



Q-interactive™

Équivalence des tâches cognitives du WISC-V[®] entre les modes d'administration papier-crayon et via Q-interactive[®]

8^e rapport technique de Q-interactive

Mark H. Daniel, PhD,

Dustin Wahlstrom, PhD,

Ou Zhang, PhD,

septembre 2014

Introduction

Q-interactive[®], application numérique de Pearson pour l'administration de tests en face à face via iPads, est conçue pour rendre l'évaluation plus pratique et précise, pour fournir aux cliniciens un accès facile à un grand nombre d'outils et pour permettre la création de nouveaux tests ne pouvant pas être administrés ou cotés sans l'aide d'une assistance informatique.

Avec Q-interactive, l'examineur et le patient utilisent des iPads synchronisés et communiquent par Bluetooth, permettant ainsi au professionnel de lire les consignes d'administration, de relever le temps de réponse du patient, de saisir et d'enregistrer ses réponses (y compris les enregistrements audio) ainsi que de regarder et les stimuli visuels présentés sur l'iPad du patient, qui enregistre par ailleurs les réponses tactiles.

Dans la phase initiale de l'adaptation des tests à la plate-forme Q-interactive, l'objectif était de conserver l'équivalence des notes brutes entre l'administration standard (papier-crayon) et l'administration informatisée, mais également entre les modes de cotation (manuelle ou par assistance informatique). Si l'équivalence est démontrée, les normes, les qualités métriques, la fiabilité et la validité des informations recueillies pour le format papier-crayon peuvent alors s'appliquer aux résultats de Q-interactive.

Il s'agit de la huitième étude d'équivalence de Q-interactive. Dans cette étude, nous avons étudié l'équivalence des résultats entre l'administration standard et l'administration digitale de l'*Échelle d'intelligence de Wechsler pour enfants*[®]-cinquième édition (WISC[®]-V ; Wechsler, 2014).

Dans les deux premières études d'équivalence, les quinze subtests de l'*Échelle d'intelligence de Wechsler pour adultes*[®]-quatrième édition (WAIS[®]-IV ; Wechsler, 2008) ainsi que treize des quinze subtests de l'*Échelle d'intelligence de Wechsler pour enfants*[®]-quatrième édition (WISC[®]-IV ; Wechsler, 2003) ont donné des résultats comparables entre une administration standard (papier-crayon) et une administration digitale via Q-interactive. Sur deux subtests du WISC-IV (Matrices et Complètement d'images), les résultats étaient légèrement supérieurs pour l'administration via Q-interactive. La troisième étude a évalué quatre subtests du *Système d'évaluation des fonctions exécutives de Delis-Kaplan*[™] (D-KEFS[™] ; Delis, Kaplan & Kramer, 2001) ainsi que les essais de rappel libre du *Test d'apprentissage et de mémoire verbale*[®]-deuxième édition (CVLT[®]-II ; Delis, Kramer, Kaplan & Ober, 2000). Elle a démontré une équivalence entre les formats papier-crayon et digital. La quatrième étude a conclu que trois subtests du *Bilan neuropsychologique de l'enfant NEPSY*[®], deuxième édition (NEPSY[®]-II ; Korkman, Kirk, & Kemp, 2007) et deux subtests de l'*Échelle de mémoire pour enfants*[™] (CMS[™] ; Cohen, 1997) étaient équivalents. La cinquième étude a évalué les subtests de Fluidité de la lecture orale et de la Répétition de phrases du *Test de rendement individuel de Wechsler*[®]-troisième édition (WIAT[®]-III ; Wechsler, 2009a) et a jugé qu'ils répondaient au critère d'équivalence. Dans l'étude suivante, tous les subtests de l'*Échelle de mémoire de Wechsler*-quatrième édition (MEM-IV ; Wechsler, 2009b) ont été jugés équivalents. Enfin, les subtests Concepts linguistiques, Exécution de directives, Formulation de phrases et Répétition de phrases, qui font partie de l'*Évaluation clinique des notions langagières fondamentales*-cinquième édition (CELF-5 ; Semel, Wiig, & Secord, 2013) ont été évalués. Les deux premiers subtests, qui

nécessitent que l'examineur enregistre et cote une longue réponse orale, ont été jugés équivalents. Les deux derniers subtests ont été les premiers à appliquer un système de cotation automatique des réponses tactiles du sujet. Cela a donné des résultats équivalents pour le subtest Exécution de directives, mais pas pour Concepts linguistiques, car, lorsque le sujet appuyait sur deux éléments à la fois, cela été considéré à tort comme des erreurs. Lorsque cette erreur de programmation a été corrigée, l'équivalence a été prouvée.

