

Pratique professionnelle

Au-delà du consensus, l'incontournable exercice du jugement professionnel dans l'utilisation du WISC-IV



Pierre Desjardins / Psychologue

Directeur de la qualité et du développement de la pratique
pdesjardins@ordrepsy.qc.ca

L'utilisation du WISC-IV soulève, depuis le début, chez les psychologues du Québec de nombreuses questions. Il y a quelque temps déjà, l'Association québécoise des psychologues scolaires (AQPS) a initié des démarches pour rencontrer les responsables de la compagnie Pearson¹ afin d'obtenir des réponses notamment sur les normes de référence franco-québécoises. Quelques rencontres ont eu lieu, impliquant des psychologues de plusieurs milieux, praticiens et chercheurs, et des spécialistes de chez Pearson. Il semble qu'il n'y ait pas encore consensus, bien que plusieurs soient maintenant convaincus de la validité des normes franco-québécoises. L'AQPS a, pour sa part, pris position et recommande aux psychologues l'utilisation de ces normes, position qu'elle soutient dans un texte qui se trouve sur son site Web.

Il faut souligner que toute méthodologie de recherche comporte des forces et des faiblesses, des avantages et des inconvénients, y compris les recherches permettant la mise au point de tests psychologiques. Par conséquent, aucun test n'est parfait en toutes circonstances. Le psychologue qui choisit un test doit en connaître la portée et les limites et il doit engager son jugement professionnel quand vient le temps de tirer des conclusions à partir des résultats dont il dispose. Quelle que soit la norme de référence qu'il juge opportun d'utiliser, il est important qu'il l'indique dans son rapport afin que le résultat qu'il en dégage et la portée que d'autres pourraient lui donner ne portent pas préjudice au client.

Nous vous invitons, pour plus d'informations, à lire dans cette édition du magazine *Psychologie Québec* l'article de madame Élisabeth Roussy, psychologue, qui résume les échanges avec les spécialistes de chez Pearson et qui fait le point sur la question des normes franco-québécoises.

Notes

- 1 Pearson : anciennement Harcourt Assessment inc., Psychological Corporation ou PsyCorp, cette compagnie se spécialise notamment en recherche et développement d'outils d'évaluation, dont le WISC-IV.

COURS DE DÉONTOLOGIE ET PROFESSIONNALISME



POUR QUI ?

Les candidats à l'admission - Les psychologues

POURQUOI ?

Réfléchir sur plusieurs situations impliquant une prise de décision éthique susceptibles de se présenter dans le cadre d'une pratique professionnelle telles que :

La confidentialité - Les conflits d'intérêts -

La dangerosité - Les tribunaux

QUAND ?

Le cours totalise **45 heures de travail** et requiert la présence des participants à **deux journées** complètes de formation de **9 h à 16 h 30**.

À MONTRÉAL

- 23 mai et 20 juin (samedis)
- 11 septembre et 9 octobre
- 16 octobre et 13 novembre
- 6 novembre et 4 décembre

OÙ ?

Dans les bureaux de l'Ordre des psychologues du Québec situés au :1100, avenue Beaumont, bureau 510, à Ville Mont-Royal

COMBIEN ?

284,88 \$ (taxes incluses)

Les personnes intéressées à s'inscrire doivent le faire via le site Internet de l'Ordre :
www.ordrepsy.qc.ca/membres



Élisabeth Roussy

Neuropsychologue

Conseillère spécialisée en psychométrie au Centre de psychométrie et de mesures en réadaptation, Centre de réadaptation Marie-Enfant (CRME), CHU Sainte-Justine.



Geneviève Duchesne

Neuropsychologue

Programme de psychiatrie, neurodéveloppement et génétique du CHU Sainte-Justine.



Dr Georges Beauséjour

Psychologue

Chef professionnel des psychologues du CHU Sainte-Justine.

