deux activités puissent évidemment être considéré comme des activités cliniques.

accommodations, être dispensé du service militaire, etc.) ou lorsque l'on suspecte que l'effort peut être incomplet ou que les informations rapportées sont erronées. En 2009, et récemment en 2021, c'est au tour de l'American Academy of Clinical Neuropsychology (AACN) de publier un consensus sur l'évaluation de la validité (initialement décrite comme l'évaluation de l'effort, le biais de réponse et la simulation ; Heilbronner et al., 2009 ; Sweet et al., 2021). Plusieurs recommandations y sont formulées, tant pour la pratique clinique que pour la recherche. Le point central du consensus de l'AACN est que toute évaluation neuropsychologique, qu'il s'agisse d'une expertise ou non, devrait évaluer le degré de validité des résultats aux tests neuropsychologiques et psychologiques (Sweet et al., 2021).

En revanche, l'utilisation des tests de validité semble nettement moins répandue au Québec et dans les milieux francophones. De fait, Richer et Forest (2023) ont réalisé une étude visant à caractériser les croyances et les pratiques actuelles de détection de la validité des neuropsychologues québécois. L'étude visait également à mesurer le niveau d'adhésion des neuropsychologues québécois aux recommandations de pratiques entourant l'administration des TVP et TVS et la communication des résultats de cette évaluation. Leur échantillon final était composé de 113 participants, soit un taux de participation de 11,4% des neuropsychologues détenant une attestation d'évaluation des troubles neuropsychologiques et étant inscrits au tableau de l'OPQ. Les résultats de leur étude montrent que comparativement aux neuropsychologues américains, les neuropsychologues québécois sont plus nombreux à entretenir des croyances erronées concernant l'évaluation de la validité. Un peu moins de la moitié de l'échantillon (soit 41,5%) était en accord avec la croyance que toute évaluation neuropsychologique devrait contenir une mesure de validité objective. Leurs analyses démontrent également que seulement 11,3% des neuropsychologues ayant répondu au sondage administrent toujours des TVP et/ou des TVS, près de 40% en administrent dans moins d'une évaluation sur dix et 15,5% n'administrent jamais de mesures objectives. Les autrices avancent également que les neuropsychologues québécois sont moins nombreux à utiliser des méthodes de détections (subjectives et objectives) et qu'ils priorisent surtout les méthodes subjectives (ex. observations comportementales indicatives d'un faible engagement à un sous-test individuel; contradictions entre le rapports médicaux, professionnels, les symptômes autorapportés et les comportements observés).

Dans le reste de la francophonie, les seuls documents publiés par une association professionnelle francophone abordant ce sujet sont, à notre

connaissance, les lignes directrices pour l'expertise de l'Association suisse des neuropsychologues (ASNP, 2011; Frei et al., 2022). L'AQNP a donc décidé de publier une prise de position sur le sujet et d'émettre des recommandations à ses membres sur les erreurs à éviter et l'utilisation des tests de validité de performance (TVP).

L'objectif du présent document est d'émettre quelques recommandations concernant les TVP et d'ainsi inciter les neuropsychologues à prendre conscience des conséquences pouvant découler de leur sous-utilisation.

Démarche de la prise de position

Des membres de l'AQNP s'intéressant à la validité de performance se sont d'abord regroupés en sous-comité en 2019, afin de réfléchir sur les outils de validation des performances et leur utilisation en pratique clinique au Québec. Deux membres de ce sous-comité ont présenté les enjeux liés aux TVP lors du congrès annuel de l'AQNP en 2019 (Élisabeth Perreau-Linck et Josie-Anne Bertrand), incluant la présentation d'outils pour aider à l'utilisation des TVP.

En 2021, un nouveau groupe de travail, composé des auteurs de l'article, a été mis sur pied afin de rédiger un document destiné aux cliniciens francophones. L'objectif principal était d'informer les neuropsychologues du Québec et de la francophonie de l'importance d'aborder la validité de performance lors des évaluations neuropsychologiques, et ce dans tous les types de pratique neuropsychologique. À la suite de la première rencontre, le 17 mai 2021, il a été décidé de rédiger un article plutôt qu'un guide, cela permettant une meilleure diffusion de l'information dans la communauté francophone. Les auteurs se sont entendus sur l'importance d'aborder l'application des TVP aux trois tranches d'âges distinctes (clientèle pédiatrique, adulte et gériatrique), ainsi que le contexte de la francophonie.

Au total, quatre rencontres virtuelles ont eu lieu entre les mois de mai 2021 et janvier 2022. Après la rédaction d'une première version de l'article, les auteurs l'ont soumis à trois neuropsychologues externes afin de récolter leurs commentaires et suggestions de modifications. La nouvelle version du document, intégrant ces commentaires, a par la suite été soumise au conseil d'administration (CA) de l'AQNP afin de recevoir à nouveau des commentaires et des suggestions de modifications. Le CA de l'AQNP a approuvé la version finale de la prise de position de façon unanime le 8 novembre 2022.

Bref historique

Comme dans bien d'autres aspects de la neuropsychologie clinique, André Rey aura été l'un des pionniers dans le domaine de l'évaluation de la validité (Rey, 1941 ; Fredericks, 2002). Dans les années 1940 et 1950, Rey conçoit différentes tâches à la demande d'assureurs qui doutent de la bonne foi de certaines personnes évaluées. Quelques-uns de ces tests demeurent populaires aujourd'hui, bien que souvent utilisés en versions modifiées, tels les 15 items de Rey et le dénombrement de points (Boone et al., 2002).

C'est dans les années 1980 et 1990 que le domaine a pris de l'ampleur, notamment avec les travaux de chercheurs comme Pankratz (p. ex. Pankratz, 1983 ; Pankratz & Erikson, 1990) et l'arrivée de mesures comme le Test of Memory Malingering (TOMM; Tombaugh, 1996) ou le Word Memory Test (WMT ; Green et al., 1996). Au Québec, un test de validité à choix forcés, le test des 48 images, a également été validé dans la même période (Chouinard & Rouleau, 1997). Tout comme les travaux d'André Rey, ces premières études portaient davantage sur le gain secondaire et l'amplification des troubles cognitifs dans un contexte d'expertise. Initialement décrits comme des tests de « simulation » (malingering), puis comme des « tests d'effort » ou de « biais de réponse », un certain consensus s'est peu à peu construit autour du concept de validité de performance. Depuis les premières études, les modèles théoriques et de validation se sont peaufinés, menant à une véritable explosion d'articles scientifiques (Lippa, 2018).

Concepts généraux

Sur le plan psychométrique, la validité représente le niveau auquel la preuve empirique et la théorie soutiennent les interprétations d'un test pour l'utilisation proposée (Standards for Educational and Psychological Testing, 2014). Sur le plan clinique, la validité réfère au degré auquel les tests utilisés mesurent réellement ce qu'ils prétendent mesurer (Kennepohl, 2018).

En principe, il s'agirait d'une estimation quantitative et mesurable. L'objectif d'un TVP est d'évaluer si les résultats obtenus sont une représentation valide du fonctionnement neurocognitif de la personne évaluée. Un résultat trop faible à un test de validité contribuerait à mettre en doute la validité des données recueillies lors de l'évaluation neuropsychologique et limiterait conséquemment leur interprétation. La complexité associée à l'interprétation des données non valides sera abordée plus loin.

Les neuropsychologues pourraient se demander si l'investissement de temps supplémentaire qu'impliquent la passation et la correction de TVP est

nécessaire. À titre de professionnels spécialistes du comportement humain, les neuropsychologues ne devraient-ils pas être en mesure de détecter un manque de collaboration ou d'effort basé sur leur expérience et leur jugement clinique? Cela ne semble pas être le cas, du moins si on se fie aux différentes études formelles qui se sont penchées sur la question (p. ex. Heaton et al., 1978 ; Dandachi-FitzGerald et al., 2017). Peu importe le contexte, ces études indiquent toutes que les neuropsychologues sont peu habiles à détecter un profil non valide sans l'utilisation de mesures objectives.

Lors d'évaluation en clinique, il importe de distinguer la validité de performance de la validité de symptômes (Larrabee, 2012). La validité de performance se définit comme le degré de correspondance entre la performance obtenue aux tests cognitifs et le fonctionnement réel de la personne. Les tests de validité de performance (TVP) ressemblent à des tests cognitifs standards, mais sont réussis aisément par la vaste majorité des gens, donc même par ceux ayant des atteintes cognitives importantes ou des atteintes altérant le fonctionnement (typiquement un traumatisme craniocérébral sévère ou une maladie dégénérative en début d'évolution). Par exemple, un TVP pourrait ressembler à un test de mémoire, mais serait facilement réussi par les gens ayant des capacités mnésiques affaiblies. En fait, de la même façon qu'un score plus faible à un test de mémoire peut être interprété comme un indicateur d'un déficit mnésique, un score plus faible à un TVP peut contribuer à la conclusion de non-validité de performance aux tests cognitifs standards (voir Kennepohl, 2018).