Dans toutes les études d'équivalence, on suppose qu'une administration digitale (via Q-interactive) peut affecter les résultats au test pour un certain nombre de raisons, notamment :

- L'interaction du sujet avec l'iPad. Pour minimiser les effets d'interaction entre le sujet et la tablette qui pourraient nuire à l'équivalence, des manipulables nécessaires à la passation (par ex., la grille de Localisation de points de l'*Échelle de mémoire pour enfants CMS*) et les cahiers de passation (par ex., le test de Traçage de pistes du *Système d'évaluation des fonctions exécutives Delis-Kaplan D-KEFS*) ont été jusqu'à maintenant utilisés dans le cas d'une administration via Q-interactive. L'objectif est, à terme, de remplacer ces éléments par des interfaces numériques interactives, mais cela impliquerait de passer d'une stratégie de démonstration d'équivalence (qui permet de se fonder sur les preuves psychométriques existantes) à celle d'une démonstration de la validité et de la fiabilité de la version électronique. Nous travaillons sur cette étude pour la première fois avec les subtests de vitesse de traitement du WISC-V, mais ce travail est encore en cours et n'est pas décrit dans le présent rapport, qui se concentre uniquement sur l'équivalence.
- L'interaction de l'examineur avec l'iPad, en particulier au cours de la saisie et de la cotation des réponses. À ce jour, la plupart des différences entre les administrations papier-crayon et via Q-interactive se sont produites sur l'interface examinateur. L'administration d'un test via Q-interactive est différente d'une administration classique car Q-interactive comprend des outils et des procédures conçus pour simplifier et aider la tâche de l'examineur. Nous avons pris soin de nous assurer que ces adaptations ne diminuaient pas la précision avec laquelle l'examineur présente les consignes et les stimuli, contrôle et minute les performances, ou saisit et cote les réponses.
- La précision avec laquelle l'application Q-interactive saisit et cote les réponses tactiles du sujet. Q-interactive a d'abord introduit la cotation automatique avec le CELF-5. Dans les précédentes versions, l'examineur devait entrer une note pour chaque item, ce qui maintenait le contrôle de l'examineur mais ne permettait pas d'exploiter pleinement des capacités de la tablette qui peut reconnaître des réponses tactiles et leur attribuer des points.
- Les effets globaux d'un système d'évaluation numérique. Les effets globaux vont bien au-delà de la simple interaction entre le sujet ou l'examineur et l'iPad. Nous avons par exemple observé l'un de ces effets globaux lors d'une étude précédente au cours de laquelle l'examineur utilisait un clavier pour saisir les réponses verbales du sujet. Le sujet semblait alors ralentir le rythme de ses réponses, ou même donner des réponses plus courtes, de manière à éviter d'avoir à attendre que l'examineur, qui tapait assez lentement, ait terminé la saisie de ses réponses textuelles. Nous avons observé un autre type d'effet global auprès de certains très jeunes enfants (âgés de 2 ou 3 ans) qui ont été distraits par la tablette qu'ils percevaient comme un jouet. Nous avons abordé ce problème en nous fondant sur des études

portant sur la problématique des facteurs humains à cet âge.

Dans les études sur Q-interactive, si une activité n'a pas été jugée équivalente entre les deux modes d'administration, nous avons recherché la cause de l'effet numérique dans la passation. Il est essentiel d'en comprendre la cause afin de pouvoir supprimer l'effet du mode d'administration. En principe, si nous trouvons que Q-interactive améliore la précision de l'administration ou de la cotation des examinateurs, cela veut dire que l'application constitue une avancée dans les outils d'évaluation et ce manque d'équivalence ne posera pas nécessairement problème. Il est tout à fait normal qu'une nouvelle technologie ait pour but de produire des résultats au moins équivalents à ceux qu'un examinateur obtiendrait en utilisant correctement le format papier-crayon. Le format électronique ne devrait pas reproduire les erreurs d'administration ou de cotation que l'on trouve avec le format papier-crayon. En revanche, s'il apparaît que l'effet électronique se traduit par une perte de précision de la part du sujet ou de l'examineur, il faudra alors en priorité essayer de modifier l'application Q-interactive afin de supprimer cette source d'erreur. Si cela n'est pas possible, il faudra alors résoudre ce problème par un ajustement des normes.

Dans les précédentes études d'équivalence, la plupart voire l'ensemble des administrations ont été enregistrées sur vidéo, montrant ainsi les interactions entre l'examineur et le sujet par l'intermédiaire de leurs iPads. Ces vidéos permettent ainsi de vérifier que l'administration, l'enregistrement et la cotation sont correctement réalisés dans les deux formats (papier-crayon et électronique). Si l'on détecte un effet de format (c'est-à-dire la non-équivalence entre les deux modes de passation), elles pourront permettre de déterminer la raison de cette différence et de la corriger. Ces enregistrements permettent également d'observer comment les examinateurs et les sujets interagissent avec le matériel de test pour chaque mode de passation. Ces informations peuvent ensuite être utilisées pour orienter les améliorations dans la conception de l'interface. Compte tenu de la vaste expérience que l'équipe de Q-interactive a acquise au cours des sept premières études s'étendant sur deux ans et comprenant plus d'un millier d'administrations, les administrations du WISC-V n'ont pas été enregistrées sur vidéo dans cette étude.

