



Q-interactive™

Equivalence des tâches cognitives de la MEM[®]–IV entre les modes d’administrations papier-crayon et via Q-interactive[®]

6^e rapport technique de Q-interactive

Mark H. Daniel, PhD

Cadre scientifique pour l'innovation et la recherche

Octobre 2013

Copyright © 2013 NCS Pearson, Inc. Tous droits réservés.

Q-interactive est une marque déposée aux États-Unis et/ou dans d'autres pays, de Pearson Education, Inc.

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 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 de 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.

Il s'agit de la sixième étude d'équivalence de Q-interactive. Dans cette étude, nous avons étudié l'équivalence des résultats entre l'administration standard et l'administration informatisée de l'*Échelle clinique de mémoire de Wechsler*®-4^e édition (MEM®-IV ; Wechsler, 2009a).

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. Pour les deux subtests du WISC-IV (Matrice 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 jugé 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 enfin 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 toutes les études d'équivalence, on suppose qu'une administration numérique (via Q-interactive) peut affecter les résultats au test pour un certain nombre de raisons, notamment :

Copyright © 2013 NCS Pearson, Inc. Tous droits réservés.

Q-interactive est une marque déposée aux États-Unis et/ou dans d'autres pays, de Pearson Education,

- 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 des cahiers de passation (par ex., le test de Traçage de pistes du Système d'évaluation des fonctions exécutives Delis-Kaplan D-KEFS) ont été utilisés dans le cas d'une administration via Q-interactive. Même si ces manipulables pourraient, à terme, être remplacés par des interfaces numériques interactives, le niveau d'adaptation nécessaire pourrait entraîner un manque d'équivalence des notes brutes. Il faudrait alors fournir davantage d'efforts de développement pour confirmer l'interprétation normative et fournir une preuve de fiabilité et de validité.
- L'interaction de l'examineur avec l'iPad, en particulier au cours de la saisie et de la cotation des réponses. La plupart des différences d'administration introduites dans la première version de Q-interactive se situent au niveau de l'interface examineur. 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 relève les temps de réponse, ou encore saisit et cote les réponses.
- 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 pour ne pas prendre d'avance sur l'examineur. Comme cela pouvait faire baisser ses notes, nous avons abandonné l'utilisation d'un clavier pour la saisie des réponses.

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 examineur obtiendrait en utilisant correctement le format papier-crayon. Le format digital 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 faudra alors résoudre ce problème par un ajustement des normes.

Il est impératif que les études d'équivalence intègrent une méthode de vérification de l'exactitude de l'administration, de l'enregistrement et de la cotation à la fois pour le mode de passation classique et digital. Ce n'est que dans ces conditions que les écarts de notes peuvent être attribués à un mode de passation ou à l'autre, ou bien à certaines caractéristiques de l'un des deux modes de passation. Dans chaque étude d'équivalence de Q-interactive, la totalité ou la plupart des administrations ont été enregistrées sur support vidéo pour établir la cotation *correcte* pour chaque item et subtest. Ces enregistrements ont également montré comment les

examineurs et les sujets ont interagi avec les matériaux de test dans chaque mode de passation.

Les études d'équivalence montrent généralement que les sujets âgés d'au moins 5 ans (correspondant à l'âge du plus jeune sujet testé), qui n'ont aucun diagnostic clinique et qui ne relèvent pas d'une éducation spécialisée, répondent de la même manière, que les stimuli soient présentés sur une tablette numérique plutôt que dans un cahier de passation ou que leurs réponses tactiles soient saisies à l'écran plutôt qu'observées par un examinateur. Nous n'avons repéré qu'une seule exception (pour les subtests Matrices et Complètement d'images) : pour les subtests impliquant le raisonnement conceptuel avec des stimuli visuels (ou une analyse visuelle détaillée de ces stimuli), les enfants réussissent mieux lorsque les stimuli sont présentés sur une tablette. La cause de cette différence reste encore inconnue. En outre, un ensemble de preuves démontre que, lorsque les examinateurs utilisent les modèles d'interfaces numériques qui ont été jusqu'ici décrites au lieu d'un formulaire d'observation papier, d'une administration manuelle et d'un chronomètre, ils obtiennent les mêmes résultats.

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. 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 plan 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 au format digital et l'autre au format papier-crayon. Le plan à groupes équivalents est décrit en détail dans les Rapports techniques 1 et 2 de Q-interactive.

Nous avons utilisé un autre plan, *la méthode test-retest*, pour l'étude de suivi sur les subtests de vitesse de traitement de la WAIS-IV (Rapport technique n°1, traduction en français disponible) ainsi que pour l'étude des subtests sur la NEPSY®-II et la CMS (Rapport technique n°4 uniquement disponible en langue anglaise). 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.