Par ailleurs, les tests de validité des symptômes (TVS) ont pour but d'évaluer la validité des difficultés rapportées par la personne évaluée. Dans la littérature portant sur les TVS, un symptôme fait référence à une description subjective que la personne évaluée rapporte elle-même. Les TVS sont typiquement issus des questionnaires ou des entrevues semi-structurées dont le but est d'évaluer si les symptômes autorapportés sont exagérés, inconstants ou atypiques (p. ex. au sein des échelles du MMPI-2-RF ou du MMPI-3 ; Burchett & Bagby, 2021 ; SRSI; Guerten et al., 2018; IOP-29 ; Viglione, D. J., Giromini, L., & Landis, P., 2017)).

Bien que corrélés, les résultats aux TVS peuvent se distinguer de ceux obtenus aux TVP (Van Dyk et al., 2013). Ainsi, une personne évaluée peut présenter un profil de symptômes tout à fait cohérent (TVS valides), mais échouer les TVP (TVP non valides). Inversement, quelqu'un pourrait bien réussir aux indicateurs de validité de performance (TVP valides), mais rapporter des symptômes manifestement exagérés ou atypiques (TVS non valides). Le présent article se concentrera sur les TVP, étant donné leur plus grande portée pour ce qui est de l'évaluation des troubles neurocognitifs

en neuropsychologie,. Bien que les TVS demeurent particulièrement pertinents en contexte d'évaluation, surtout dans le contexte où il s'avère nécessaire d'interpréter les raisons pouvant expliquer la non-validité (Heilbronner et al., 2009 ; Sweet et al., 2021), une description plus détaillée de leur utilisation nous semblait au-delà de l'objectif du présent article.

Il existe plusieurs TVP permettant d'évaluer la validité des résultats obtenus lors d'une évaluation neuropsychologique. D'abord, on retrouve les TVP dits « spécifiques » ou « indépendants » (“stand-alone” dans la littérature anglophone). Bien que typiquement plus sensibles que les TVP dits « imbriqués » (voir plus bas), les TVP indépendants doivent être administrés en plus des épreuves choisies par le clinicien au cours de son évaluation, ce qui peut entraîner des désavantages tels qu'allonger le temps de passation ou encore devoir éviter une interférence en administrant un TVP à un moment inopportun en rencontre (p. ex. passer un test de mémoire en même temps qu'un TVP ressemblant à un test de mémoire tel que le TOMM, le WMT ou le Memory Validity Profile).

D'autre part, on note l'apparition grandissante de TVP dits « imbriqués », c'est-à-dire des indices dans les tests neuropsychologiques traditionnels qui, au-delà d'un seuil déterminé, peuvent servir d'indicateur de non-validité (p. ex. des indices dans le sous-test de Séquences de chiffres de la WAIS-IV, la reconnaissance en choix forcés du California Verbal Learning Test ou les temps de réponse du Conners Continuous Performance Test 3rd Edition; voir par exemple, Webber et Soble, 2018; Schwartz et al., 2016 et Robinson et al, 2023). Bien que typiquement considérés moins sensibles que les TVP indépendants, ils ne rajoutent pas de temps de passation et sont généralement plus résistants à l'entraînement (“coaching”) par un tiers (qui consiste à recevoir des enseignements pour pouvoir mieux déjouer les mesures de validité). Pour une discussion plus détaillée des différents types de TVP et leurs avantages et inconvénients, se référer à des écrits plus exhaustifs (p. ex. Boone, 2021; Lippa, 2018; Schroeder & Martin, 2022).

Quelques considérations interprétatives

Outre les trois énoncés de position décrits précédemment (NAN, 2005; AACN, 2009, 2021), plusieurs articles et ouvrages abordent de façon détaillée les enjeux liés à l'administration et à l'interprétation des TVP (Boone, 2021; Kennepohl, 2018; Kirkwood, 2015; Lippa, 2018; Sherman et al., 2020 ; Schroeder & Martin, 2022 ; voir aussi Meulemans & Seron, 2004 et Rainville, 2020). Il est tout de même pertinent de survoler quelques concepts et enjeux que les cliniciens devraient connaître.

Statistiques cliniques

L'interprétation des tests de validité de performance implique des connaissances de base en statistiques cliniques, incluant les concepts de sensibilité et de spécificité, introduits par Yerushalmy (1947), puis appliqués en psychologie clinique et dans la pratique basée sur les données probantes en médecine (Glaros & Kline, 1988; Sackett et al., 1997).

- La sensibilité réfère à la capacité d'un test de détecter toutes les personnes qui ont la condition d'intérêt (vrais positifs). Un indice ou un test de validité ayant une sensibilité de 60% parviendra à identifier 60% des personnes qui émettent un résultat non valide. On dira donc qu'il ne détecte pas 40% des personnes qui ont une performance non valide (faux négatifs).
- La spécificité réfère à la capacité d'un test à rejeter toutes les personnes qui n'ont pas la condition d'intérêt (vrais négatifs). Ainsi, un test de validité qui a une spécificité de 90% identifiera correctement 90% des personnes qui ne présentent pas de signes d'une performance non valide. Inversement, il indiquera faussement un test non valide chez 10% des personnes évaluées (faux positifs).
- Le score seuil (“cut-off”) correspond à la valeur à partir de laquelle il y aura une décision quant à la validité de la performance. Le score seuil peut être établi pour optimiser la classification globale des patients. Dans le cas des TVP, on vise habituellement une spécificité élevée (p. ex. spécificité typiquement > 90%). Ces seuils sont habituellement dépendants de plusieurs facteurs, dont la problématique clinique. Ainsi, il se pourrait qu'un seuil adapté pour des personnes qui ont subi un traumatisme craniocérébral (TCC) soit inadéquat pour une population psychiatrique.

La recherche portant sur les mesures de validité de performance appuie fortement l'utilisation de scores seuils qui ont une valeur minimale de 90% de spécificité (Lippa, 2018) afin de minimiser le nombre de faux positifs. Rappelons aussi que les valeurs de sensibilité et de spécificité ne permettent pas de se prononcer sur une probabilité pour un individu en particulier (voir comme exemple la Figure 1).

Nombre de tests de validité de performance à utiliser

L'administration de plusieurs mesures de validité à différents moments de l'évaluation, et sollicitant différents domaines cognitifs, a été recommandée à plusieurs reprises au cours des dernières années (p. ex. Boone, 2009 ; Chafetz et al., 2015). Cette pratique, qui permet notamment de réduire les chances d'obtenir un faux négatif, s'est répandue chez les neuropsychologues américains, qui administreraient en moyenne quatre à six mesures de validité de