_WISC-IV : choisir les normes pour la population québécoise

Les guides de pratique en évaluation nous incitent à choisir des instruments psychométriques non seulement bien normalisés, valides et fidèles, mais également appropriés au contexte ethnique et culturel. Malheureusement, parmi la panoplie de tests utilisés au Québec, peu répondent à ce dernier critère. La publication en 2007 des normes québécoises du WISC-IV (Wechsler, 2007) a donc suscité beaucoup d'intérêt et, à cet égard, il faut souligner les efforts faits par Pearson pour adapter les tests psychométriques à notre contexte francophone. Les psychologues ont rapidement constaté que ces normes modifient l'interprétation des résultats, notamment à la baisse, au plan du raisonnement. Puisqu'il est plus difficile de concéder une baisse de points qu'une hausse, des preuves convaincantes sont nécessaires pour se résoudre à utiliser ces normes qui bousculent les référents cliniques. Or le manuel du test donne peu d'informations sur l'échantillonnage et les indices de validité et de fidélité, ce qui contribue au questionnement que soulève le petit échantillon utilisé, soit 174 sujets.

Le présent article vise à donner un éclairage sur les normes à privilégier en partageant des informations données par l'équipe de recherche de Pearson lors d'une conférence¹ organisée par le Centre de psychométrie et de mesure en réadaptation du Centre de réadaptation Marie-Enfant (CRME) affilié au CHU Sainte-Justine. L'article résume également les critiques de quatre consultants externes² ayant des connaissances avancées en statistiques et qui ont examiné la méthode de normalisation utilisée.

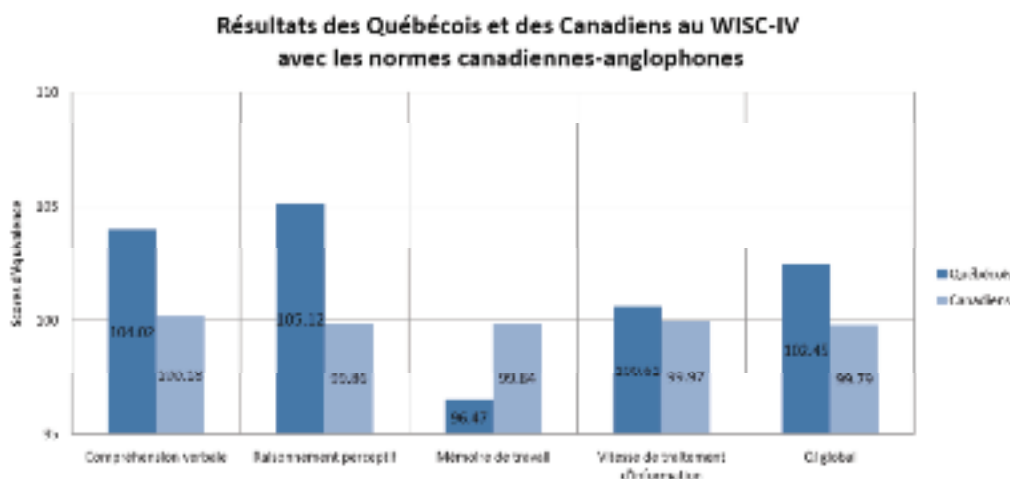
_LES NORMES CANADIENNES NE REFLÈTENT PAS LA POPULATION QUÉBÉCOISE

Selon Gendron & Zhu (2009), les normes canadiennes du WISC-IV ne seraient pas appropriées pour évaluer la clientèle québécoise, puisqu'il existe des différences significatives entre les Canadiens et les Québécois. Au test WISC-IV, les Québécois, comparés aux Canadiens, réussissent moins bien à l'Index Mémoire de travail et mieux aux Index Compréhension verbale, Raisonnement perceptif et au QI global. Les résultats sont comparables à l'Index Vitesse de traitement (voir graphique 1).