Les études d'équivalence montrent généralement que les sujets âgés de 5 ans et plus (correspondant à l'âge des plus jeunes individus testés) répondent de la même manière, que les stimuli soient présentés sur une tablette numérique plutôt que dans un livret papier ou que leurs réponses tactiles soient saisies à l'écran plutôt qu'observées par un examinateur. Nous avons repéré une seule exception concernant les subtests Matrices et Complètement d'images du WISC-IV. Pour ces deux subtests impliquant le raisonnement conceptuel avec des stimuli visuels complexes, les études semblent indiquer que les enfants réussissent mieux lorsque les stimuli sont présentés sur une tablette. (L'étude actuelle du WISC-V nous permettra de déterminer s'il est possible de reproduire cette conclusion.) En outre, l'ensemble des preuves apportées par ces études indique que, lorsque les examinateurs utilisent les différentes interfaces de Q-interactive, ils obtiennent les mêmes résultats qu'avec une administration papier-crayon.

Conception des études d'équivalence

Nous avons utilisé plusieurs plans expérimentaux pour les études d'équivalence sur Q-interactive. Dans la majorité des cas, chaque sujet ne passe qu'une seule fois un subtest, soit au format digital soit au format standard (papier-crayon). Nous avons opté pour cette solution afin que l'interaction d'un sujet avec la tâche ne puisse pas être affectée par le fait de l'avoir déjà effectuée auparavant. Nous essayons idéalement de détecter les effets que le mode de passation peut avoir sur la façon dont le sujet interagit avec la tâche lors de sa première présentation au sujet. Les plans d'étude dans lesquels chaque sujet ne passe qu'une seule fois un test sont très proches des conditions réelles de test.

Les études sur la WAIS-IV et le WISC-IV reposent essentiellement sur un protocole à *groupes équivalents*, avec une répartition aléatoire ou non aléatoire des sujets dans les groupes. Cette méthode compare les performances de deux groupes, l'un passant le test dans le format électronique et l'autre dans le format papier-crayon. Le protocole à groupes équivalents est décrit en détail dans les Rapports techniques 1 et 2 de Q-interactive. Nous avons utilisé ce protocole, avec répartition aléatoire, pour l'étude du WISC-V.

Il convient également d'utiliser un autre type de protocole de passation unique, appelé *double saisie*, lorsque le format électronique influe sur la façon dont l'examineur saisit et cote les réponses, sans pour autant influencer sur le comportement du sujet. Un nombre relativement restreint de sujets ne passe le test qu'une seule fois, mais l'administration est enregistrée sur vidéo du point de vue de l'examineur afin qu'elle puisse être regardée par un certain nombre de correcteurs, qui la cotent en utilisant soit le format papier-crayon soit le format électronique. Une comparaison des résultats moyens pour les deux modes de passation permet d'indiquer si le format influe sur les processus de saisie et de cotation des réponses. Vous pouvez trouver davantage de détails sur cette méthode dans les rapports techniques n°3 (CVLT-II et D-KEFS), n°5 (WIAT-III) et n°6 (MEM-IV-seul ce dernier est disponible en français).

Dans la troisième méthode, appelée *test-retest*, chaque sujet passe deux fois le subtest, une fois dans chaque format (en ordre contrebalancé). Lorsque la méthode test-retest est possible, elle est très efficace, car le sujet est son propre témoin. Ce protocole est adapté lorsque le processus de réponse est peu susceptible de changer au test-retest, parce que le sujet n'apprend ni solutions ni nouvelles stratégies pour aborder la tâche ou résoudre le problème. Nous avons utilisé la méthode test-retest pour l'étude de suivi des subtests de vitesse de traitement de la WAIS-IV (Rapport technique n°1) ainsi que pour les études de la NEPSY-II (Rapport technique n°4), de la MEM-IV (Rapport technique n°6) et de la CELF-5 (Rapport technique n°7).

Pour toutes les études d'équivalence, nous avons utilisé une taille d'effet inférieure ou égale à 0,2 comme norme d'équivalence. La taille d'effet correspond à la différence moyenne de notes entre une administration papier-crayon et une administration via Q-interactive, divisée par l'écart-type des notes de la population. Une taille d'effet de 0,2 est légèrement supérieure à un demi-point de note étalonnée par rapport à la métrique de subtest couramment utilisée, avec une moyenne de 10 et un écart-type de 3.

Mode de sélection des sujets

Les études d'équivalence sur Q-interactive (y compris celle-ci) ont utilisé des échantillons de sujets sains afin de maintenir l'accent sur l'estimation de la présence et de la taille d'effets avec le format électronique. Comme les effets possibles de la passation assistée par ordinateur sur les individus ayant des conditions cliniques particulières ne sont pas connus, l'inclusion de patients souffrant de troubles divers dans l'échantillon pourrait compromettre la bonne interprétation des résultats. Même si, dans un second temps, il sera important pour les applications cliniques de Q-interactive de comprendre l'interaction du mode de passation avec des conditions cliniques particulières, la recherche initiale doit se concentrer principalement sur la question de savoir si le format électronique influe sur les résultats obtenus par des sujets sains.