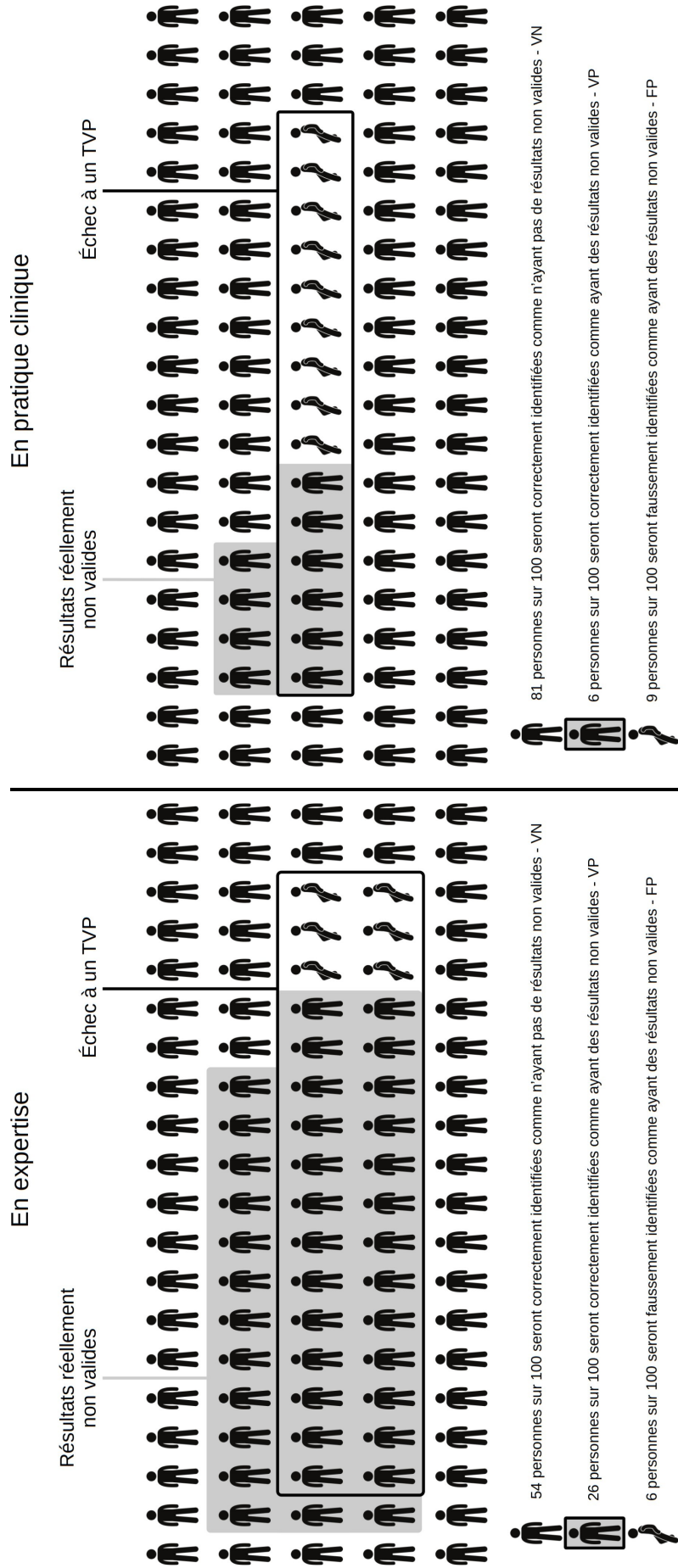


Figure 1. Que signifie un échec à un TVP? Ces deux séries d'icônes représentent les résultats obtenus (encadrés) à un TVP ayant une sensibilité de 0.65 et une spécificité de 0.90. Les données sont calculées à l'aide de prévalences hypothétiques de résultats non valides de 40% en expertise et de 10% en pratique clinique. En pratique clinique, un résultat positif, soit un échec à un TVP, annoncera plus souvent une conclusion erronée (faux positif) qu'une conclusion juste (vrai positif). La valeur prédictive positive (VPP) désigne la probabilité qu'une personne ait réellement des résultats neuropsychologiques non valides, sachant qu'elle a obtenu un échec à un TVP. Selon les données précédentes, en expertise, la VPP serait de 81%, alors qu'elle serait de 42% en pratique clinique. VN: vrai négatif; VP: vrai positif; FP: faux positif.

performance lors d'une évaluation neuropsychologique clinique régulière, du moins avec une clientèle adulte (Martin et al., 2015). Lorsque plusieurs tests de validité sont administrés et qu'il y a un échec à deux mesures ou plus, on atteint un très haut niveau de spécificité pour se prononcer sur la présence de résultats non valides (Chafetz, 2011; Larrabee, 2014a; Meyers & Volbrecht, 2003; Victor et al., 2009). Ce constat a mené à une règle fréquemment véhiculée et utilisée par les neuropsychologues, soit que deux échecs ou plus appuient la présence de données non valides, et ce, peu importe s'il s'agit de TVP imbriqués ou indépendants (Lippa, 2018 ; Sweet et al., 2021).

Des questionnements demeurent toutefois quant à la quantité optimale de mesures de validité à utiliser et à l'impact d'un trop grand nombre de ces mesures au sein d'une même évaluation. Deux études, qui ont eu recours à des simulations de données, ont soulevé la possibilité d'une relation positive entre la quantité de mesures de validité administrées et le taux de faux positifs (Berthelson et al., 2013; Odland et al., 2015).

Cette hypothèse aurait été réfutée par des études menées sur plusieurs ensembles de données cliniques. Celles-ci ont démontré qu'il n'y avait pas de relation notable entre la quantité de mesures administrées et les faux positifs, du moins jusqu'à un maximum de 10 mesures, puisque ces données cliniques comportaient tout au plus 10 mesures de validité (Davis & Millis, 2014; Larrabee, 2014b; Davis, 2018; Larrabee et al., 2019).

Même si Sherman et al. (2020) suggèrent de miser sur le ratio de tests échoués (p. ex. 4 échecs sur 14), plutôt que le nombre absolu, les données empiriques actuelles ne semblent pas suffisantes pour supporter une telle démarche (Sweet et al., 2021).

Concrètement, il faut retenir que l'administration de plusieurs mesures de validité au sein d'une même batterie neuropsychologique est recommandée, mais qu'au-delà d'une dizaine de mesures, il n'existe pas de données empiriques qui permettent de bien guider la réflexion quant au risque d'erreurs sur le plan de faux positifs. Au-delà du nombre de mesures, il est également recommandé d'administrer des mesures de validité qui donnent l'impression d'évaluer diverses fonctions cognitives (Boone, 2013).

L'interprétation de plusieurs résultats aux TVP

Plusieurs méthodes d'interprétation des TVP ont été proposées, comme les analyses multivariées, l'enchaînement des probabilités pour chaque test échoué et le décompte des échecs aux TVP (voir Davis, 2021, pour une revue détaillée des méthodes et

des enjeux). Parmi ces méthodes, le décompte du nombre d'échecs semble être non seulement la plus simple, mais aussi la plus fréquemment adoptée dans la littérature (p. ex. Rickards et al., 2018; Sherman et al., 2020). Essentiellement, cette méthode est une règle simple qui consiste à déterminer qu'un ensemble de données est non valide lorsqu'on atteint un nombre minimal d'échecs aux TVP, soit habituellement deux échecs ou plus obtenus à des TVP distincts. Il est à noter que certains TVP contiennent plusieurs indices ou scores. De multiples échecs au sein d'un même TVP doivent donc être comptabilisés comme un seul TVP échoué. Concrètement, si une personne échoue à trois indices d'un même TVP, le critère de deux échecs ou plus n'est pas satisfait, puisque ces indices ne représentent qu'un seul échec.

Il est essentiel de mentionner que la "détection" d'un profil non-valide (typiquement basée sur la présence de deux échecs ou plus aux TVP) ne fournit pas d'informations sur les intentions ou sur les causes précises de ces échecs (voir par ex., Sweet et al., 2021). En fait, ceci ne constitue que la première étape et nous permet d'établir s'il est possible ou non de procéder à une interprétation plus standard du profil de résultats aux tests neuropsychologiques. Dans la mesure où cela s'avère utile ou nécessaire, la détermination de la cause ou des raisons qui pourraient avoir mené à ces échecs fait partie d'une deuxième étape distincte au sein du processus décisionnel (Kennepohl, 2018 ; Rickards et al., 2018).

Les premiers critères visant à déterminer la présence de simulation aux tests cognitifs, mieux connus sous le nom des "critères de Slick", ont été publiés il y a plus d'une vingtaine d'années (Slick et al., 1999). Une mise à jour plus récente a été publiée en 2020, qu'on nomme les "critères de Sherman". Cette mise à jour vise à tenir compte des dernières données de la recherche et à adresser les critiques qui avaient été formulées à la première version des critères. Avant de conclure à de la simulation, les critères de Sherman suggèrent aux cliniciens de retrouver les points suivants: 1) la présence d'un gain secondaire, 2) la présence de résultats non valides (p. ex. scores non valides aux TVP), 3) la présence de discordances marquées (p. ex., entre les tests cognitifs, entre les symptômes rapportés au dossier et à l'entrevue, entre les observations et les plaintes, etc.), et 4) l'absence d'une condition développementale, médicale ou psychiatrique qui pourrait expliquer les comportements observés au point 2. Bien que ces critères permettent d'avoir de nouvelles balises et de guider le processus décisionnel, il serait mal avisé d'adopter les "critères de Sherman" sans regard critique. En effet, les auteurs de cette révision, tout comme Slick et al., 1999, soulignent que ces

nouveaux critères devront être testés et validés dans les milieux cliniques et à l'aide de futures recherches. Ils réitérent également qu'il s'agit d'une proposition de modèle.

Les TVP les plus utilisés dans la pratique clinique comprennent souvent une méthode à choix forcés (Martin et al., 2015). Habituellement, ces tests comportent des items où la personne évaluée doit choisir parmi deux choix de réponses. Lorsqu'un test de validité affiche un nombre d'erreurs qui dépasse statistiquement le seuil de la chance (p. ex. 23 erreurs ou plus sur 40 items), ce résultat devrait être considéré comme une intention de fournir les mauvaises réponses, même s'il s'agit du seul test échoué (Frederick & Speed, 2007 ; Binder al., 2014). Binder et al. (2014) fournissent d'ailleurs des tableaux permettant de déterminer si un résultat se retrouve sous le niveau de la chance, et ce en fonction du nombre d'items au test. Bien que rare en pratique, cela constitue la seule situation où un résultat à un seul TVP peut être considéré de façon quasi certaine comme une intention de la personne de fausser les données.

