



Q-interactive™

Équivalence des tâches cognitives de la WAIS-IV® entre les modes d'administration papier-crayon et via Q-interactive®

1^{er} rapport technique de Q-interactive

Mark H. Daniel, PhD

Cadre scientifique pour l'innovation et la recherche

Juin 2012

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 permettra 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 sujet utilisent des iPads synchronisés et communiquant par Bluetooth, permettant ainsi au professionnel de lire les consignes d'administration, de relever le temps de réponse du sujet, 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 sujet, 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.

Pour cette raison, nous avons utilisé pour l'administration via Q-interactive les manipulables nécessaires à la passation (notamment, les cubes du subtest Cubes des échelles de Wechsler) et les cahiers de passation (subtests de Vitesse de traitement des échelles de Wechsler). Même si ces manipulables pourraient éventuellement être remplacés par des interfaces numériques interactives, il serait alors improbable de pouvoir conserver l'équivalence des notes brutes à cause du niveau d'adaptation nécessaire.

Le but de cette étude était d'évaluer l'équivalence des résultats entre une administration classique et une administration via Q-interactive de l'*Échelle d'intelligence de Wechsler pour adultes-4^e édition* (WAIS-IV ; Wechsler, 2008). Le mode d'administration digital (via Q-interactive) pourrait affecter les résultats au test pour de multiples raisons, notamment :

- L'interaction du sujet avec l'iPad ;
- L'interaction de l'examineur avec la tablette, en particulier au cours de l'enregistrement et de la cotation des réponses ;
- Les effets globaux d'un système d'évaluation digitale.

La plupart des différences introduites dans la première version de Q-interactive se situent au niveau de 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 observé l'un de ces effets globaux dans une étude précédente portant sur un modèle d'interface dans lequel l'examineur utilisait un clavier pour saisir les réponses verbales du sujet. Le sujet semblait alors ralentir le rythme de ses réponses pour ne pas prendre d'avance sur l'examineur. Ce type d'interface a donc été abandonné.

Dans cette étude, si une activité n'a pas été jugée équivalente entre les deux formats d'administration (ce que l'on appelle un effet numérique), nous avons recherché la cause de l'effet numérique dans la passation. 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 numérique se traduit par une perte de précision de la part du sujet ou de l'examineur, il faudra donc en priorité essayer de modifier l'application Q-interactive afin de supprimer cette source d'erreur. Si cela n'est pas possible, il conviendra alors de résoudre ce problème par un ajustement des normes. La collecte d'informations par l'intermédiaire d'enregistrements vidéo s'est révélée essentielle pour évaluer l'administration et la précision de la cotation lors de l'étude d'équivalence.

Conception des études d'équivalence

Nous avons utilisé plusieurs plans expérimentaux pour la présente étude. La plupart ont fait appel à un plan à groupes équivalents avec répartition aléatoire ou non aléatoire dans lequel chaque sujet n'a passé qu'une seule fois un test, soit sous format classique (papier-crayon) soit sous format digital, de sorte que leur expérience était très semblable à celle de conditions cliniques réelles. Ce plan exige de plus grands échantillons qu'une méthode test-retest ou des formes parallèles, mais atténue le risque d'effets de pratique. Certains subtests de la WAIS-IV induisent des effets de pratique lorsqu'ils sont administrés dans leur format final. Lorsqu'un sujet a résolu une fois un problème, le résoudre une seconde fois implique un processus différent (peu de tests disponibles sur Q-interactive présentent des formes parallèles). En passant un test une seconde fois, les processus cognitifs utilisés peuvent être radicalement différents, réduisant ainsi les risques de détecter un effet du mode d'administration. Un plan à groupes équivalents, qui compare les performances des deux groupes, l'un passant le test dans sa version digitale et l'autre dans sa version papier-crayon, permet d'éviter ces problèmes.

Plan à groupes équivalents avec répartition aléatoire

L'échantillon devrait être représentatif de la population générale en termes de sexe, d'origine ethnique et de niveau d'étude. La répartition du nombre de sujets par groupe d'âge devrait refléter les questions issues de la recherche (par exemple, les tranches d'âge pour lesquelles on suspecte un risque particulier de non-équivalence devraient être surreprésentées). À l'intérieur de chaque cellule démographique (combinaison de sexe, origine ethnique et niveau d'étude), nous avons réparti aléatoirement la moitié des sujets dans chaque mode de passation.

Juste après la passation d'essai, tous les sujets ont passé une série de tests de covariables au format papier-crayon qui mesurent les mêmes construits que ceux du test au format électronique (le test de convergence).

Les résultats de chaque test ou subtest de convergence sont ensuite analysés séparément, en utilisant une régression multiple (ou une analyse de covariance, ANCOVA). Dans la méthode de régression, les prédicteurs sont les scores de référence, ajustés selon l'âge, obtenus aux tests de covariables, aux variables démographiques et à une variable muette qui représente le mode de passation. La variable dépendante est le score de référence, ajusté selon l'âge, obtenu au test de convergence. Le poids de régression non normalisé concernant le mode de passation est une mesure de l'effet du mode d'administration, exprimé en un score de référence du test de convergence.