Des différences de moyenne au WISC-IV entre des populations ont souvent été observées (Gendron & Zhu, 2009) et des dissemblances culturelle, langagière et démographique sont invoquées pour les expliquer. Au Québec, la grande homogénéité de la culture et du système d'éducation est notamment avancée comme explication. Les auteurs ajoutent que, selon les rapports du Conseil des ministres de l'Éducation du Canada (2001), les Québécois auraient aussi des résultats supérieurs à ceux des Ontariens et des Canadiens anglais en lecture, mathématiques et sciences.

Ce qu'il importe de souligner, c'est qu'en l'absence d'ajustement des moyennes, les différences entre les composantes sont faussées, accentuées, puisque certaines d'entre elles sont plus

Graphique 1



élevées, alors que d'autres le sont moins. Concrètement, les normes canadiennes-anglaises laissent croire qu'un enfant québécois francophone dans la moyenne a des habiletés de raisonnement meilleures que son potentiel, associées à un léger décalage sur le plan de la mémoire de travail.

LA NORMALISATION INFÉRENTIELLE

Puisque la fidélité et la validité du WISC-IV ont déjà été établies à partir de la population américaine, canadienne et franco-ontarienne, une normalisation traditionnelle ne serait pas requise, selon Gendron & Zhu (2009) et « Questions fréquentes... » (en ligne, 2009). De plus, les analyses effectuées sur les données de la population québécoise montrent que les courbes de performance suivent de façon parallèle celles obtenues auprès des autres populations. Ceci confirmerait que le WISC-IV est valide chez les Québécois, puisqu'il mesurerait les mêmes concepts que chez leurs voisins.

LA NORMALISATION INFÉRENTIELLE

Une nouvelle méthode sous instance de brevet, la normalisation inférentielle (Wilkins, Rolfhus, 2004), a été utilisée afin d'effectuer les ajustements de moyennes. Cette méthode permettrait d'utiliser des normes reconnues pour leur valeur statistique et générées auprès d'un grand échantillon afin d'effectuer un ajustement des normes auprès d'un petit échantillon.

La méthode de normalisation inférentielle a l'avantage de requérir jusqu'à quatre fois moins de sujets que la normalisation traditionnelle (Wilkins, Rolfhus, 2004). Ceci s'explique par le fait que la distribution de la population pour un certain âge est estimée à partir des données de tous les groupes d'âge, notamment des âges voisins, et non pas seulement à partir des données collectées pour ce groupe d'âge. Deux critiques émanent de ces éléments d'information. D'abord, Achim (communications personnelles, février et mars 2009) souligne que les sujets du bout (c'est-à-dire les individus âgés de 6 et 16 ans) n'ont pas de voisins d'un côté. Par conséquent, il est possible que l'erreur de mesure soit plus grande pour ces âges. Ensuite, Achim et Cousineau (communication personnelle, 26 février 2009) soulignent que l'étude de validité de cette méthode comportait 50 sujets par groupe (Wilkins, Rolfhus, 2004). Or aucune étude n'aurait démontré que la méthode est efficace avec 16 sujets par groupes, comme c'est le cas pour les normes québécoises.

La normalisation inférentielle permet de générer des moments statistiques (p. ex. moyenne, écarts types, erreurs types, etc.) pour chacun des groupes d'âge. Une fois ces statistiques obtenues, des courbes idéales sont tracées à travers ces moments. Les courbes correspondent à des équations de régressions polynomiales allant du degré 1 à 5. La principale critique des consultants concerne l'absence de motivation sur le degré idéal. Choisir un polynôme de degré trop élevé entraînerait le risque que la courbe idéale épouse les spécificités de l'échantillon, rendant les normes moins généralisables, selon Cousineau. Les consultants ajoutent

que sans contrainte théorique une courbe polynomiale peut prendre une infinité de formes arbitraires. Selon les informations données par Pearson Assessment (Wechsler, 2007; Gendron & Zhu, 2009), il y aurait néanmoins contrainte théorique, puisque ce sont l'expérience clinique de l'instrument et l'expertise acquises avec des normalisations antérieures (p. ex. les courbes de croissance observées dans l'échantillon canadien, franco-ontarien et américain) qui ont permis d'ajuster les données.