Il est important de réitérer que dans plusieurs cas, notamment dans une situation clinique, il n'est pas nécessaire de procéder à la détermination ou non de la simulation (comme les critères de Sherman). En ce qui concerne l'établissement d'un processus décisionnel plus adapté à la pratique clinique, Rickards et al. (2018) fournissent des recommandations quant aux scores seuils à utiliser, le choix du nombre de TVP à administrer, de même que des modèles d'interprétation clinique. Davis (2021) se montre toutefois critique face aux modèles de prise de décision comme celui de Rickards et al. (2018) parce que de tels modèles doivent être testés empiriquement afin de démontrer leur utilité clinique. Dans tous les cas, la transmission des résultats non-valides nécessite une approche à la fois nuancée et adaptée à la clientèle. À cet égard, il existe quelques articles suggérant des méthodes appropriées pour communiquer les résultats non-valides, notamment dans un contexte clinique (par ex., Carone et al, 2010, Postal & Armstrong, 2013).

Puisque les TVP et les scores seuils proposés pour différentes mesures ont été validés principalement auprès d'une population anglophone, une certaine prudence est de mise quant à leur utilisation auprès de la clientèle francophone, particulièrement en ce qui concerne les tests faisant appel aux capacités langagières. Certains auteurs suggèrent que les facteurs linguistiques et culturels pourraient avoir un impact sur les résultats aux TVP (voir Salazar et al., 2021 ; Strutt & Stinson, 2022). La plupart des études répertoriées ont été réalisées auprès d'une population

hispanophone, notamment aux États-Unis. En Europe, bien qu'il existe de plus en plus d'études qui étudient l'impact de la non-validité, l'utilisation des TVP demeure une pratique qui semble se limiter à quelques pays (Merten et al., 2021). Évidemment, ceci n'exclut pas l'utilisation judicieuse des TVP dans les milieux francophones, surtout si l'on favorise des seuils et une approche plus "conservatrice". Toutefois, il est également évident que davantage d'études de ce type devront être réalisées auprès de populations francophones.

Évaluation de la validité chez l'adulte

Tel que nous l'avons exposé plus haut, l'évaluation de la validité tire ses origines du contexte de l'expertise psycholégale et l'utilisation des TVP constitue dorénavant un élément incontournable dans ce contexte. Toutefois, au fil des années, il est devenu de plus en plus reconnu que la non-validité pouvait également être un enjeu dans le contexte de la pratique clinique. En fait, plusieurs études ont démontré que la non-validité pouvait être assez fréquente, notamment pour certaines problématiques plus complexes (comme l'épilepsie ou les crises non épileptiques psychogènes, p. ex. Tyson et al., 2018), ou pour lesquelles il existe des enjeux liés aux accommodements (comme chez les clients présentant un possible trouble du déficit de l'attention/hyperactivité). À ce sujet, les études répertoriées par Harrison et al. (2021) présentent des prévalences de résultats non valides allant de 11 à 48% chez les personnes qui consultent pour une évaluation du trouble du déficit de l'attention/hyperactivité. Dans le contexte de clients ayant subi un TCC léger, certaines études suggèrent que la proportion d'individus présentant un profil non valide peut être assez élevée (allant de 27 % à 43% selon les études), et ce même dans un contexte clinique (p. ex. Stulemeijer et al., 2007 ; Sherry et al., 2021).

La présence de résultats non valides peut avoir des répercussions très importantes sur le plan de l'interprétation. Entre autres, des résultats plus faibles aux TVP sont souvent corrélés à de faibles résultats à l'ensemble des tests neuropsychologiques administrés (p. ex. Green et al., 2001 ; Rohling et al., 2021). Par exemple, un score faible à un TVP qui semble solliciter la mémoire n'affecte pas seulement l'interprétation des tests de mémoire, mais toutes les épreuves neuropsychologiques. En d'autres mots, la présence d'une performance non valide plaide en faveur d'une interprétation conservatrice selon laquelle même les performances d'apparence valide peuvent sous-représenter les capacités réelles (Sweet et al., 2021).

Au-delà des difficultés d'interprétation, le fait d'ignorer la question de la validité dans un contexte clinique comporte des risques importants. Schroeder & Martin (2022) abordent quelques-unes des conséquences possibles. Celles-ci incluent l'impact d'un diagnostic erroné avec des enjeux psychoaffectifs importants (p. ex. l'annonce que la personne présente des troubles neurocognitifs sévères et persistants). D'autre part, ignorer les signes de non-validité peut servir à renforcer la symptomatologie non crédible et même contribuer à des effets iatrogènes (p. ex. Merckelbach & Merten, 2012 ; Carone, 2018).

La fausse attribution de la symptomatologie à une condition plus grave peut également mener à une multiplication d'exams inutiles (p. ex. IRM) ou à la prescription de traitements pouvant s'avérer dangereux, comme certaines médications (Schroeder & Martin, 2022). Enfin, un mauvais diagnostic pourrait mener à une série de restrictions qui limitent l'autonomie et la participation aux activités de la vie quotidienne, comme la cessation de la conduite automobile ou un arrêt de travail prolongé (Schroeder et Martin, 2022).

De plus, il est contre-indiqué d'interpréter les TVP de façon homogène et en ayant toujours recours aux mêmes scores seuils, indépendamment du type de problématique. Notamment, pour certaines conditions neurologiques, l'utilisation de scores seuils non adaptés à la population évaluée peut provoquer une mauvaise classification chez un patient fournissant un effort adéquat. Par exemple, lorsqu'ils sont utilisés avec leurs scores seuils « habituels », certains TVP peuvent avoir une validité discutable auprès de populations cliniques qui présentent un faible rendement intellectuel ou une atteinte cognitive plus importante, comme la déficience intellectuelle ou le trouble neurocognitif majeur (voir Lippa, 2018). Chez les patients qui sont aux prises avec un trouble psychotique, une méta-analyse récente démontre que lorsqu'il n'y a pas de litige légal, le taux d'échec aux TVP serait de 16% lorsqu'on utilise les scores seuils « habituels » (Ruiz et al., 2020). Selon les auteurs de l'étude, il est peu probable que la majorité de ces échecs soit expliquée par des enjeux motivationnels. Dans l'étude de Bodner et al. (2019), l'utilisation de certains TVP provoque des taux de faux positifs élevés chez des patients aphasiques, allant jusqu'à 73 %. Certains auteurs suggèrent également que les scores seuils habituels devraient être révisés pour les adultes qui souffrent d'épilepsie (Maiman et al., 2019). En résumé, lorsque les TVP sont utilisés auprès de populations cliniques qui présentent des troubles cognitifs sévères (dits « à risque » d'un taux anormalement élevé de faux positifs), il est suggéré d'administrer plusieurs TVP et de tenir compte de toutes les informations objectives présentes dans le

dossier afin de se prononcer sur la validité du profil (Ruiz et al., 2020; Bodner et al., 2019).

Un autre enjeu à considérer en pratique clinique est la prévalence des profils non valides variant selon le type de population ou le contexte clinique (p. ex. Dandachi-Fitzgerald et Martin, 2022). En situation d'expertise, la prévalence de résultats non valides peut varier entre 30 et 54% selon certaines études (p. ex. Ardolf et al., 2007 ; Larrabee, 2003; Larrabee et al., 2009 ; Mittenberg et al., 2002). Alors que Larrabee et al. (2009) ont suggéré une prévalence de $40\pm 10\%$, Young (2015) a critiqué ce nombre pour diverses raisons méthodologiques et a plutôt proposé une prévalence de $15\pm 15\%$. En pratique clinique, loin d'être considérée comme un phénomène rare, la prévalence de résultats non valides peut varier entre 4% et près de 22%, selon les problématiques (Dandachi-Fitzgerald et al., 2011 ; Hill et al., 2003 ; Kemp et al., 2008 ; Locke et al., 2008).

Puisque la prévalence a un impact non négligeable sur le taux de faux positifs, comme le présente la Figure 1, l'obtention d'un échec à un TVP ne doit pas être considérée de la même façon en contexte d'expertise qu'en pratique clinique (voir Elwood, 1993, pour une explication du théorème de Bayes en neuropsychologie). Prenons un exemple fictif à la Figure 1 où l'on estime les taux d'échecs (prévalence) à un seul TVP à 40% en expertise et à 10% en pratique clinique, puisque plusieurs études empiriques démontrent un plus haut taux de base en expertise (Dandachi-Fitzgerald et Martin, 2022). En contexte d'expertise, un faux positif surviendrait dans un cas sur six chez ceux qui échouent un TVP. En contexte de pratique clinique, toutefois, cette proportion de faux positifs passerait à plus de la moitié, soit 58%. En d'autres termes, dans un contexte où la prévalence de profils non valides est moindre (comme dans la plupart des milieux cliniques), un plus grand nombre d'échecs serait théoriquement nécessaire avant de se prononcer en faveur d'un ensemble de données non valides. Adapter le score seuil peut également aider à minimiser la probabilité d'un faux positif, bien que cela puisse diminuer la sensibilité de l'outil. Il incombe donc aux cliniciens de bien connaître les populations cliniques auxquelles ils sont confrontés, d'utiliser des scores seuils adaptés et d'interpréter de la façon la plus appropriée la signification d'un échec à un TVP en fonction du contexte de pratique et de la littérature scientifique appropriée.