L'équipe de Q-interactive a choisi d'utiliser une taille d'effet inférieure à 0,2 comme norme d'équivalence. (Cela correspond à une taille légèrement supérieure à un demi-point de note étalonnée par rapport à la mesure de subtest de l'échelle de Wechsler, avec une moyenne de 10 et un écart-type de 3). Si la combinaison des données démographiques et des tests de covariables obtient une corrélation multiple de 0,7 avec le score au test en cours d'analyse (une valeur classique pour la WAIS-IV), il est alors nécessaire d'avoir 200 sujets par mode de passation (soit un total d'environ 400) pour obtenir une puissance de 0,8, indispensable pour pouvoir détecter une taille d'effet de 0,2 (avec un alpha de 0,05).

L'avantage du plan à groupes équivalents avec répartition aléatoire, c'est que l'assignation aléatoire des sujets au mode de passation a tendance à rendre les sous-échantillons comparés équivalents au niveau de toutes les caractéristiques qui pourraient influencer sur la performance du test, y compris celles qui ne sont pas mesurées (ou ne peuvent pas être mesurées). Cet avantage n'est toutefois possible qu'au prix d'un échantillon relativement vaste.

Plan à groupes équivalents avec répartition non aléatoire

Cette méthode tire parti du large échantillon témoin soigneusement stratifiée qui existe pour chaque test et réduit la taille de l'échantillon nécessaire. Elle peut être utilisée lorsque le test de convergence contient au moins deux subtests qui mesurent le même construit de compétence ou que les sujets de l'échantillon témoin ont passé des tests de covariables externes. Lorsque deux subtests mesurent le même construit, ils servent de tests de covariables.

Dans ce plan, l'échantillon témoin existant sert d'échantillon pour la passation papier-crayon ; seul l'échantillon pour l'administration digitale doit être recueilli. L'échantillon digital doit avoir les mêmes caractéristiques démographiques que l'échantillon témoin. La grande taille de l'échantillon témoin permet de réduire la taille de l'échantillon électronique nécessaire pour atteindre le degré de puissance voulu. On considère que les résultats statistiques de l'échantillon témoin constituent des paramètres de population, permettant ainsi l'utilisation d'un test à échantillon unique pour déterminer si les résultats statistiques de l'échantillon digital diffèrent de ces paramètres.

Parmi les méthodes d'analyse des données à partir de ce plan, on peut établir des équations de prédiction pour les scores du test de convergence d'après les données démographiques et les résultats aux tests de covariables, en utilisant les données de l'échantillon témoin. On applique ces équations à l'échantillon de la version digitale et la différence moyenne entre les résultats observés et estimés au test de convergence sert à estimer l'effet du mode d'administration. Si la corrélation multiple des tests de covariables et des variables démographiques avec les résultats au test de convergence est égale à 0,75 (valeur classique pour la WAIS-IV), un échantillon du mode digital d'environ 100 cas permet alors d'obtenir une puissance de 0,8 pouvant détecter une taille d'effet de 0,2 (avec un alpha de 0,05).

Autres plans

La nature d'un test se prête parfois à un type de plan plus efficace dans lequel les sujets sont leurs propres témoins, comme le test-retest et la double saisie. (Un plan à formes parallèles n'a pas été possible car les subtests de la WAIS-IV n'ont pas d'autres formes.)

Méthode test-retest

Dans la méthode test-retest, chaque sujet passe le test deux fois et l'ordre d'administration est contrebalancé dans l'échantillon. Ce plan est adapté lorsque le processus de réponse est peu susceptible de changer au retest, parce que le sujet n'apprend pas de solutions spécifiques à un problème ni de stratégies pour résoudre le problème. Parmi ce genre de test, nous trouvons par exemple des mesures de la vitesse de traitement ou de la mémoire à court terme pour des stimuli abstraits.

Lorsque la méthode test-retest est possible, elle est très efficace, car le sujet est son propre témoin. Dans cette méthode, chaque sujet passe le test deux fois, une fois dans chaque format, et l'ordre d'administration est contrebalancé, de sorte que la moitié des sujets passe d'abord un format tandis que l'autre moitié passe d'abord l'autre format.

On analyse une étude d'équivalence test-retest en calculant pour chaque sujet la différence entre les scores de la première passation et ceux de la deuxième passation. Si nous n'observons aucun effet du mode d'administration, la valeur moyenne de ces écarts sera identique indépendamment de l'ordre de passation. Toutefois, si nous observons un effet du mode d'administration, les valeurs moyennes de ces écarts entre les deux groupes d'ordre de passation correspondront au double ou à la moitié de la taille de l'effet : dans un groupe, l'effet multipliera par deux la valeur moyenne de ces écarts tandis que, dans l'autre groupe, il la divisera par deux. L'utilisation de paires de sujets démographiquement appariés dans les deux groupes d'ordre de passation permet de produire une puissance statistique élevée avec des échantillons de petite taille. En supposant une corrélation au retest de 0,8, il faut un échantillon de 30 cas (15 paires appariées) pour obtenir une puissance de 0,8 pouvant détecter une taille d'effet de 0,2 (avec un alpha de 0,05).

Plan à double saisie

Dans un plan à double saisie, chaque sujet ne passe le test qu'une seule fois, mais l'administration est enregistrée sur un support vidéo pour saisir les réponses du sujet et tous les éléments audio. Un certain nombre d'examineurs regardent indépendamment chaque vidéo pour saisir et coter les réponses, en utilisant soit le format papier-crayon soit le format électronique.