Le nombre de variables démographiques nécessaires pour contrôler l'échantillon dépend du type de protocole. Dans un protocole à groupes équivalents, il est important que les échantillons comparés soient représentatifs de la population générale (sexe, origine ethnique, niveau socio-économique [niveau d'études]) et que les deux groupes soient démographiquement semblables. Dans les méthodes test-retest et à double saisie, qui mettent l'accent sur les comparaisons entre sujets, les caractéristiques des sujets sont moins significatives. Il est toutefois important que l'échantillon propose des niveaux de compétence et des styles de réponses suffisamment diversifiés pour produire des réponses variées afin de pouvoir évaluer les différentes fonctions de l'interface numérique.

Les examinateurs participant aux études d'équivalence ont été formés aux procédures d'administration standard des tests (papier-crayon). Les examinateurs ont également reçu une formation théorique et pratique suffisante dans les procédures d'administration digitale et de cotation pour être en mesure de mener à bien la passation et de saisir sans problème les réponses, sans avoir à trop se concentrer sur le format. D'après notre expérience, il ressort qu'il faut beaucoup de pratique pour se familiariser complètement à un nouveau mode d'administration.

Étude d'équivalence du WISC-V

Méthode

Mesures

Le WISC-V est un instrument complet utilisé pour évaluer la capacité intellectuelle de sujets âgés de 6 à 16 ans 11 mois. Un grand nombre des subtests proposent une administration et une cotation presque identiques à celles du WISC-IV, dont l'équivalence entre formats papier-crayon et Q-interactive a déjà été évaluée. Toutefois, comme plusieurs des subtests du WISC-V sont nouveaux et que certains ont subi de légères modifications dans l'administration ou la cotation (y compris les modifications apportées à l'interface Q-interactive de l'examineur), nous avons réévalué l'équivalence de la totalité des subtests. Notez que certaines des épreuves du WISC-V américain n'ont pas été adaptées en français.

La version électronique des trois subtests de Vitesse de traitement (Code, Symboles et Barrage), dans lesquels le sujet répond en touchant la tablette ou en dessinant dessus, a été administrée dans l'étude d'équivalence du WISC-V. Afin de réduire les différences non pertinentes par rapport au construit entre les versions papier-crayon et électronique, nous avons remanié le matériel de test papier du WISC-IV afin qu'il ressemble au format Q-interactive. Néanmoins, nous ne nous attendions pas à une équivalence totale des résultats bruts entre les versions papier-crayon et électronique à cause de la différence du mode de réponse. D'après l'objectif du projet de développement, les deux versions de chaque subtest devaient mesurer le même construit et avoir les mêmes propriétés psychométriques afin qu'elles soient comparables et puissent être considérées comme deux formes équivalentes à des fins d'interprétation clinique. Les résultats ont toutefois révélé que les versions papier-crayon et digitale n'étaient pas suffisamment proches pour pouvoir répondre à cet objectif. Des travaux complémentaires de conception et de développement de ces subtests sont en cours. La version initiale du WISC-V continue d'utiliser les cahiers de passation pour les subtests de Vitesse de traitement, en complément de l'interface examinateur de Q-interactive pour la prise de temps et l'enregistrement, qui est identique à celle utilisée dans les études du WISC-IV et de la WAIS-IV, dont l'équivalence au format papier-crayon a été démontrée. Comme le sujet travaille sur un cahier de passation et que l'interface examinateur est identique à celle du WISC-IV, ces subtests n'ont pas été réévalués.

Sujets

Cette étude a été réalisée dans le cadre de la normalisation du WISC-V. Presque toutes les administrations ont été réalisées en avril et en mai 2014 ; seule une faible part d'administrations via Q-interactive a été réalisée aux Etats Unis durant les quelques mois précédents. L'échantillon des sujets était composé d'enfants américains âgés de 6 à 16 ans 11 mois. L'équipe Pearson chargée de la recherche sur le terrain a recruté les sujets et les a indemnisés pour leur participation. Les sujets potentiels ont été présélectionnés en fonction de leurs caractéristiques démographiques et de facteurs d'exclusion, notamment des troubles moteurs ou perceptifs ou bien des conditions cliniques. Le plan d'échantillonnage exige un nombre à peu près équivalent de garçons et de filles, une distribution sensiblement équivalente des sujets par groupe d'âge ainsi qu'une bonne représentativité de chaque groupe ethnique et de chaque niveau socioéconomique (niveau d'études des parents).

Les sujets potentiels dont les caractéristiques correspondaient aux exigences démographiques à la fois des échantillons papier-crayon et Q-interactive ont été répartis aléatoirement à l'un des deux modes de passation. Après la fin de la collecte des données, l'équipe de recherche a sélectionné en vue d'analyse des paires de sujets (un membre pour chaque mode de passation) appariés par tranche d'âge, par sexe, par origine ethnique et par niveau d'études des parents.

Les examinateurs étaient basés à divers endroits aux Etats Unis. Tous les examinateurs ont reçu une formation à distance sur le WISC-V et sur l'administration via Q-interactive afin qu'ils aient le droit d'administrer le test dans l'un ou l'autre mode de passation. Ils se sont donc entraînés à faire des passations et nous leur avons fait un retour sur leurs erreurs d'administration ou de cotation. Les examinateurs qui ne faisaient pas partie du personnel de Pearson ont été indemnisés pour leur participation.