Le troisième type de protocole, plan à administration unique appelé *double saisie*, a été utilisé pour les études sur le CVLT®-II, le D-KEFS® et le WIAT®-III (rapports techniques de Q-interactive n°3 et 5, uniquement disponibles en langue anglaise). Il convient d'utiliser cette méthode lorsque le format digital 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 digital. Une comparaison des notes moyennes pour les deux modes de passation permet d'indiquer si le format influe sur les processus de saisie et de cotation des réponses.

La présente étude de la MEM[®]-IV utilise l'ensemble de ces trois plans, parce que les différents subtests se prêtent à l'usage de diverses méthodes. Les méthodes test-retest et de double saisie nécessitent moins de cas que le plan à groupes équivalents, si bien que nous avons de préférence utilisé ces deux premières méthodes. Nous avons utilisé le plan à groupes équivalents lorsque les caractéristiques du subtest rendaient inutilisables les méthodes plus efficaces.

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 correspond à un peu plus d'un demi-point par rapport à la métrique de note étalonnée utilisée pour les subtests de la MEM-IV (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 digital. 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 digital 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 plan à groupes équivalents, il est important que les échantillons comparés soient représentatifs de la population générale (sexe, origine ethnique, niveau socio-économique [niveau d'étude]) et que les deux groupes soient démographiquement semblables. Pour les méthodes test-retest et de 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 subtests (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 faut avoir réalisé au moins trois administrations d'entraînement pour bien se familiariser et maîtriser un nouveau format d'administration.

Étude d'équivalence de la MEM-IV

Mesures

La MEM-IV évalue de manière approfondie, à partir d'une passation individuelle en face à face, différents aspects de la mémoire, notamment le rappel immédiat, le rappel différé et la reconnaissance de stimuli visuels et verbaux. Nous avons estimé que tous les subtests de la MEM-IV devaient faire l'objet d'une évaluation d'équivalence, car les interfaces de Q-interactive présentent des caractéristiques qui pourraient influencer sur la performance du sujet ou sur la façon dont l'examineur saisit et cote les réponses. De plus, ces fonctionnalités n'ont pas encore été étudiées et nous ne savons pas si elles entraînent des effets du mode d'administration.

Les subtests de la MEM-IV sont présentés dans le tableau 1. Nous avons utilisé la tablette numérique du sujet pour présenter tous les stimuli visuels. Les sujets ont répondu directement sur l'écran au subtest Mémoire de symboles, mais ont répondu sur des cahiers de passation aux subtests Dessin d'horloge et Reproduction visuelle et ont utilisé la grille mémoire et des cartes pour les subtests Dessins et Addition spatiale.

Tableau 1 Description des subtests de la MEM-IV

Subtest Description d'étude		Type de plan
Subtest	Description d'étude	Type de plan
Examen cognitif rapide du sujet (ECRS™)	Ensemble de tâches simples : orientation, contrôle mental, dessin d'horloge, rappel incident, contrôle mental, inhibition et production verbale	Test-retest et Double saisie
Mémoire logique	On fait entendre au sujet deux histoires courtes et il doit les raconter à nouveau de mémoire.	Double saisie
Mots couplés	On fait entendre au sujet un ensemble de mots couplés, puis il doit restituer le deuxième mot en entendant le premier.	Groupes équivalents
Dessins	On présente au sujet pendant 10 secondes une grille sur laquelle figurent des symboles disposés dans certaines cases, puis il doit reproduire la même disposition.	Test-retest
Reproduction visuelle	On présente au sujet une figure abstraite pendant 10 secondes, puis il doit la dessiner sur papier de mémoire.	Groupes équivalents
Addition spatiale	On présente successivement au sujet deux grilles sur lesquelles figurent des cercles colorés, puis il doit placer les cercles sur une grille vierge en suivant des règles d'addition et de soustraction.	Test-retest
Mémoire de symboles	On présente au sujet une suite de symboles abstraits, puis il doit reproduire la suite en pointant les symboles dans l'ordre.	Test-retest

Comme indiqué dans le tableau 1, les différents subtests se prêtaient à différents plans d'étude. Le plan à double saisie est adapté lorsque le sujet n'interagit pas directement avec Q-interactive (c'est-à-dire lorsque la tablette du sujet n'est pas utilisée) et qu'il y a peu voire aucune interaction entre l'examineur et le sujet au cours de l'administration d'un item. Le subtest Mémoire logique et la plupart des tâches de l'*Examen cognitif rapide du sujet* (ECRS) répondaient à ces critères. La méthode test-retest est adaptée lorsque le sujet est peu susceptible de se rappeler en détail la réponse à un item et que l'apprentissage d'une stratégie pour un test ne constitue pas un élément

déterminant de la performance. Nous avons utilisé la méthode test-retest pour les subtests Dessins, Addition spatiale et Mémoire de symboles ainsi que pour la partie Inhibition de l'ECRS. Nous avons étudié les deux subtests restants, Reproduction visuelle et Mots couplés, à l'aide d'un plan à groupes équivalents. Pour les subtests contenant à la fois un essai immédiat (partie 1) et un essai différé (partie 2), nous avons uniquement évalué l'essai immédiat. Il n'est pas nécessaire d'étudier les essais différés car les interfaces examinateur et sujet ne présentent aucune caractéristique complexe ne figurant pas déjà dans les essais immédiats.