Ce plan est adapté pour les subtests pour lesquels le format électronique n'affecte pas le comportement du sujet, soit directement (par l'affichage ou le mode de réponse sur la tablette) soit indirectement (par les retours que l'examineur fait au sujet pendant la réalisation de chaque item). Les administrations sont enregistrées sur un support vidéo du point de vue de l'examineur, sans pour autant montrer la saisie et la cotation des réponses par l'examineur. Un ensemble d'examineurs cote chaque enregistrement vidéo : la première moitié utilise le format papier-crayon tandis que l'autre moitié utilise le format digital. Si l'hypothèse selon laquelle le comportement de l'examineur n'est pas affecté par le format digital est vérifiée, nous pouvons en conclure que cela équivaut à une méthode test-retest dans laquelle les résultats du sujet sont identiques pour chaque mode de passation. Ce plan se concentre ainsi entièrement sur l'effet du format digital sur la capacité de l'examineur à saisir et à coter la performance d'un sujet. Il faut un nombre suffisant de sujets pour fournir un éventail de tâches cognitives différentes et chaque protocole doit être coté

par plusieurs examinateurs pour chaque mode de passation.

Mode de sélection des sujets

Les premières études d'équivalence sur Q-interactive ont utilisé des échantillons de sujets sains ayant des caractéristiques démographiques semblables à celles de la population générale. Les patients présentant des profils cliniques ont été exclus de ces études afin de se focaliser sur l'estimation de la présence et de la taille d'un possible effet du mode d'administration. Comme il est difficile de prévoir les effets de l'administration assistée par tablette sur les individus présentant une pathologie avérée, l'inclusion d'un échantillon arbitrairement déterminé d'individus présentant divers troubles aurait des effets inconnus sur les résultats et pourrait interférer avec l'objectif, à savoir si le format digital a un effet sur le comportement du sujet ou de l'examineur.

Les examinateurs participant aux études d'équivalence sur le mode de passation devaient maîtriser les procédures d'administration des tests standardisés et ont reçu une formation théorique et pratique suffisante dans les procédures d'administration digitale pour être en mesure de mener à bien la passation, sans avoir à trop se concentrer sur le format. D'après notre expérience, il faut beaucoup de pratique pour se familiariser complètement à un nouveau format d'administration.

Études d'équivalence de la WAIS-IV

À ce jour, deux études sur la WAIS-IV ont été menées.

Étude 1 : WAIS-IV

Méthode

Nous avons utilisé un plan à groupes équivalents avec répartition non aléatoire pour la première étude sur la WAIS-IV via Q-Interactive, réalisée en août-septembre 2011.

Sujets

L'échantillon de l'étude 1 via Q-interactive était composé d'individus âgés de 16 à 77 ans, qui avaient été recrutés par l'équipe Pearson chargée de la recherche sur le terrain ou par une entreprise d'études de marché. 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 graves (nous avons recruté et testé certains individus présentant des conditions cliniques bénignes, mais nous les avons exclus des analyses). Le plan d'échantillonnage exigeait une surreprésentation des individus les plus âgés et de ceux ayant au maximum le Baccalauréat. En effet, ces groupes étaient logiquement les plus susceptibles d'être affectés par le mode de passation sous format digital.

Nous avons réparti aléatoirement les individus ayant accepté de participer à l'une des deux sous-études (Étude 1A ou Étude 1b). Nous avons rémunéré tous les individus ayant participé.

Le tableau 1 indique les caractéristiques démographiques des deux sous-échantillons, qui sont semblables sauf pour le sexe (l'échantillon de l'Étude 1b ayant une proportion plus élevée d'hommes). Comme les sous-études sont analysées de façon indépendante (pas de comparaison entre elles), cette différence n'affecte pas les analyses.

L'analyse de puissance effectuée indique que, avec un nombre de 39 ou 40 individus passant la version digitale de chaque subtest et une corrélation multiple entre les tests de covariables et les variables démographiques avec les scores au test de convergence étant égale à 0,75, le plan à groupes équivalents avec répartition non aléatoire a une probabilité de 0,46-0,47 de trouver un effet statistiquement significatif (alpha de 0,05) si la taille de l'effet réel est de 0,2.

Les examinateurs étaient des psychologues cliniciens ou des psychologues scolaires qualifiés et expérimentés dans l'administration de la WAIS-IV ou du WISC-IV. Ils ont reçu deux jours de formation sur site pour administrer la WAIS-IV via Q-interactive et se sont entraînés à faire plusieurs passations avant le début de l'étude. Les essais se sont déroulés sur quatre sites aux États Unis : San Antonio, Chicago, San Francisco et Newark, NY (près de Rochester). À San Antonio, les essais ont été effectués dans le bureau de Pearson. Pour les autres sites, les entreprises d'études de marché qui ont recruté les candidats ont fourni les salles nécessaires aux essais. Toutes les administrations ont été enregistrées sur support vidéo (avec l'accord des intéressés) afin que l'exactitude des administrations digitales et papier-crayon puisse être évaluée au cas où un effet du mode d'administration serait trouvé. Les examinateurs qui ne faisaient pas partie du personnel

de Pearson ont été indemnisés pour leur participation.