Procédure

Les examinateurs ont saisi les éléments de réponse en employant la méthode classique, telle qu'elle est déployée lors des phases de recueil de données pour la normalisation des tests : ils ont notamment transcrit fidèlement les réponses pour chaque item verbal de subtest et coté tous les items. L'équipe de recherche Pearson a vérifié les cas d'administration papier-crayon pour s'assurer du bon respect des règles d'administration (notamment les interrogations, les règles d'entrée et les critères d'arrêt), mais n'a pas recoté les items. Tous les résultats bruts aux subtests ont été automatiquement calculés, soit par les équipes Pearson à l'aide des critères de cotation classiques ou via l'application Q-interactive.

L'effet du mode d'administration sur chaque subtest a été calculé en utilisant une méthode de régression multiple, dans laquelle les variables indépendantes correspondaient aux données démographiques (âge, sexe, origine ethnique et niveau d'études des parents), aux résultats à d'autres indices ou subtests du WISC-V et au mode de passation tandis que la variable dépendante correspondait à la note étalonnée au subtest. Bien que l'utilisation de paires de sujets appariés démographiquement et répartis aléatoirement en mode de passation assure un bon contrôle expérimental, l'utilisation de covariables améliore encore la fiabilité statistique. Nous avons utilisé les subtests du WISC-V plutôt que des tests externes comme covariables parce que l'étude sur le WISC-IV avait déjà démontré que, à quelques exceptions près, le mode de passation n'influe que très peu sur les subtests. L'analyse a été effectuée en deux étapes. Lors de la première étape, nous avons utilisé comme variables prédictives les subtests ou les indices qui avaient présenté peu d'effets du mode d'administration dans l'étude d'équivalence du WISC-IV : Cubes, Arithmétique, Mémoire des chiffres et l'Indice de Compréhension verbale. (Pour l'analyse de chaque subtest verbal, l'Indice de Compréhension verbale a été remplacé par les subtests Similitudes et Vocabulaire.) Nous avons utilisé les résultats de cette phase pour affiner l'ensemble des variables prédictives et nous avons répété l'analyse durant la deuxième étape.

Comme le format papier-crayon a été codé 0 et le format Q-interactive 1, le poids de régression non normalisé pour le mode de passation a permis d'évaluer directement l'effet du mode d'administration au niveau des unités de score normalisé pour ce subtest. Ce chiffre a été converti en taille d'effet en le divisant par l'écart-type du score normalisé (3 ou 15). Si nous supposons une corrélation multiple de 0,5 entre l'ensemble des prédicteurs démographiques et de résultat au

subtest et le résultat au subtest cible, un échantillon total de 350 sujets a une puissance de 0,56 pour détecter une taille d'effet de 0,20 (alpha = 0,05).

Afin de déterminer s'il pouvait y avoir un effet du mode de passation pour les sous-groupes de la population selon l'âge, le sexe, le niveau socio-économique, l'appartenance ethnique ou le niveau d'efficacité cognitive globale, nous avons calculé un indicateur d'effet du mode d'administration pour chaque sujet utilisant Q-interactive et nous avons analysé la relation de cet indicateur pour chacune des caractéristiques individuelles. L'indicateur d'effet du mode d'administration correspondait à la différence entre le score réel du sujet au subtest (avec une administration digitale) et le score estimé du sujet pour le format papier-crayon en fonction des variables démographiques et d'autres résultats au subtest du WISC-V.

Résultats

Nous avons sélectionné un total de 350 sujets (175 paires appariés) pour l'analyse. Sur les 113 examinateurs participants, 58 % ont fait une passation dans les deux formats ; 38 % ont fait une passation seulement via Q-interactive et 4 % ont uniquement fait une passation papier-crayon. Le tableau 1 dresse les caractéristiques des échantillons de sujets ayant fait une passation papier-crayon et via Q-interactive. Les deux échantillons ont des variables démographiques très similaires. Dans l'ensemble, la représentativité de l'échantillon en termes d'âge et de proportion des groupes ethniques à l'échelle nationale est presque assurée. Les filles et les enfants dont les parents sont diplômés de l'université sont légèrement surreprésentés par rapport à la population générale.

Tableau 1 : Caractéristiques démographiques des échantillons

Caractéristiques démographiques		Mode de passation		
		Papier-crayon	Q-interactive	
Nombre de paires		175	175	
Âge (en années)	6	14	16	
	7	18	15	
	8	15	16	
	9	15	15	
	10	12	12	
	11	20	20	
	12	14	14	
	13	17	17	
	14	18	18	
	15	16	16	
	16	16	16	
		Moyenne	11,1	11,1
		Écart-type	3,2	3,2
Sexe	Filles	101	102	
	Garçons	74	73	
Origine ethnique	Afro-américain	18	19	
	Hispanique	29	32	
	Blanc	117	115	
	Autres	11	9	
Niveau d'étude des parents	< 12 ans	17	12	
	Bac	30	29	
	Post-bac	54	55	
	Licence	74	79	

Le tableau 2 indique les moyennes et les écarts-types des résultats obtenus aux subtests du WISC-V pour chaque format. Étant donné la grande similitude des caractéristiques démographiques des deux groupes établis en fonction du mode de passation et le fait que le format ait été attribué aléatoirement aux sujets, nous ne nous attendions pas à observer des différences nettes ou systématiques dans les résultats entre les groupes.