En résumé, il demeure important de tenir compte de plusieurs facteurs dans l'interprétation des TVP chez l'adulte, incluant:

- la présence de facteurs externes pouvant affecter le niveau de participation (par ex., recherche d'accommodations, litige potentiel);

VALIDITÉ DE PERFORMANCE ET NEUROPSYCHOLOGIE

- La nature de la clientèle (par ex., problématiques plus complexes);
- Le contexte de l'évaluation neuropsychologique (par ex., expertise ou clinique, taux de base de l'échec des TVP).

Évaluation de la validité chez l'enfant et l'adolescent

L'évaluation de la validité de la performance est un sujet relativement nouveau dans le domaine de la neuropsychologie pédiatrique. Le sujet commence à percer dans la littérature pédiatrique depuis une quinzaine d'années (DeRight & Carone, 2015; Emhoff et al., 2018; Kirkwood, 2015). Ainsi, le choix de tests concernant la population pédiatrique et les recommandations de pratique sont à un stade moins avancé que ce que l'on retrouve auprès des adultes.

Il est d'ailleurs intéressant de constater que l'avis des neuropsychologues quant à l'importance d'utiliser des TVP diffèrent selon la population avec laquelle ils travaillent. De fait, l'utilisation de mesures de validité auprès de la population pédiatrique n'est pas aussi répandue qu'auprès de la population adulte. Par exemple, Hirst et al. (2017) ont sondé 654 neuropsychologues américains. Au total, 59% des neuropsychologues travaillant en pédiatrie croyaient que toutes les batteries de tests devraient inclure au moins une mesure de validité (contre 76% pour ceux travaillant auprès d'adultes), et 3,5 mesures de validité étaient utilisées en moyenne (contre huit avec les adultes).

Soulignons par ailleurs que cette étude a été réalisée aux États-Unis et que l'utilisation de mesures de validité qu'on y trouve n'est pas nécessairement équivalente à celle du Québec. Un sondage mené dans le cadre d'un travail dirigé au Québec (Gauthier, 2020) suggère que la majorité des neuropsychologues pédiatriques (61%) utilisent des TVP dans le cadre de leurs évaluations cliniques, une proportion quasi identique à celle de Hirst et al. (2017) pour les neuropsychologues pédiatriques (59%). Hirst et al. ont souligné qu'il était plus probable que les neuropsychologues travaillant auprès d'adultes mentionnent la présence de non-validité dans leur rapport, comparativement à ceux travaillant auprès d'enfants. Selon un sondage de Brooks et al. (2016) mené auprès de 282 cliniciens, la majorité des neuropsychologues pédiatriques se limiteraient principalement sur les observations comportementales et les incohérences des informations (p. ex. entre les données au dossier, les résultats aux tests et les informations autorapportées) afin de déterminer la validité de performance. Ils notent d'ailleurs que

l'utilisation de tests de validité de performance (TVP) arrive seulement au septième rang (sur dix) dans la liste des méthodes utilisées pour déterminer la crédibilité de la performance.

Pourtant, les enfants peuvent aussi avoir des performances non valides. Kirkwood (2015) a réalisé une revue de littérature et souligne que dans un contexte non litigieux, la prévalence d'une performance non valide en milieu pédiatrique serait de 3% à 5%. Il rapporte une proportion plus importante de performances non valides chez les enfants vus à la suite d'un TCC léger (12 à 20%) et les enfants dont les parents font une demande d'allocations/prestations sociales d'invalidité (US Social Security Disability benefits). Dans ce cas, 60% des enfants présentaient des signes de performances non valides.

Depuis la publication des lignes directrices de la NAN (2005) et de l'AACN (2009), la littérature portant sur la validité de performance chez l'enfant a suffisamment progressé (Kirkwood 2015, Emhoff, 2018) pour appuyer la recommandation d'utiliser des TVP de façon routinière lors d'évaluations neuropsychologiques auprès d'enfants d'âge scolaire et d'adolescents (Sweet et al., 2021). Notamment, les auteurs soulignent que les enfants sont certainement capables de présenter des comportements de simulation et que les professionnels les détectent difficilement (Peterson & Peterson, 2015). Les auteurs soulignent aussi que les motivations d'un enfant/adolescent à présenter une performance non valide diffèrent de ce qui est typiquement rencontré chez l'adulte.

L'enfant serait davantage motivé par le fait de pouvoir éviter des travaux académiques, d'avoir des accommodations scolaires, d'éviter certains stressseurs (p. ex. intimidation) ou de se soustraire de certaines activités (p. ex. sports). Certains enfants admettent également avoir fourni peu d'efforts lors de leur évaluation parce qu'ils n'ont pas aimé les tests, ceux-ci leur rappelant des tâches de nature scolaire (Flaro & Green, 2000).

Les auteurs soulignent également qu'il faut considérer la simulation par procuration ("malingering by proxy") comme explication. Ceci apparaît lorsque l'enfant est contraint par son parent ou figure parentale de fournir une performance non valide dans le but d'obtenir un gain externe (p. ex., des allocations/prestations sociales d'invalidité). Des enjeux comportementaux et affectifs (p. ex. opposition passive, ennui, anxiété de séparation, etc.) peuvent aussi influencer la validité de performance chez les enfants et les adolescents (Baker & Kirkwood, 2015).

Des recueils sommaires des TVP disponibles chez l'enfant existent (voir, par exemple, Boone 2021 et le

site internet de l'AQNP à cet effet : section Membres/ outils). Le clinicien peut s'y référer, mais tout comme pour la clientèle adulte, il doit être prudent dans les outils sélectionnés afin de s'assurer de leur applicabilité chez la clientèle évaluée. Notamment, les facteurs suivants devraient être pris en considération lors du choix d'un TVP chez les enfants :

- L'étendue d'âge couverte par le TVP (certains tests ou indices sont utilisables à partir de 8 ans ou 10 ans, plutôt que dès 6 ans);
- Prioriser les tests développés spécifiquement pour les enfants plutôt que les tests développés pour les adultes appliqués aux enfants;
- Prioriser des tests qui n'impliquent pas de capacité de lecture, de dénombrement ou de connaissance des lettres et des chiffres afin qu'ils puissent être utilisables auprès d'enfants ayant des troubles d'apprentissage.

L'évaluation de la validité chez la clientèle gériatrique

L'évaluation de la validité chez la clientèle gériatrique est particulièrement difficile en raison des troubles cognitifs relativement sévères et réels que l'on y retrouve, notamment dans le cas des troubles neurocognitifs majeurs (TNCM). Cela fait en sorte que de nombreuses études ont tenté d'identifier des TVP imbriqués, mais les valeurs psychométriques qui en résultent sont typiquement plus faibles (voir Bortnik & Dean, 2021). Notamment, la spécificité diminue avec la sévérité du TNCM et fluctue en fonction du type de TNCM (p. ex. Burton et al., 2015). Heureusement, le taux de base d'aînés présentant des performances non valides est estimé comme étant nettement plus faible comparativement à la population adulte non gériatrique (Bortnik & Dean,

2021), potentiellement en raison des répercussions significatives d'un diagnostic de TNCM sur les libertés de la personne. Tout de même, on peut retrouver des cas de non-validité, par exemple dans les troubles factices, une condition dans laquelle la motivation expliquant les plaintes cognitives/somatiques est de recevoir une attention médicale ou familiale. Par ailleurs, de nombreux facteurs autres que la simulation ou l'amplification de symptômes peuvent avoir un impact significatif sur la validité de performance chez cette clientèle, notamment la fatigue, la médication, un inconfort physique ou le manque de mobilisation du patient anosognosique qui n'est pas en mesure d'apprécier les raisons de l'évaluation (Bortnik & Dean, 2021). Ces facteurs peuvent venir biaiser les conclusions de l'évaluation et

affecter la qualité des services rendus. Il est donc tout aussi pertinent de se questionner quant à la validité de performance chez cette clientèle.