Tableau 1 : Caractéristiques démographiques des échantillons de l'étude 1 (WAIS-IV)

Caractéristique démographique		Étude 1a	Étude 1b
<i>N</i>		39	40
Âge (en années)	Étendue	16-77	17-75
	Moyenne	46,1	45,1
	Écart-type	19,1	17,8
Sexe	Filles	20	14
	Garçons	19	26
Origine ethnique	Afro-américain	8	8
	Hispanique	13	17
	Blanc	17	15
	Autres	1	0
Niveau d'études	< 9 ans	2	2
	9-11 ans	11	7
	Bac	8	11
	Post-bac	8	12
	Licence	10	8
Région des États-Unis	Nord-Est	9	9
	Centre Nord	6	5
	Sud	21	24
	Ouest	3	3

Procédure

Dans cette étude d'équivalence non randomisée, les normes américaines établies pour la WAIS-IV pour une passation papier-crayon (2200 sujets) ont été utilisées. La moitié des subtests de la WAIS-IV ont été évalués au cours de chacune des deux sous-études indépendantes (1a et 1b). Dans chaque sous-étude, la moitié des subtests ont servi de tests de covariables administrés sous format papier-crayon. Ces subtests ont en effet été administrés sous format papier-crayon à la fois à l'échantillon témoin et au nouvel échantillon (Q-interactive). L'autre moitié des subtests (ceux qui ont été évalués pour l'effet du mode d'administration) ont été administrés sous format digital pour l'échantillon Q-interactive et sous format papier-crayon pour l'échantillon témoin. Nous avons formé les deux moitiés en essayant de maximiser la corrélation multiple de chaque subtest de l'une des moitiés (à savoir, un subtest analysé pour observer l'effet du mode d'administration) avec la série de subtests de l'autre moitié (c'est-à-dire les covariables administrés au format papier-crayon), de

manière à maximiser la puissance statistique.

Le tableau 2 montre la composition des subtests pour les deux moitiés. Les 15 subtests mesurent quatre domaines d'aptitude, avec 3 à 5 subtests par domaine. Pour l'étude, les subtests au sein de chaque domaine ont été divisés en deux moitiés aussi équitables que possible afin d'équilibrer la représentation des construits et des modes de passation entre les deux moitiés. Comme il était important de conserver l'ordre classique d'administration des subtests, les subtests de chaque moitié ont été répartis à l'intérieur de la séquence en groupes de deux à quatre subtests consécutifs, pour réduire le nombre de changements de formats au cours de l'administration.

Tableau 2 : Ordres et modes d'administration des subtests pour l'étude 1 (WAIS–IV)

Subtests	Étude 1a	Étude 1b
Cubes	Papier-crayon	Digitale
Similitudes	Papier-crayon	Digitale
Mémoire des chiffres	Papier-crayon	Digitale
Matrices	Papier-crayon	Digitale
Vocabulaire	Papier-crayon	Digitale
Arithmétique	Papier-crayon	Digitale
Symboles	Papier-crayon	Digitale
Puzzles visuels	Papier-crayon	Digitale
Information	Papier-crayon	Digitale
Code	Papier-crayon	Digitale
Séquence lettres-chiffres	Papier-crayon	Digitale
Balances	Papier-crayon	Digitale
Compréhension	Papier-crayon	Digitale
Barrage	Papier-crayon	Digitale
Complètement d'images	Papier-crayon	Digitale

Chaque sujet a passé l'ensemble de la WAIS-IV dans l'ordre classique, avec la moitié des subtests administrés au format papier-crayon et l'autre moitié au format digitale (Q-interactive). L'examineur a changé de mode de passation entre les groupes.

Pour tous les subtests à l'exception des subtests de vitesse de traitement, nous avons retenu pour l'analyse le choix de cotation des items par les examinateurs (même si les équipes de Pearson ont corrigé toute erreur dans le calcul des notes brutes des subtests). L'interface examinateur de Q-interactive peut affecter la manière dont les examinateurs cotent les items ; leurs décisions constituent donc une partie importante de l'étude. D'autre part, comme les subtests de vitesse de traitement sont cotés après administration de manière identique pour les modes de passation papier-crayon et électronique, les équipes Pearson ont de nouveau coté les cahiers de passation pour ces

subtests pour éviter toute erreur de cotation.

La première étape de l'analyse consistait à utiliser l'échantillon témoin de la WAIS-IV pour calculer une équation de régression multiple pour chaque subtest. Nous avons estimé les notes étalonnées des subtests à partir des notes étalonnées des subtests de l'autre série et des variables démographiques (sexe, origine ethnique et niveau d'étude). Cette équation a ensuite été appliquée à chaque subtest administré au format digital dans les échantillons Q-interactive, pour générer une note estimée pour chaque administration électronique d'un subtest. Les résidus (différences entre les valeurs observées et les valeurs estimées) représentent l'effet numérique plus l'erreur. L'analyse de chaque subtest a utilisé un test t pour échantillon unique ou test de l'hypothèse nulle selon laquelle la moyenne des résidus est égale à zéro.

Résultats

Le tableau 3 fournit des informations permettant de déterminer si les équations de prédiction dérivées de l'échantillon témoin de la WAIS-IV sont bien adaptées à l'échantillon Q-interactive. Les corrélations de l'échantillon témoin indiquées dans ce tableau correspondent aux valeurs de validation croisée estimée de la corrélation multiple de l'ensemble des prédicteurs avec chaque test de convergence. Les corrélations étaient très semblables aux échantillons témoin et Q-interactive, qui confirment l'utilisation de notes estimées comme critères d'évaluation de notes obtenues lors d'administrations digitales.