Une autre réserve émise par Cousineau est que la forme de la distribution à ses extrémités n'ayant pas été prise en compte, les individus ayant un QI très haut ou très bas risquent d'être plus marginalisés. Notons cependant que les données d'enfants atteints d'un retard mental ou dotés d'une douance intellectuelle déjà identifiés, les 2,2 % à l'extrémité inférieure et supérieure de la courbe normale, ont été ajoutées à l'échantillon normatif (Wechsler, 2007; Gendron & Zhu, 2009). Ceci permet, en partie du moins, de combler cette lacune.

L'une des dernières étapes de la normalisation inférentielle implique des procédures de lissage qui permettent d'éliminer les données marginales et la variance échantillonnale de façon à ce qu'il y ait progression des scores à l'intérieur de chaque groupe et à travers tous les groupes d'âge. Selon Cousineau et Achim, quand un échantillon contient peu de sujets, il est difficile de déterminer ce qu'est une donnée marginale. De plus, Achim observe dans quelques données brutes québécoises que la courbe de performance pour certains âges, dans certains sous-tests, descend plutôt que d'augmenter progressivement en fonction du développement. La comparaison des deux tables de normes montre aussi un décalage inconstant entre les normes canadiennes et québécoises à travers les âges, ce qui soulève des questionnements. Bien que cette particularité puisse être attribuée à un phénomène unique au Québec, Achim estime qu'une erreur d'échantillonnage pourrait en être la cause.

LA COMPOSITION DE L'ÉCHANTILLON ET LES VARIABLES DE STRATIFICATION

Outre le nombre, la composition de l'échantillon normatif mérite d'être explicitée suivant trois variables importantes : le statut économique, la provenance géographique et l'origine ethnique.

Premièrement, le niveau d'éducation des parents aurait été utilisé à la place du statut socioéconomique, le rationnel étant que ces deux variables sont fortement corrélées (Gendron & Zhu, 2009). D'après Statistique Canada, les personnes ayant moins d'éducation auraient des revenus significativement moins élevés que les personnes ayant plus d'éducation.

Deuxièmement, en raison de la petitesse de l'échantillon, la provenance géographique n'a pas été stratifiée. Selon les recherches antérieures de la compagnie Pearson, les différences régionales disparaissent lorsqu'une stratification basée sur le niveau de scolarité est effectuée. La performance intellectuelle serait peu influencée par la variable géographique au Québec, alors que le niveau de scolarité des parents aurait un impact significatif.

Troisièmement, les chercheurs expliquent que la stratification par groupe ethnique a été omise, le Québec comprenant moins de 10 % de minorités visibles dans sa population (Gendron & Zhu, 2009). Cette absence de stratification peut être une lacune, d'autant plus que les psychologues évaluent fréquemment des francophones dont l'origine ethnique n'est pas caucasienne.

_LE CHOIX DES NORMES

Pearson a utilisé une méthode statistique sous instance de brevet qui n'a pas été décrite dans le manuel. Les psychologues québécois n'ont pas endossé ces normes sans se questionner, ce qui démontre qu'ils sont critiques, vigilants et que les principes scientifiques dans leur pratique de la psychométrie leur tiennent à cœur.

Mais le choix des normes doit être fait, et à priori, afin de minimiser les risques que ce choix, s'il est fait a posteriori, ne soit influencé par le désir de favoriser le client. Un consensus entre les psychologues est également important afin d'utiliser un langage commun quand vient le temps de partager des informations. Un dilemme se pose, puisque le choix entre les normes québécoises ou canadiennes s'avérerait imparfait dans un cas comme dans l'autre.

D'une part, l'utilisation des normes québécoises peut se révéler un choix délicat. D'abord, à la lumière des données qui leur ont été fournies, trois des quatre consultants ont évoqué de sérieuses réserves quant à la méthode statistique utilisée. Des critiques sont aussi émises quant au nombre de sujets, et ce, même en utilisant la normalisation inférentielle. De plus, des inquiétudes sont soulevées quant à de possibles erreurs d'échantillonnage. Dans cette éventualité, des mises en garde spécifiques pour certains sous-tests ou sous-groupes d'âge pourraient ultérieurement être publiées.