On retrouve certains TVP qui ont été grandement étudiés chez la population gériatrique et qui auraient de bonnes valeurs psychométriques (voir par exemple le Medical Symptom Validity Test (MSVT) dans Singhal et al., 2009). Une liste exhaustive de TVP a été dressée dans Boone (2021) et dans la revue de la littérature de McGuire et al. (2019). Plutôt que d'utiliser un score seuil, il semble d'autant plus important chez cette clientèle d'analyser les patrons de réponses. Par exemple, les patients avec un TNCM performant généralement moins bien aux tâches difficiles qu'aux tâches simples, alors qu'un patron inverse pourrait indiquer que la personne n'offre pas son plein potentiel.

Une performance sous le seuil de la chance en choix forcés pourrait également indiquer une non-validité. Dans certains TVP indépendants, un "profil de démence" a été développé afin d'optimiser l'évaluation de la validité, soit en permettant aux cliniciens de différencier les profils qui émergent d'un TNCM de ceux qui proviennent d'un résultat non valide. On le retrouve, entre autres dans le Word Memory Test (WMT), le MSVT et le Nonverbal Medical Symptom Validity Test (NV-MSVT). Après une revue exhaustive de la littérature chez cette clientèle, McGuire et al. (2019) concluent que ces trois tests sont les TVP les plus appropriés à utiliser dans les contextes de l'évaluation d'un TNCM. Ils concluent également que d'autres indices imbriqués, comme celui du Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), sont très prometteurs. Toutefois, il faut demeurer prudent quant aux conclusions tirées à la suite de l'administration de ces TVP, puisque lorsque les scores correspondent à un "profil de démence", cela n'indique pas sans équivoque la présence d'un TNCM et le profil pourrait représenter de la non-validité.

Une approche plus globale est essentielle et inclut notamment la cohérence des données. L'anosognosie, souvent présente à différents degrés chez la population âgée cognitivement atteinte, expliquerait une meilleure sensibilité des outils de validation de symptômes chez cette clientèle. De plus, s'assurer de mesurer la validité de symptômes, tel qu'avec le Memory Complaints Inventory (Green, 2019), pourrait s'avérer plus importante en gériatrie que de mesurer la validité de performance, les patients avec symptômes non crédibles rapportant nettement plus de symptômes cognitifs que la population âgée avec troubles cognitifs réels (dépressifs ou avec trouble neurocognitif léger ou avec TNCM) (Huber et al., 2022).

VALIDITÉ DE PERFORMANCE ET NEUROPSYCHOLOGIE

Les TVP s'appliquant à la population gériatrique peuvent également être des outils très intéressants pour soutenir la thèse d'un problème de validité auprès d'autres clientèles (Bortnik & Dean, 2021). Par exemple, les performances d'une personne ayant subi un TCC léger qui a un fonctionnement relativement bon au quotidien, mais qui performe moins bien que les patients avec un TNCM, peuvent être très évocatrices d'une non-validité de performance.

Prise de position de l'AQNP

Après avoir consulté les informations disponibles à ce jour sur l'évaluation de la validité, l'AQNP constate que l'évaluation de la validité est essentielle pour assurer la qualité de nos évaluations et la crédibilité de notre profession. L'AQNP en vient à formuler la prise de position suivante :

- “Toute évaluation neuropsychologique devrait comprendre une évaluation de la validité de performance (minimalement avec des indices imbriqués), à moins qu'il y ait présence d'une contre-indication, telle qu'une population peu étudiée ou un contexte particulier.”

Étant donné l'évolution dans le domaine de l'évaluation de la validité de performance et la crédibilité que cela procure à la profession, la prise de position de l'AQNP représente, à notre avis, des exigences minimales qui devraient être adoptées par l'ensemble des cliniciens. Elle ne devrait pas être perçue comme étant restrictive, mais plutôt comme reflétant les bonnes pratiques dans le domaine. Rappelons qu'il existe des raisons autres que la simulation pouvant mener à de faibles performances aux TVP, comme de l'opposition ou un trouble cognitif sévère (voir également Slick & Sherman, 2012). Il est aussi possible que certaines populations (p. ex. population gériatrique) et certains contextes (p. ex. patient évalué au chevet, en phase aiguë alors qu'il existe une instabilité dans son traitement ou sa condition médicale, en délirium, en psychose ou qui présente des troubles cognitifs sévères) limitent l'évaluation plus formelle de la validité. À notre avis, peu importe le contexte, tous les neuropsychologues devraient réfléchir à la validité de leur évaluation afin de minimiser l'impact de biais potentiels sur les conclusions qui peuvent être tirées. Lorsqu'une évaluation de la validité objective à l'aide de TVP (imbriqués ou indépendants) n'a pas été possible, les neuropsychologues devraient écrire dans leur rapport les raisons qui ont empêché l'utilisation de cette procédure.

Par ailleurs, nous appuyons les recommandations émises par l'AACN et nous recommandons notamment que :

- Les évaluations neuropsychologiques réalisées dans un contexte psycholégal comprennent plus d'un TVP indépendant.
- Les TVP et les scores seuils utilisés aient idéalement été étudiés et adaptés à la problématique évaluée.
- Les neuropsychologues soient prudents dans la description des tests de validité et des résultats obtenus (incluant le score seuil) cités dans les rapports afin
- d'assurer la protection des TVP/TVS (voir à cet effet la récente prise de position de l'AACN sur la protection des tests ; Boone et al., 2022).
- Les conclusions soient appuyées sur un ensemble d'informations et formulées de manière prudente et nuancée lorsqu'il y a présence de résultats non valides.

Conclusion

Le but de cet article était d'émettre des recommandations pour les neuropsychologues quant à l'utilisation des mesures de validité de performance. Alors que l'introduction des mesures de validité de performance remonte aux années 1940, le développement de celles-ci a réellement pris de l'ampleur dans les années 1990. Pourtant, encore à ce jour, de nombreux neuropsychologues francophones n'en font pas l'utilisation ou ne connaissent pas l'importance d'en faire usage. Typiquement, les neuropsychologues réalisant des expertises sont plus sensibles à leur nécessité. Toutefois, la littérature démontre la présence non équivoque de performances non valides dans une variété de populations cliniques et possiblement au sein des échantillons normatifs des tests neuropsychologiques (Erdodi & Lichtenstein, 2017). Récemment, l'AACN a d'ailleurs pris position quant à l'importance d'utiliser des mesures de validité dans toutes les évaluations, même dans un contexte clinique.

L'utilisation de tels outils ne se limite pas à la clientèle adulte, et devrait donc également être de mise en pédiatrie et en gériatrie. Bien que l'utilisation de ces outils prenne de l'ampleur depuis les dernières années, il importe d'introduire les notions liées à la validité dans la formation des cliniciens. De fait, dans le sondage mené par Gauthier (2020), les participants ont répondu en majorité (78%) qu'ils ne se sentaient pas suffisamment outillés quant à l'utilisation et l'interprétation de ces mesures et 74% indiquaient qu'ils n'avaient pas abordé les tests de validité de performance dans le cadre de leur formation universitaire. Davantage de formations sur la validité de performance devraient

donc être offertes afin de favoriser son utilisation et de contribuer à l'avancement de la profession.

Bien que la détection de la non-validité demeure une démarche complexe et délicate, les neuropsychologues demeurent très souvent les professionnels à qui incombe la responsabilité de soulever la possibilité que les résultats obtenus ou que les symptômes rapportés ne soient pas une représentation valide du fonctionnement réel de la personne concernée. À cet égard, il faut retenir que la neuropsychologie demeure une des spécialités les mieux placées pour exercer ce jugement, ne serait-ce que par la présence de tests objectifs (TVP) et la présence d'une littérature scientifique bien étoffée et solide dans le domaine

Conflits d'intérêts

Les auteurs déclarent ne pas avoir de conflits d'intérêts en rapport avec cet article.

Contributions des auteurs

Tous les auteurs ont lu et approuvé la version finale de ce manuscrit.

Remerciements

Les auteurs tiennent à remercier tous les membres du Conseil d'administration de l'AQNP qui ont relu ce texte de même que ces relecteurs désignés qui ont offert leurs commentaires à la suite de la première ébauche du texte: Geneviève Forest, Constant Rainville et Isabelle Rouleau .

Approbation éthique et consentement à participer

Non applicable.

Consentement pour publication

Non applicable.