Tableau 3 : applicabilité des équations de prédiction aux échantillons Q-interactive

Subtest	Corrélation de la valeur estimée avec la valeur observée	
	Échantillon Q-interactive	Échantillon témoin
Arithmétique	0,83	0,73
Cubes	0,68	0,72
Barrage	0,28	0,51
Code	0,68	0,57
Compréhension	0,67	0,77
Mémoire des chiffres	0,81	0,74
Balances	0,78	0,71
Information	0,77	0,72
Séquence lettres-chiffres	0,65	0,71
Matrices	0,82	0,67
Complètement d'images	0,57	0,55
Similitudes	0,82	0,76
Symboles	0,52	0,57
Puzzles visuels	0,77	0,71
Vocabulaire	0,84	0,81

Les résultats de l'étude 1 sont présentés dans le tableau 4. Certains subtests administrés au format digital présentaient des scores manquants ou inutilisables pour des raisons techniques ou de planification. Cela se reflète ainsi dans la taille de plusieurs échantillons lorsqu'elle est inférieure à 39 (Étude 1a) ou 40 (Étude 1b). Aucune des données manquantes n'était due à la difficulté d'un sujet à utiliser le format digital.

La moyenne des résidus est la différence moyenne entre les valeurs observées et les valeurs étalonnées estimées pour un subtest administré au format électronique. Une valeur positive signifie que les scores pour le mode de passation digital étaient plus élevés que prévu, tandis qu'une valeur négative indique une performance inférieure aux prévisions. La taille de l'effet correspond à la moyenne des résidus divisée par 3 (l'écart-type des scores étalonnés pour l'échantillon témoin).

Tableau 4 : Effets du mode d'administration pour l'étude 1 (WAIS-IV)

Subtest	Résidus			<i>t</i>	Taille d'effet
	<i>N</i>	Moyenne	Écart type		
Arithmétique	39	0,13	1,90	0,43	0,04
Cubes	36	-0,47	1,94	-1,45	-0,16
Barrage	32	0,48	2,65	1,02	0,16
Code	37	1,02	2,11	2,94**	0,34
Compréhension	39	-0,35	2,39	-0,91	-0,12
Mémoire des chiffres	36	-0,24	1,60	-0,9	-0,08
Balances	36	-0,06	2,35	-0,15	-0,02
Information	40	-0,83	1,98	-2,65*	-0,28
Séquence lettres-chiffres	36	-0,11	2,21	-0,30	-0,04
Matrices	40	0,30	1,82	1,04	0,10
Complètement d'images	40	-1,28	2,39	-3,39**	-0,43
Similitudes	36	-0,34	1,93	-1,06	-0,11
Symboles	37	0,54	2,44	1,35	0,18
Puzzles visuels	40	0,53	1,83	1,83	0,18
Vocabulaire	40	-0,14	1,65	-0,54	-0,05

Remarque : Taille d'effet = moyenne des résidus divisée par 3

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

Trois subtests ont montré des effets du mode d'administration statistiquement significatifs qui sont au-delà du critère du 0,2. Les notes obtenues lors d'administrations digitales des subtests Information et Complètement d'images se sont avérées plus faibles que prévu et les notes au subtest Code plus élevées que prévu. Les deux autres subtests de vitesse de traitement (Barrage et Symboles) n'ont montré aucun effet du mode d'administration significatif, contrairement au subtest Code.

Discussion

Le résultat selon lequel 12 des 15 subtests n'ont montré aucune taille d'effet du mode d'administration supérieure à 0,2 a confirmé l'objectif de cette étude, visant à démontrer l'équivalence entre le mode d'administration via Q-interactive et la procédure classique. Les trois subtests qui ont montré un effet du mode d'administration (ainsi que les deux autres subtests de vitesse de traitement) ont fait l'objet d'une enquête approfondie pour en découvrir les causes possibles. Cette investigation comprenait l'analyse des enregistrements vidéo des administrations papier-crayon et digitales ainsi que l'examen des cas présentant une variabilité exceptionnellement grande des notes aux subtests afin de détecter toute donnée incorrecte pouvant avoir influé sur les résultats.

L'évaluation du subtest Complètement d'images a suggéré deux causes possibles de l'effet numérique. Pour certains items, le dessin affiché sur la tablette du sujet était légèrement flou. Bien que ces effets aient été discrets et n'aient pas eu d'incidence sur la perception des caractéristiques essentielles des illustrations, la tâche proposée au sujet l'invite à rechercher des défauts et il est donc plausible que le sujet ait été perturbé par les zones floues du dessin et qu'il ait passé du temps à inspecter ces zones. Deuxièmement, certains examinateurs ont constaté que l'organisation des boutons de saisie de réponses prêtait à confusion sur la tablette de l'examineur. (Nous n'avons observé aucune erreur de saisie de réponse sur les enregistrements vidéo.)

En revanche, aucune explication apparente n'a été trouvée pour expliquer les effets du mode d'administration pour les subtests Information et Code. Pour ces deux subtests, la tablette du sujet n'est pas utilisée et l'interface examinateur est très simple. Les enregistrements vidéo réalisés lors des administrations papier-crayon et digitales de ces subtests n'ont montré aucun comportement du sujet ou de l'examineur ni aucune erreur de l'examineur qui pourraient expliquer ces résultats. Nous avons réalisé une analyse du fonctionnement différentiel des items du subtest Information pour déterminer si l'un des items était devenu plus difficile depuis l'étalonnage de la WAIS-IV il y a environ six ans, mais les résultats ont été négatifs. Pour Code et les autres subtests de vitesse de traitement, nous n'avons repéré qu'une seule différence entre les modes de passation : dans l'administration digitale, l'iPad affiche les consignes à l'examineur (au lieu de se trouver dans le cahier de passation) et contrôle le minutage.