Tableau 2 : Statistiques descriptives pour les subtests du WISC-V, par mode de passation

Subtest	Papier-crayon		Q-interactive	
	Moyenne	Éc	Moyenne	Éc
Arithmétique	10,3	2,6	9,9	2,4
Cubes	9,9	2,5	10,6	2,5
Compréhension	9,9	2,6	9,5	2,6
Mémoire des chiffres	10,1	2,7	10,6	2,5
Balances	10,0	2,6	10,6	2,9
Information	10,2	3,0	10,2	2,6
Séquence lettres-chiffres	10,3	2,5	10,7	2,4
Matrices	9,9	2,5	10,6	2,9
Identification de concepts	9,9	2,9	10,1	3,2
Mémoire des images	10,3	2,5	10,7	2,7
Similitudes	10,2	2,8	10,3	2,6
Puzzles visuels	9,8	2,6	10,0	2,7
Vocabulaire	10,0	3,0	9,8	2,6
Association immédiate de symboles	99,4	13,4	100,7	13,2
Association différée de symboles-rappel libre	99,9	13,2	100,8	13,9
Association différée de symboles-rappel indicé	101,7	12,6	102,5	13,2
Dénomination rapide de lettres-chiffres	100,7	13,9	103,0	14,2
Dénomination rapide de quantités	101,4	14,5	102,2	12,6

Les résultats issus de la première étape de l'analyse de régression ont conduit à la sélection d'un nouveau jeu de variables prédictives qui présentent un effet du mode d'administration relativement faible en vue de leur utilisation dans la deuxième étape : Puzzles visuels, Identification de concepts, Mémoire des images, Séquence lettres-chiffres et l'Indice de Compréhension verbale (uniquement les subtests Similitudes ou Vocabulaire pour l'analyse des subtests verbaux). Le tableau 3 montre les résultats de la deuxième série d'analyses de régression, y compris la corrélation multiple avec les variables prédictives, le poids de régression non normalisé pour le mode de passation, la valeur *t* et la significativité statistique associée au mode de passation comme prédicteur et la taille d'effet

pour le mode de passation. Les résultats de l'étude d'équivalence sur le WISC-IV sont indiqués à titre de comparaison.

Des corrélations multiples allant de 0,35 à 0,71 (médiane = 0,52), indiquant que les données démographiques et les variables d'aptitude représentaient, en moyenne, environ un quart de la variance dans les résultats aux subtests. Le but de ces variables prédictives était d'accroître la puissance des analyses en réduisant la quantité de variance à expliquer. Ainsi, des corrélations multiples élevées sont certes souhaitables, mais pas essentielles pour la validité des analyses.

Tableau 3 : Taille d'effet du mode de passation via Q-interactive sur chaque subtest du WISC-V

Subtest	R	Poids de régression		Taille d'effet	Taille d'effet du WISC-IV
		non normalisé	t		
Arithmétique	.55	-0,49	-2,11*	-0,16	0,10
Cubes	.58	0,59	2,66**	0,20	0,02
Compréhension	.55	-0,59	-2,51*	-0,20	0,00
Mémoire des chiffres	.54	0,25	1,04	0,08	0,13
Balances	.52	0,49	1,95	0,16	—
Information	.71	-0,15	-0,68	-0,05	0,07
Séquence lettres-chiffres	.50	0,26	1,13	0,09	0,18
Matrices	.48	0,51	1,99	0,17	0,27
Identification de concepts	.40	0,07	0,22	0,02	0,21
Mémoire des images	.42	0,21	0,83	0,07	—
Similitudes	.66	0,11	0,50	0,04	0,02
Puzzles visuels	.52	0,11	0,46	0,04	—
Vocabulaire	.66	-0,39	-1,69	-0,13	0,05
Association immédiate de symboles	.57	0,52	0,44	0,03	—
Association différée de symboles-rappel libre	.50	0,21	0,16	0,01	—
Association différée de symboles-rappel indicé	.44	-0,07	-0,06	0,00	—
Dénomination rapide de lettres-chiffres	.41	1,73	1,24	0,12	—
Dénomination rapide de quantités	.35	-0,37	-0,27	-0,02	—

Remarque : Une taille d'effet positive indique des scores plus élevés avec Q-interactive. Le poids de régression non normalisé est une mesure du score normalisé au subtest (écart-type de 3 pour les 13 premiers subtests et de 15 pour les cinq derniers).