Références

- Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist*, 21, 899–916.
- Arnett, P. A. (Ed.). (2013). *Secondary influences on neuropsychological test performance: Research findings and practical applications*. Oxford University Press.
- Baker, D. A., & Kirkwood, M. W. (2015). Motivations behind noncredible presentations: Why children feign and how to make this determination. In M. W. Kirkwood (Ed.), *Validity testing in child and adolescent assessment: Evaluating exaggeration, feigning, and noncredible effort* (pp. 125–144). The Guilford Press.
- Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury*, 27(7-8), 909-916. doi.org/10.3109/02699052.2013.793400
- Binder, L. M., Larrabee, G. J., & Millis, S. R. (2014). Intent to fail: Significance testing of forced choice test results. *The Clinical Neuropsychologist*, 28(8), 1366-1375. doi.org/10.1080/13854046.2014.978383
- Bodner, T., Merten, T., & Benke, T. (2019). Performance validity measures in clinical patients with aphasia. *Journal of clinical and experimental neuropsychology*, 41(5), 476-483. doi.org/10.1080/13803395.2019.1579783
- Boone, K. B. (Ed.). (2021). *Assessment of feigned cognitive impairment*. Guilford Publications.
- 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. doi:10.1080/13854040802427803
- Boone, K. B. (2013). *Clinical practice of forensic psychology: An evidence-based approach*. Guilford Publications.
- Boone, K. B., Lu, P., & Herzberg, D. (2002). *Rey Dot Counting Test*. Los Angeles, CA: Western Psychological Services.
- Boone, K. B., Sweet, J. J., Byrd, D. A., Denney, R. L., Hanks, R. A., Kaufmann, P. M., . . . Suhr, J. A. (2022). Official position of the American Academy of Clinical Neuropsychology on test security. *The Clinical Neuropsychologist*. doi:10.1080/13854046.2021.2022214
- Bortnik, K. E., & Dean, A. C. (2021). Performance Validity Testing in Patients with Dementia. Dans K. B. Boone (Éd.), *Assessment of feigned cognitive impairment* (2e éd., pp. 525-558).
- Brooks, B. L., Ploetz, D. M., & Kirkwood, M. W. (2016). A survey of neuropsychologists' use of validity tests with children and adolescents. *Child Neuropsychology*, 22(8), 1001-1020. doi:10.1080/09297049.2015.1075491
- Burchett, D., & Bagby, R. M. (2021). Assessing Negative Response Bias: a Review of the Noncredible Overreporting Scales of the MMPI-2-RF and MMPI-3. *Psychological Injury and Law*. doi:10.1007/s12207-021-09435-9.
- Burton, R.L., Enright, J., O'Connell, M.E., Lanting, S., & Morgan, D. (2015) RBANS embedded measures of suboptimal effort in dementia: Effort scale has a lower failure rate than the effort index. *Archives of Clinical Neuropsychology*, 30(1), 1-6.

- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology*, 20(4), 419-426. doi.org/10.1016/j.acn.2005.02.002
- Carone, D. (2018). Medical and Psychological Iatrogenesis in Neuropsychological Assessment Dans J. Morgan et J. P. Ricker (Éds.), *Textbook of Clinical Neuropsychological* (2e éd., pp. 1018-1031).
- Chafetz, M. (2011). Reducing the Probability of False Positives in Malingering Detection of Social Security Disability Claimants. *The Clinical Neuropsychologist*, 25(7), 1239-1252. doi:10.1080/13854046.2011.586785
- Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., Larrabee, G. J., & Ord, J. S. (2015). Official position of the American Academy of Clinical Neuropsychology Social Security Administration policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist*, 29(6), 723-740. doi.org/10.1080/13854046.2015.1099738
- Chouinard, M. J., & 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. et Martin, P. K. (2022). Clinical Judgment and Clinically Applied Statistics: Description, Benefits, and Potential Dangers When Relying on Either One Individually in Clinical Practice. , Dans R.W. Schroeder et P. K. Martin (Éds.), *Validity assessment in clinical neuropsychological practice* (pp.).
- Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology*, 39(3), 257-264. doi:10.1080/13803395.2016.1223278
- Dandachi-FitzGerald, B., Ponds, R. W. H. M., Peter, M. J. V., & Merckelbach, H. (2011). Cognitive underperformance and symptom over-reporting in a mixed psychiatric sample. *The Clinical Neuropsychologist*, 25, 812-828.
- Davis, J. J., & Millis, S. R. (2014). Reply to commentary by Bilder, Sugar, and Helleman (2014 this issue) on minimizing false positive error with multiple performance validity tests. *The Clinical Neuropsychologist*, 28(8), 1224-1229. doi.org/10.1080/13854046.2014.987167
- Davis, J. J. (2018). Performance validity in older adults: Observed versus predicted false positive rates in relation to number of tests administered. *Journal of clinical and experimental neuropsychology*, 40(10), 1013-1021. doi.org/10.1080/13803395.2018.1472221
- Davis, J. J. (2021). Interpretation of data from multiple performance validity tests. Dans K. B. Boone (Éd.), *Assessment of feigned cognitive impairment* (2e éd., pp. 525-558).
- DeRight, J., & Carone, D. A. (2015). Assessment of effort in children: A systematic review. *Child Neuropsychology*, 21(1), 1-24. doi: 10.1080/09297049.2013.864383
- Erdodi LA, Lichtenstein JD. Invalid before impaired: an emerging paradox of embedded validity indicators. *Clin Neuropsychol*. 2017 Aug-Oct;31(6-7):1029-1046. doi: 10.1080/13854046.2017.1323119. Epub 2017 May 27. PMID: 28552028.
- Elwood, R. W. (1993). Clinical discriminations and neuropsychological tests: An appeal to Bayes' theorem. *Clinical Neuropsychologist*, 7(2), 224-233. doi.org/10.1080/13854049308401527
- Emhoff, S. M., Lynch, J. K., & McCaffrey, R. J. (2018). Performance and Symptom Validity Testing in Pediatric Assessment: A Review of the Literature. *Developmental Neuropsychology*, 43(8), 671-707. doi: 10.1080/87565641.2018.1525612
- Flaro, L., & Green, P. (2000). Symptom validity test results with children: CARB and WMT. *Archives of Clinical Neuropsychology*, 15, 840.
- Frederick, R. I., & Speed, F. M. (2007). On the interpretation of below-chance responding in forced-choice tests. *Assessment*, 14(1), 3-11. doi.org/10.1177/1073191106292009
- Fredericks, R. (2002). A review of Rey's strategies for detecting malingered neuropsychological impairment. *Journal of Forensic Neuropsychology*, 2, 1-25.
- Gauthier, A. K. (2020). Les tests de validité de performance dans le cadre d'une évaluation neuropsychologique pédiatrique : leur utilisation, leur disponibilité et leur efficacité [Travail dirigé non publié]. Université de Montréal.
- Glaros, A. G., & Kline, R. B. (1988). Understanding the accuracy of tests with cutting scores: The sensitivity, specificity, and predictive value model. *Journal of clinical psychology*, 44(6), 1013-1023.
- Green, P. (2019). *Users' manual for the Memory Complaints Inventory (MCI)*. Green's Publishing.
- Green, P., Allen, L. M., & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer-administered forms, US version 1.1*. Durham, NC: CogniSyst.
- Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. A. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain injury*, 15(12), 1045-1060. doi.org/10.1080/02699050110088254
- Harrison, A. G., Lee, G. J., & Suhr, J. A. (2021). Use of performance validity tests and symptom validity tests