Étude 2 : Suivi de la WAIS-IV

Méthode

Une deuxième étude de la WAIS-IV a été réalisée en novembre 2011 afin de vérifier les effets du mode d'administration trouvés dans l'étude 1 pour les subtests Complètement d'images, Information et Code (ainsi que pour les subtests de vitesse de traitement). Les dessins numériques ont été remplacés par des images plus nettes et l'écran de saisie de l'examineur a été remanié et simplifié, dans l'espoir de supprimer l'effet du mode d'administration pour ce subtest. Les autres subtests ont été administrés sans changement par rapport à l'étude 1. Comme aucune piste d'amélioration ou de correction n'est apparue pour ces subtests, nous avons conçu cette partie de l'étude pour confirmer ou infirmer les résultats initiaux.

Sujets

Nous avons recruté les sujets de la même manière que pour l'étude 1, hormis le fait que nous n'avons pas accepté de sujets présentant des profils cliniques. Nous avons en outre contrôlé plus étroitement les caractéristiques démographiques que durant l'étude 1 pour obtenir une meilleure adéquation avec la population générale. Le tableau 5 indique les répartitions par sexe, origine ethnique, niveau d'études et région dans l'échantillon de l'étude 2. L'échantillon d'étude de vitesse de traitement des 30 cas comprenait 15 paires démographiquement appariées et formait un sous-ensemble du plus large échantillon d'étude sur l'Information et le Complètement d'images. L'étude 2 a utilisé les mêmes sites d'essai et les mêmes examinateurs que l'étude 1.

Tableau 5 : Caractéristiques démographiques des échantillons de l'étude 2 (WAIS–IV)

Caractéristique démographique		Étude sur la Vitesse de Traitement	Étude sur les subtests Information & Complètement d'images
N		30	99
Âge (en années)	Étendue	16-67	16-82
	Moyenne	37,5	40,8
	Écart-type	14,0	17,0
Sexe	Filles	15	49
	Garçons	15	50
Origine ethnique	Afro-américain	8	15
	Hispanique	6	25
	Blanc	16	55
	Autres	0	4
Niveau d'études des parents	< 9 ans	0	1
	9-11 ans	0	4
	Bac	6	21
	Post-bac	13	41
	Licence	11	32
Région des États-Unis	Nord-Est		0
	Centre Nord		16
	Sud		83
	Ouest		0

Procédure

L'étude 2 était une combinaison de deux plans d'étude : le plan à échantillons équivalents avec répartition non-aléatoire a été répliqué pour les subtests Information et Complètement d'images et la méthode test-retest a été utilisée pour les subtests de vitesse de traitement. Comme cette dernière méthode nécessite moins de cas que la première, nous l'avons menée sur les 30 premiers cas et nous avons utilisé les cas restants uniquement pour l'analyse des subtests Information et Complètement d'images. La méthode test-retest a été jugée appropriée pour les subtests de vitesse de traitement car relativement peu d'apprentissage se produit d'une administration à l'autre. Les subtests utilisent des tâches qui sont faciles à exécuter et permettent d'évaluer le nombre de tâches de ce type que le sujet peut réaliser correctement en un temps donné. Si le sujet se souvient des associations entre symboles et chiffres pour le subtest Code, cela augmentera sa note lors de la deuxième passation, mais nous avons jugé que cela ne modifierait pas fondamentalement le processus cognitif nécessaire pour effectuer ce subtest.

Les 30 premiers sujets ont tous passé les subtests dans l'ordre suivant :

- Code, Barrage et Symboles soit au format papier-crayon soit au format digital ;
- Cubes et Vocabulaire au format papier-crayon (pour servir de covariables) ;
- Complètement d'images et Information au format digital ;
- Code, Barrage et Symboles dans le format non choisi auparavant.

Les autres sujets ont suivi ce même ordre, mais n'ont pas passé les subtests de vitesse de traitement. Dans la partie de l'étude sur la vitesse de traitement, dans chacun des 15 paires de sujets appariés démographiquement, un sujet a passé ces subtests tout d'abord au format papier-crayon puis au format digital tandis que l'autre sujet les a passés dans l'ordre inverse. Cela nous a permis d'utiliser pour l'analyse un test de Student pour paires appariées.

Nous avons utilisé les mêmes procédures que dans l'étude 1 pour l'analyse des subtests Complètement d'images et Information. Nous avons en effet utilisé des équations de régression basées sur l'échantillon témoin de la WAIS-IV pour prédire les notes obtenues avec une administration digitale de ces subtests et nous avons interprété les résidus entre les valeurs observées et les valeurs estimées comme des mesures de l'effet du mode d'administration.

Résultats

Les résultats de l'étude test-retest sur les subtests de vitesse de traitement sont indiqués dans le tableau 6. Nous avons obtenu chacun des effets du mode d'administration en calculant le score de variation moyenne entre le temps 1 et le temps 2 pour chaque groupe d'ordre de passation, puis en soustrayant la moyenne du groupe ayant d'abord passé les subtests au format digital à la moyenne du groupe les ayant d'abord passés au format papier-crayon et en divisant enfin le résultat par deux. En général, les résultats à l'étude test-retest sont semblables à ceux de l'étude 1, à savoir que les notes étaient plus élevées avec l'administration digitale.