D'autre part, l'utilisation des normes canadiennes-anglaises n'est pas appropriée au Québec, car elle entraîne une erreur de mesure systématique lors des évaluations. Il s'agit d'une donnée significative dont il faut tenir compte pour éviter tout risque de surestimation du rendement intellectuel, qui se répercuterait par exemple sur le classement scolaire et sur l'allocation de services dont les enfants pourraient bénéficier.

Il s'agit de l'argument principal qui a convaincu l'Association québécoise des psychologues scolaires (AQPS) et le Regroupement des neuropsychologues pédiatriques du Québec (RNPQ), (qui implique des cliniciens de divers milieux) de recommander à leurs membres l'utilisation des normes québécoises. Cette recommandation a aussi été faite au CHU Sainte-Justine et au CRME, bien que chaque psychologue reste libre d'utiliser les normes qu'il juge les plus adéquates pour sa clientèle. Cette suggestion se base de plus sur les informations complémentaires recueillies lors de la conférence (Gendron & Zhu, 2009). Le choix a également été fait de croire en la bonne foi des chercheurs qui ont rectifié le tir en acceptant de présenter publiquement leurs données de recherches. Des analyses comparatives de protocoles québécois

corrigés avec les normes québécoises et canadiennes ont aussi apporté un éclairage clinique. En effet, les membres du RNPQ qui se sont prêtés à cet exercice concluent que les normes québécoises reflètent souvent mieux leurs observations cliniques, notamment en ce qui concerne le vocabulaire.

Enfin, il faut souligner que les faiblesses statistiques des normes québécoises ont été mises en évidence parce que les critiques les ont scrutées à la loupe. Un pareil exercice de révision, s'il était fait pour les autres tests que nous utilisons, nous permettrait de prendre conscience de leurs limites et ainsi, d'être plus prudents dans les conclusions qu'on en tire.

En utilisant les normes québécoises, il faudra faire preuve de discernement, comme nous devons le faire avec tous les outils qui comportent peu de sujets. Il est recommandé dans les rapports d'évaluation de spécifier les normes utilisées tout en présentant les résultats avec leur intervalle de confiance et, peut-être même, de préciser que les conclusions pourraient être différentes si on utilisait d'autres normes. L'utilisation au besoin d'outils complémentaires est également recommandée devant un profil particulier pour valider les résultats obtenus.

_Références

- Gendron, M.-J. & Zhu, J. (octobre 2008). *Normes québécoises du WISC-IV; présentation des travaux de recherche et démarche de consensus concernant le choix des normes*. Conférence donnée au CRME à Montréal.
- Pearson Assessment. « Questions fréquentes en ce qui concerne l'étude de validation québécoise du WISC-IV ^{CDNF} ». En ligne, 2009, <http://pearsonassess.ca/hai/images/support/wisc-iv-questions-frequentes-normes.pdf> (25mars 2009).
- Wechsler, D. (2007). *WISC-IV ^{CDNF} Manuel de normes québécoises*. Toronto (Ontario) : Harcourt Assessment.
- Wilkins, C., Rolfhus, E. (2004). *A simulation study of efficacy of inferential norming compared to traditional norming*, San Antonio (Texas) : Harcourt Assessment.

_Notes

- 1 Suite à cette conférence de trois heures, un document a été publié et est disponible sur le site de Pearson. Il est possible de visionner le film de la conférence en en faisant la demande à la compagnie Pearson.
- 2 Remerciements au Dr Louis Picard, psychologue, CHU Sainte-Justine, Denis Cousineau, Ph. D., professeur agrégé au Département de psychologie de l'Université de Montréal, ainsi que messieurs Jean Bégin, agent de recherche et statisticien, et André Achim, Ph. D. professeur agrégé au Département de psychologie de l'UQAM.