- in assessment of specific learning disorders and attention-deficit/hyperactivity disorder. Dans K. B. Boone (Éd.), *Assessment of feigned cognitive impairment* (2e éd., pp. 525-558).
- Heaton, R. K., Smith, H. H., Jr., Lehman, R. A., & 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., & 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. doi.org/10.1080/13854040903155063
- Hill, S. K., Ryan, L. M., Kennedy, C. H., & Malamut, B. L. (2003). The relationship between measures of declarative memory and the test of memory malingering in patients with and without temporal lobe dysfunction. *Journal of Forensic Neuropsychology*, 3(3), 1-18.
- Hirst, R. B., Han, C. S., Teague, A. M., Rosen, A. S., Gretler, J., & Quittner, Z. (2017). Adherence to Validity Testing Recommendations in Neuropsychological Assessment: A Survey of INS and NAN Members. *Archives of Clinical Neuropsychology*, 32(4), 456-471. doi:10.1093/arclin/acx009
- Huber, B. N., Jones, R. G., Capps, S. C., & Buchanan, E. M. (2022). Memory complaints inventory profiles: Differentiating neurocognitive impairment, depression, and non-credible performance. *Applied neuropsychology. Adult*, 29(2), 234-243. https://doi.org/10.1080/23279095.2020.1735388
- Kemp, S., Coughlan, A. K., Rowbottom, C., Wilkinson, K., Teggart, V., & Baker, G. (2008). The base rate of effort test failure in patients with medically unexplained symptoms. *Journal of Psychosomatic Research*, 65, 319-325.
- Kennepohl, S. (2018). Validité et crédibilité dans le contexte de l'expertise en neuropsychologie clinique. *Revue québécoise de psychologie*, 39(3), 51-74. doi.org/10.7202/1058184ar
- Kirkwood, M. W. (2015). *Validity testing in Children and adolescents assessment: Evaluating exaggeration, feigning, and noncredible effort*. New-York: Guilford Press.
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17, 410-425.
- Larrabee, G. J. (2012). Assessment of malingering. In G. J. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (pp. 117-159). New York, NY : Oxford University Press.
- Larrabee, G. J. (2014a). Minimizing false positive error with multiple performance validity tests: Response to Bilder, Sugar, and Hellemann (2014 this issue). *The Clinical Neuropsychologist*, 28(8), 1230-1242. doi.org/10.1080/13854046.2014.988754
- Larrabee, G. J. (2014b). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology*, 29(4), 364-373. doi.org/10.1093/arclin/acu019
- Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. *The Clinical Neuropsychologist*, 23(5), 841-849.
- Larrabee, G. J., Rohling, M. L., & Meyers, J. E. (2019). Use of multiple performance and symptom validity measures: Determining the optimal per test cutoff for determination of invalidity, analysis of skew, and inter-test correlations in valid and invalid performance groups. *The Clinical Neuropsychologist*, 33(8), 1354-1372. doi.org/10.1080/13854046.2019.1614227
- Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, 32(3), 391-421. doi.org/10.1080/13854046.2017.1406146
- Locke, D. E., Smigielski, J. S., Powell, M. R., & Stevens, S. R. (2008). Effort issues in post-acute outpatient acquired brain injury rehabilitation seekers. *NeuroRehabilitation*, 23, 273-281.
- Maiman, M., Del Bene, V. A., MacAllister, W. S., Sheldon, S., Farrell, E., Arce Rentería, Slugh, M., S. N. Siddhartha, & Barr, W. B. (2019). Reliable digit span: does it adequately measure suboptimal effort in an adult epilepsy population? *Archives of Clinical Neuropsychology*, 34(2), 259-267. doi.org/10.1093/arclin/acy027
- Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist*, 29(6), 741-776. doi.org/10.1080/13854046.2015.1087597
- McGuire, C., Crawford, S., Evans, J.J. (2019) Effort Testing in Dementia Assessment: A Systematic Review. *Archives of Clinical Neuropsychology*, 34(1), 114-131. doi.org/10.1093/arclin/acy012
- Merckelbach, H., & Merten, T. (2012). A note on cognitive dissonance and malingering. *The Clinical Neuropsychologist*, 26(7), 1217-1229. doi.org/10.1080/13854046.2012.710252
- Merten, T., Dandachi-FitzGerald, B., Hall, V., Bodner, T., Giromini, L., Lehrner, J., Gonzalez-Ordi, H., Santamaria, P., Schmand, B., & Di Stefano, G. (2021). Symptom and performance validity assessment in European countries: an update.

- Psychological Injury and Law. <https://doi.org/10.1007/s12207-021009436-8>
- Meulemans T, & Seron X (2004) L'examen neuropsychologique dans le cadre de l'expertise médico-légale. Mardaga, Sprimont.
- Meyers, J. E., & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology*, 18(3), 261–276. doi.org/10.1093/arclin/18.3.261
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094–1102.
- Odland, A. P., Lammy, A. B., Martin, P. K., Grote, C. L., & Mittenberg, W. (2015). Advanced administration and interpretation of multiple validity tests. *Psychological Injury and Law*, 8(1), 46–63. doi.org/10.1007/s12207-015-9216-4
- Pankratz, L. (1983). A new technique for the assessment and modification of feigned memory deficits. *Perceptual and Motor Skills*, 57, 367–372.
- Pankratz, L., & Erickson, R. C. (1990). Two views of malingering. *The Clinical Neuropsychologist*, 4(4), 379–389. <https://doi.org/10.1080/13854049008401832>
- Peterson, E., & Peterson, R. L. (2015). Understanding deception from a developmental perspective. In M. W. Kirkwood (Ed.), *Validity testing in child and adolescent assessment: Evaluating exaggeration, feigning, and noncredible effort* (pp. 42–61). The Guilford Press.
- Rainville, C. (2020). *La validation des observations en expertise médico-légale : l'expertise en psychologie-neuropsychologie*. Éditions Yvon Blais.
- Rickards, T. A., Cranston, C. C., Touradji, P., & Bechtold, K. T. (2018). Embedded performance validity testing in neuropsychological assessment: Potential clinical tools. *Applied Neuropsychology: Adult*, 25(3), 219–230.
- Rohling, M., Lanhinrichsen-Rohling, J., & Meyers, J. E. (2021). Effects of premorbid ability, neuropsychological impairment, and invalid test performance on the frequency of low scores. Dans K. B. Boone (Éd.), *Assessment of feigned cognitive impairment* (2e éd., pp. 264–282).
- Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. *Archives de Psychologie*, 28, 286–340.
- Ruiz, I., Raugh, I. M., Bartolomeo, L. A., & Strauss, G. P. (2020). A Meta-Analysis of Neuropsychological Effort Test Performance in Psychotic Disorders. *Neuropsychology review*, 30(3), 407–424. doi:10.1007/s11065-020-09448-2
- Sackett, D. L., Richardson, W., Rosenberg, W., & Haynes, R. (1997). *Evidence-based medicine: how to practice and teach EBM*. Edinburgh: Churchill Livingstone.
- Salazar, X. F., Lu, P. H., & Boone, K. B. B. (2021). The use of performance validity tests in ethnic-minority and non-English-dominant populations. Dans K. B. Boone (Éd.), *Assessment of feigned cognitive impairment* (2e éd., pp. 578–608).
- Schroeder, R. W., & Martin, P. K. (2022). Validity assessment in clinical settings: How it differs from forensic settings and why it is important. Dans R.W. Schroeder et P. K. Martin (Éds.), *Validity assessment in clinical neuropsychological practice* (pp. 3–11).
- Sherman, E. M., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology*, 35(6), 735–764.
- Sherry, N., Ernst, N., French, J. E., Eagle, S., Collins, M., & Kontos, A. (2021). Performance Validity Testing in Patients Presenting to a Specialty Clinic With a Mild Traumatic Brain Injury. *The Journal of Head Trauma Rehabilitation*. doi: 10.1097/HTR.0000000000000692
- Singhal, A., Green, P., Ashaye, K., Shankar, K., & Gill, D. (2009). High specificity of the Medical Symptom Validity Test in patients with very severe memory impairment. *Archives of Clinical Neuropsychology*, 24(8), 721–728. doi.org/10.1093/arclin/acp074
- Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13(4), 545–561. doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545
- Slick, D. J., & Sherman, E. M. (2012). Differential diagnosis of malingering and related clinical presentations. *Pediatric forensic neuropsychology*, 113–135.
- Standards for Educational and Psychological Testing (2014). American Educational Research Association: Washington DC.
- Strutt, A. M. & Stinson, J. M. (2022). Performance validity testing with culturally diverse individuals and non-native English speakers. Dans R.W. Schroeder et P. K. Martin (Éds.), *Validity assessment in clinical neuropsychological practice* (pp. 211–232).
- Stulemeijer, M., Andriessen, T.M.J.C., Brauer, J. M. P., Vos, P. E. & Van Der Werf, S. (2007) Cognitive performance after Mild Traumatic Brain Injury: The impact of poor effort on test results and its relation to distress, personality and litigation, *Brain Injury*, 21(3), 309–318,
- Sweet, J. J., Benson, L. M., Nelson, N. W., & Moberg, P. J. (2015). *The American Academy of Clinical*

- Neuropsychology, National Academy of Neuropsychology, and Society for Clinical Neuropsychology (APA Division 40) 2015 TCN Professional Practice and 'Salary Survey': Professional Practices, Beliefs, and Incomes of U.S. Neuropsychologists. *The Clinical Neuropsychologist*, 29 (8), 1069-1162. doi:10.1080/13854046.2016.1140228
- Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A., & Conference Participants. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*. doi.org/10.1080/13854046.2021.1896036
- Tombaugh T. N. (1996) *Test of Memory Malingering : TOMM*, New York : MHS
- Tyson, B. T., Baker, S., Greenacre, M., Kent, K. J., Lichtenstein, J. D., Sabelli, A., & Erdodi, L. A. (2018). Differentiating epilepsy from psychogenic nonepileptic seizures using neuropsychological test data. *Epilepsy & Behavior*, 87, 39-45. doi.org/10.1016/j.yebeh.2018.08.010
- Van Dyk, S. A., Millis, S. R., Axelrod, B. N., & Hanks, R. A. (2013). Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist*, 27(8), 1234–1246.
- Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist*, 23(2), 297-313. <https://doi.org/10.1080/13854040802232682>
- Yerushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Public Health Reports*, 62, 1432-1449. <https://doi.org/10.2307/4586294>
- Young, G. (2015) Malingering in Forensic Disability-Related Assessments: Prevalence 15±15 %. *Psychol. Inj. and Law*, 8,188–199.

Reçu le 23 Janvier 2023
Révision reçue le 12 octobre 2023
Accepté le 11 décembre 2023 ■