Tableau 6 : Résultats à l'étude 2 pour les subtests de Vitesse de Traitement, Information et Complètement d'images de la WAIS-IV

Effet du mode d'administration

Subtest	Moyenne	Écart-type	t	Taille d'effet
Barrage	0,40	2,31	1,34	0,13
Codes	0,37	2,05	1,38	0,12
Symboles	0,80	3,31	1,87	0,27

Remarque : Un effet du mode d'administration positif indique des notes plus élevées pour l'administration électronique.
Taille d'effet = effet du mode d'administration divisé par 3

Peu de temps après la fin de l'étude 2, les développeurs de l'application Q-Interactive ont découvert que la manière dont le départ du chronométrage avait été programmé avait introduit un décalage d'environ 2 secondes.

Par conséquent, les sujets avaient passé environ 2 % de temps supplémentaire pour effectuer les tâches de vitesse de traitement. La réduction des notes brutes de 2 % a conduit à l'ajustement suivant de la taille d'effet du mode d'administration électronique : Barrage, 0,13 ; Code, 0,07 et Symboles, 0,13. Les développeurs de l'application ont depuis corrigé cette erreur de chronométrage via Q-interactive.

Comme indiqué dans le tableau 7, le résultat du subtest Information de l'étude 2 était presque identique à celui de l'étude 1 (taille d'effet de -0,29 et de -0,28, respectivement). Comme nous n'avons rien modifié dans l'administration du subtest Information, cela a donc confirmé les résultats initiaux. D'autre part, l'effet du mode d'administration pour le subtest Complètement d'images était bien inférieur par rapport à l'étude 1 (-0,17 vs. -0,43) et se situait à l'intérieur du seuil d'équivalence autorisé, indiquant que l'amélioration de la qualité des illustrations et / ou les changements apportés à l'écran de saisie de l'examineur avaient eu un effet positif.

Tableau 7 : Résultats à l'étude 2 pour les subtests Information et Complètement d'images de la WAIS-IV

Résidus					
Subtest	N	Moyenne	Écart-type	t	Taille d'effet
Information	99	-0,88	2,02	-4,33**	-0,29
Complètement d'images	99	-0,50	2,24	-2,22*	-0,17

Remarque : Taille d'effet = moyenne des résidus divisée par 3

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

Suite à l'étude 2, une analyse complémentaire du subtest Information a permis d'identifier la source de l'effet du mode d'administration électronique. Dans l'étude 1, 40 sujets ont passé le subtest Information au format papier-crayon (pour servir de covariable). Nous avons analysé leurs notes en utilisant la même méthode que celle appliquée pour l'administration électronique, à savoir en utilisant comme covariables d'autres subtests administrés au format papier-crayon (Vocabulaire, Matrices, Puzzles visuels et Mémoire des chiffres). La corrélation multiple de ces subtests à la note du subtest Information au format papier-crayon était de 0,74 pour l'échantillon témoin de la WAIS-IV et de 0,84 pour l'échantillon Q-interactive. La moyenne des résidus était - 0,51 (taille d'effet de -0,17), indiquant que, même lorsque le subtest Information est administré au format papier-crayon, ses notes ont tendance à être moins élevées que prévu. La raison de cet effet est inconnue ; cela permet toutefois d'expliquer plus de la moitié des effets du mode d'administration trouvés dans les études 1 et 2. La taille d'effet restante, égale à -0,12 (-0,29 moins -0,17), est relativement faible et se révèle similaire à celle des autres subtests de compréhension verbale de la WAIS-IV.

Conclusion

La figure 1 offre un résumé des tailles d'effet des études 1 et 2. Tous les subtests ont des tailles d'effet inférieures à 0,2, seuil initialement fixé comme critère d'équivalence. Cela confirme que les praticiens sont en mesure d'interpréter les notes de test obtenues à l'aide de la plate-forme de Q-interactive de la même manière que les résultats obtenus par une administration classique (papier-crayon). Il convient de garder à l'esprit que ces études ont utilisé des échantillons d'individus sains et que les effets potentiels de l'utilisation de l'interface digitale avec des patients présentant des profils cliniques particuliers ne sont pas encore connus.

Les subtests de compréhension verbale ont tendance à présenter de très faibles effets du mode d'administration, qui vont toujours dans le même sens (notes légèrement inférieures avec une administration digitale). De même, les subtests de vitesse de traitement présentent des effets du mode d'administration très légèrement positifs. Ces effets ne sont pas importants ni cliniquement significatifs, mais nous devons continuer à en rechercher la cause. L'identification de ces causes nous permettra de mieux comprendre comment le modèle d'interface numérique peut influencer sur le comportement.

Ces études nous ont fourni des informations utiles sur les procédures à suivre pour étudier les effets du mode de passation en général et du mode de passation digital en particulier. La réplication des résultats pour les subtests Information et de vitesse de traitement sur des échantillons indépendants et (pour les subtests de vitesse de traitement) en suivant un autre plan expérimental confirme la validité et la fiabilité de ces protocoles.

L'intérêt d'avoir des enregistrements vidéo des administrations a souvent été démontré au cours de cette étude. Sans un moyen de déterminer comment un examinateur avait effectivement administré le test ou comment la performance du sujet aurait dû être cotée, il aurait été bien plus difficile d'évaluer les diverses hypothèses pouvant expliquer les causes des effets du mode d'administration que nous avons observés. Les enregistrements permettent également de renseigner les développeurs sur la façon dont les sujets et les examinateurs utilisent réellement les interfaces numériques.