*p < 0,05, **p < 0,01

Trois des dix-huit subtests ont montré un effet du mode d'administration statistiquement significatif autour de 0,05, mais aucune des tailles d'effet ne dépassait le critère de 0,20, correspondant à la norme dans les études d'équivalence sur Q-interactive. Parmi les dix subtests communs au WISC-IV et au WISC-V, les tailles d'effet ont tendance à être relativement semblables dans les deux études ($r = 0,44$), mais plusieurs différences importantes sont apparues. Les deux subtests ayant les tailles d'effet les plus élevées sur le WISC-V (Cubes et Compréhension), présentaient des tailles d'effet proches de 0 dans l'étude sur le WISC-IV. D'autre part, les deux subtests ayant des tailles d'effet supérieures à 0,20 dans l'étude sur le WISC-IV, présentaient des tailles d'effet modérées (Matrices, 0,17) à faibles (Identification de concepts, 0,02) dans l'étude sur le WISC-V. En outre, alors que tous les tailles d'effet du WISC-IV (à l'exclusion de l'Indice de Vitesse de traitement) étaient positives, les tailles d'effet dans l'étude du WISC-V étaient plus équilibrées (cinq négatives et treize positives).

Nous avons cherché à déterminer dans une série d'analyses s'il existait des effets différentiels entre les modes d'administration en fonction du niveau d'aptitude ou des caractéristiques démographiques des sujets. Nous avons calculé un indicateur de l'effet du mode d'administration pour chaque sujet ayant passé un subtest via Q-interactive. Cet indicateur correspondait à la différence entre le score réel du sujet au subtest et le score estimé du sujet pour le format papier-crayon, en utilisant comme prédicteurs les variables démographiques et les résultats au subtest obtenus lors des analyses de la deuxième étape. La mesure de la relation correspondait, pour les variables prédictives continues (aptitude, âge et niveau socio-économique), au coefficient de corrélation entre chaque variable et l'indicateur d'effet du mode d'administration ; pour le sexe, à la statistique t ; et pour l'origine ethnique, à la statistique F à partir de l'analyse de variance. Dans cette analyse, le score d'aptitude correspondait au résultat estimé pour une administration papier-crayon. Les résultats sont indiqués dans le tableau 4.

Tableau 4 : Relation de l'effet du mode d'administration sur le niveau d'aptitude et les données démographiques

Subtest	Corrélation			Sexe (<i>t</i>) ^b	Origine ethnique (<i>F</i>)
	Capacité ^a	Age	Niveau socio- économique		
Arithmétique	-.01	.03	-.05	-0,40	0,06
Cubes	.02	.04	.03	0,08	0,67
Compréhension	.00	-.05	-.02	0,63	0,88
Mémoire des chiffres	-.03	.06	-.01	-0,80	0,33
Balances	.03	-.03	-.04	-0,46	0,42
Information	-.10	-.07	-.04	-0,21	0,78
Séquence lettres-chiffres	-.14	.03	-.06	0,83	1,08
Matrices	.02	-.01	-.01	0,55	0,17
Identification de concepts	.04	.04	-.02	-0,06	0,59
Mémoire des images	.04	.05	.09	-0,20	0,66
Similitudes	.04	.07	-.04	-0,60	0,47
Puzzles visuels	.09	-.07	.11	0,24	0,91
Vocabulaire	-.11	-.01	-.07	0,37	0,69
Association immédiate de symboles	-.02	-.06	-.01	-0,70	1,26
Association différée de symboles-rappel libre	-.02	-.10	-.07	-0,35	2,17
Association différée de symboles-rappel indicé	-.02	-.07	-.01	-0,13	0,98
Dénomination rapide de lettres-chiffres	-.10	.03	-.08	-1,80	0,10
Dénomination rapide de quantités	-.06	.15*	-.12	-2,00*	0,64

Remarque : $N = 175$. Pour cette analyse, l'effet du mode d'administration correspond à la différence entre le score réel au subtest avec une administration via Q-interactive et le score estimé pour un mode de passation papier-crayon en fonction des variables démographiques et des résultats à d'autres subtests.

^a Score estimé au subtest pour un mode de passation papier-crayon.

^b Une valeur positive de t signifie que l'effet du mode d'administration est plus élevé chez les filles.

* $p < 0,05$.

Parmi 90 tests statistiques, nous pensions avoir une chance d'obtenir quatre ou cinq résultats statistiquement significatifs (autour de 0,05), mais en réalité seulement deux des résultats présentés dans le tableau 4 étaient statistiquement significatifs. Les deux concernaient le subtest Dénomination rapide de quantités (non intégré dans la version française de l'outil), pour lequel une administration via Q-interactive avait tendance à favoriser les sujets plus âgés par rapport aux plus jeunes et les garçons par rapport aux filles. Dans l'ensemble, les résultats indiqués dans le tableau 4 confirment les résultats déjà obtenus dans l'étude sur le WISC-IV : aucun effet différentiel avec l'administration via Q-interactive selon le niveau d'aptitude, l'âge, le niveau socio-économique, le sexe ou l'origine ethnique.