Nous avons également appris qu'il était important que les examinateurs s'exercent plusieurs fois à l'administration. Même si les examinateurs pensaient généralement avoir bien compris les procédures d'administration via Q-interactive à la fin de la deuxième journée de formation, ils se sont vite aperçus que seulement quelques administrations en conditions réelles n'étaient pas suffisantes pour maîtriser le nouveau système et qu'ils devaient s'exercer davantage.

L'expérience menée sur les subtests de vitesse de traitement illustre parfaitement les risques que l'on encourt lorsque l'on estime qu'un nouveau modèle d'interface numérique va toujours fonctionner comme prévu. Il s'agissait de la première expérimentation avec des tests cotés en fonction de la vitesse de traitement et elle a permis d'identifier une faille technique qui n'aurait pas été décelée sans la réalisation de cette étude.

Enfin, ces études nous ont permis de commencer à comprendre comment les fonctionnalités d'un modèle d'interface influent (ou n'influent pas) sur la performance des sujets et sur le mode de saisie et de cotation des réponses par les examinateurs. À mesure que nos connaissances se

développent, nous devrions pouvoir généraliser ces résultats à d'autres tests du même type et présentant des fonctionnalités similaires.

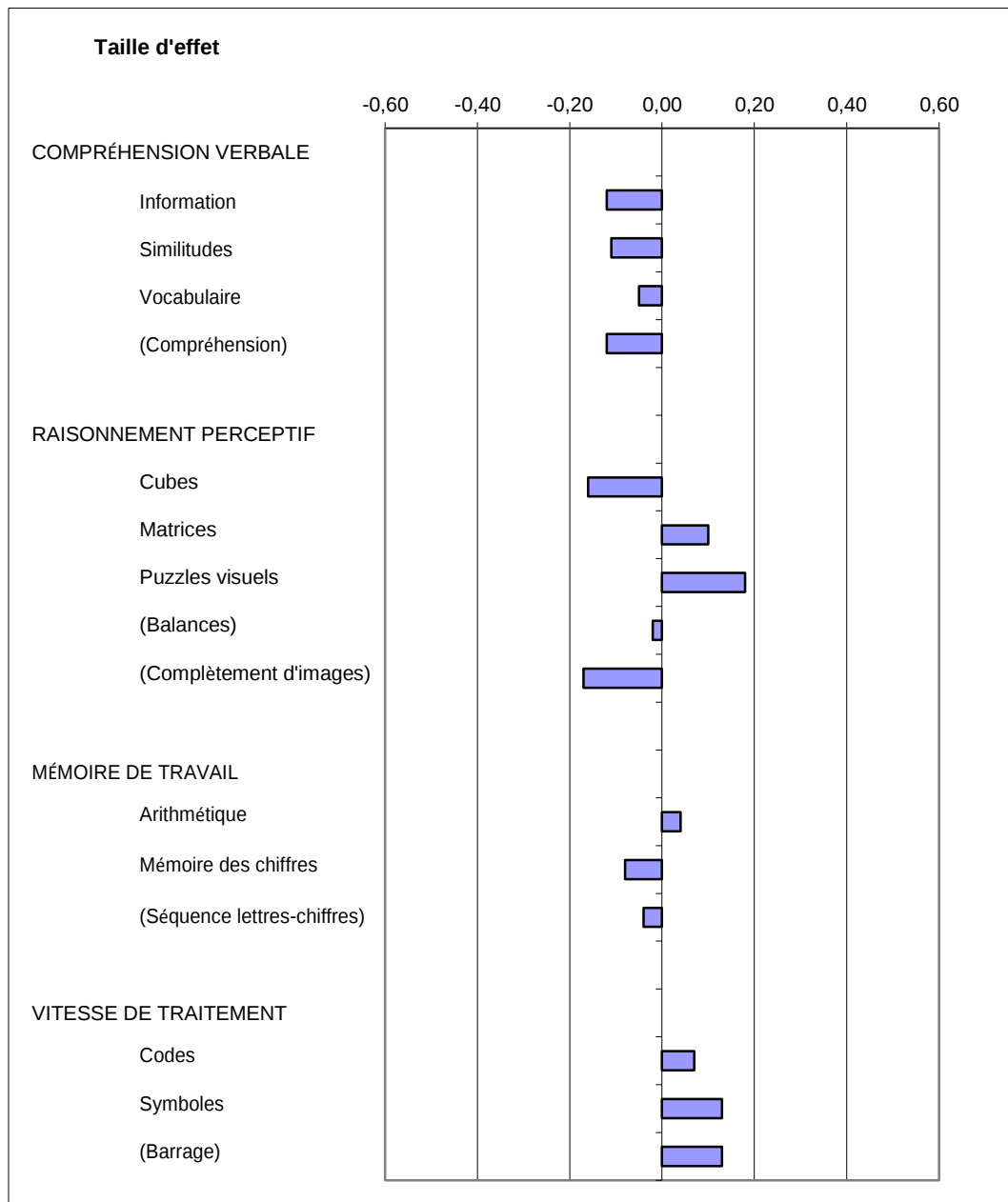


Figure 1 Résumé des résultats des études 1 et 2 de la WAIS-IV (subtests supplémentaires entre parenthèses)