Discussion

Cette étude réplique l'étude d'équivalence du WISC-IV pour dix des subtests du WISC-V ainsi que l'étude initiale des huit nouveaux subtests. En total accord avec les résultats des sept études d'équivalence précédentes, toutes les tailles d'effet du mode d'administration entraient dans le critère établi pour l'équivalence de Q-interactive (à savoir, au maximum 0,20). Les deux subtests qui avaient montré des tailles d'effet supérieures à 0,20 au WISC-IV (Matrices et Identification de concepts) présentaient des tailles d'effet inférieures dans cette étude. Sur les deux subtests qui montraient les tailles d'effet les plus élevées dans l'étude sur le WISC-V (+0,20 pour Cubes et -0,20 pour Compréhension), le sujet n'a pas interagi ou très peu avec la tablette : pour Cubes, la tablette du sujet se contente d'afficher les modèles à reproduire et, pour Compréhension, la tablette du sujet n'est pas du tout utilisée. En outre, ces subtests montraient respectivement des tailles d'effet de seulement 0,02 et de 0,00, dans l'étude sur le WISC-IV.

Au cours de la série d'études d'équivalence sur Q-interactive, il est devenu clair que, bien que de faibles effets puissent être observés dans certaines études, à moins qu'une cause ne puisse être détectée (par exemple, en observant les vidéos prises pendant les passations), ces effets ne sont généralement ni reproductibles ni systématiques et sont vraisemblablement dus en grande partie au fait qu'il n'existe pas de plan d'étude idéal pour évaluer parfaitement ces questions. Néanmoins, les études ont permis de relever à plusieurs reprises de réelles failles qui ont été corrigées avant publication, mais qui auraient pu autrement passer inaperçues (notamment, l'erreur dans l'enclenchement du chronométrage pour les subtests de Vitesse de traitement de la WAIS-IV ou le léger problème de qualité d'image pour le subtest Complètement d'images de la WAIS-IV). C'est la combinaison de résultats empiriques et d'un examen minutieux des procédures d'administration et de cotation qui offre un niveau de confiance dans l'équivalence qui ne pourrait pas être obtenue seulement à partir de l'une ou l'autre méthode.

La présente étude a également montré qu'il n'y avait pratiquement aucune différence statistiquement significative dans l'effet du mode d'administration entre les sous-groupes selon l'âge, le sexe, l'origine ethnique, le niveau socio-économique ou le niveau d'aptitude. De même, ces résultats ont toujours été corroborés par les précédentes études d'équivalence sur Q-interactive. D'après cet ensemble de résultats, nous pouvons en conclure que l'absence d'effet du mode d'administration s'applique globalement à la population générale saine.

Références bibliographiques

- Cohen, M. (1997). *Children's memory scale*. Bloomington, MN: Pearson.
- Daniel, M. H. (2012a). *Equivalence of Q-interactive administered cognitive tasks: WAIS®-IV*. (Q-interactive Technical Report 1). Bloomington, MN: Pearson.
- Daniel, M. H. (2012b). *Equivalence of Q-interactive administered cognitive tasks: WISC®-IV*. (Q-interactive Technical Report 2). Bloomington, MN: Pearson.
- Daniel, M. H. (2012c). *Equivalence of Q-interactive administered cognitive tasks: CVLT®-II and selected D-KEFS® subtests* (Q-interactive Technical Report 3). Bloomington, MN: Pearson.
- Daniel, M. H. (2013a). *Equivalence of Q-interactive and paper administrations of cognitive tasks: Selected NEPSY®-II and CMS subtests* (Q-interactive Technical Report 4). Bloomington, MN: Pearson.
- Daniel, M. H. (2013b). *Equivalence of Q-interactive and paper scoring of academic tasks: Selected WIAT®-III subtests*. (Q-interactive Technical Report 5). Bloomington, MN: Pearson.
- Daniel, M. H. (2013c). *Equivalence of Q-interactive and paper administration of WMS®-IV cognitive tasks* (Q-interactive Technical Report 6). Bloomington, MN: Pearson.
- Daniel, M. H., Wahlstrom, D., & Zhou, X. (2014). *Equivalence of Q-interactive® and paper administrations of language tasks: Selected CELF®-5 tests* (Q-interactive Technical Report 7). Bloomington, MN: Pearson.
- Delis, D., Kaplan, E., & Kramer, J. (2001). *Delis-Kaplan executive function system®*. Bloomington, MN: Pearson.
- Delis, D., Kramer, J., Kaplan, E., & Ober, B. (2000). *California verbal learning test®*, second edition. Bloomington, MN: Pearson.
- Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY®*—second edition. Bloomington, MN: Pearson.
- Semel, E., Wiig, E. H., & Secord, W. A. (2013). *Clinical evaluation of language fundamentals®*—fifth edition. Bloomington, MN: Pearson.
- Wechsler, D. (2003). *Wechsler intelligence scale for children®*—fourth edition. Bloomington, MN: Pearson.
- Wechsler, D. (2008). *Wechsler adult intelligence scale®*—fourth edition. Bloomington, MN: Pearson.
- Wechsler, D. (2009a). *Wechsler individual achievement test®*—third edition. Bloomington, MN: Pearson.
- Wechsler, D. (2009b). *Wechsler memory scale®*—fourth edition. Bloomington, MN: Pearson.
- Wechsler, D. (2014). *Wechsler intelligence scale for children®*—fifth edition. Bloomington, MN: Pearson.