Par souci d'efficacité opérationnelle, nous avons divisé cette recherche en deux études : L'étude A combine les méthodes test-retest et à double saisie, tandis que l'étude B fait appel au plan à groupes équivalents.

Étude A : Test-retest et Double saisie

Méthode

Sujets

L'échantillon de sujets était composé de 30 individus répartis en 15 paires appariées démographiquement. Nous avons utilisé tous ces sujets dans les analyses test-retest et dix d'entre eux également pour les analyses à double saisie.

Les équipes PEARSON chargées de la recherche sur le terrain ont recruté les sujets et les ont 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. Aucun des sujets n'avait de diagnostic clinique ni ne relevait d'une éducation spécialisée. Le plan d'échantillonnage exigeait un nombre presque égal d'hommes et de femmes, une distribution équitable par âge, une diversité ethnique et une diversité de niveau socio-économique (niveau d'étude du sujet ou des parents du sujet, pour les sujets âgés de 16 à 24 ans).

Le tableau 2 dresse les caractéristiques de l'échantillon. Tant les échantillons du plan à double saisie que test-retest présentaient une bonne répartition des caractéristiques démographiques. Dans l'échantillon test-retest, les sujets au sein de chaque paire étaient rigoureusement appariés par groupe d'âge, par sexe et par origine ethnique. Le niveau d'étude était rigoureusement apparié dans neuf des paires, mais différait d'un niveau dans quatre autres paires et de deux niveaux dans une autre paire.

Huit examinateurs ont mené les administrations test-retest. Toutes ces personnes étaient qualifiées et expérimentées dans l'administration de tests psychologiques. Les examinateurs ont reçu une formation sur place pour administrer et coter les subtests avec un matériel papier et ont effectué plusieurs administrations d'entraînement ainsi qu'une administration « de certification ». Ils ont également reçu une formation pour administrer et coter les subtests via Q-interactive. Les examinateurs qui ne faisaient pas partie des équipes PEARSON ont été indemnisés pour leur participation.

Les dix administrations réalisées dans la partie à double saisie de l'étude ont été effectuées par quatre des examinateurs. Dix personnes (dont deux de ces examinateurs) ont également coté les

enregistrements vidéo de ces administrations. Ces évaluateurs étaient expérimentés et avait reçu une formation spécifique à l'étude, notamment pour la cotation des entraînements, comparable à celle des examinateurs.

Tableau 2 Méthodes test-retest et à double saisie de l'étude A : caractéristiques démographiques de l'échantillon

Caractéristique démographique		Test-retest		
		Papier-crayon / Digital	Digital / Papier-crayon	Double saisie
Total		15	15	10
Âge	16-24	4	4	3
	25-54	8	8	4
	55-69	3	3	3
Sexe	Femmes	9	9	6
	Hommes	6	6	4
Origine ethnique	Afro-américain	1	1	1
	Hispanique	6	6	3
	Blanc	7	7	6
	Autres	1	1	0
Niveau d'études^a	< 12 ans	3	1	0
	Bac	3	5	3
	Post-bac	4	4	3
	Licence	4	5	4

^a Pour la méthode test-retest, le niveau d'étude d'un sujet appartenant au groupe d'ordre de passation papier-crayon / digital n'était pas disponible.

Procédure

Chaque sujet a passé les subtests dans l'ordre suivant :

- Format 1 (pour l'analyse test-retest)
 - Dessins
 - Addition spatiale
 - Partie Inhibition de l'ECRS
 - Mémoire de symboles
- Tâches de non-inhibition de l'ECRS (format papier-crayon, pour l'analyse à double saisie)
- Mémoire logique (format papier-crayon, pour l'analyse à double saisie)
- Format 2 (pour l'analyse test-retest)
 - Dessins
 - Addition spatiale
 - Partie Inhibition de l'ECRS
 - Mémoire de symboles

Les sujets n'étaient pas informés à l'avance qu'ils devraient passer certains des subtests une deuxième fois. Tous les sujets ont passé les tâches de non-inhibition de l'ECRS et le subtest Mémoire logique, qu'ils participent ou non à l'étude à double saisie, pour rendre l'intervalle de test-retest uniforme pour tous les sujets. Pour la moitié des sujets (un membre de chaque paire appariée), le format 1 correspondait au papier-crayon et le format 2 à Q-interactive. Pour l'autre moitié, nous avons inversé les modes de passation.

Les examinateurs ont effectué toutes les procédures habituelles de cotation et de saisie, y compris la cotation post-administration des réponses des sujets. Nous avons utilisé toutes ces notes dans les analyses test-retest et mais nous ne les avons pas incluses pour les analyses à double saisie. Dans tous les cas d'administration papier-crayon, les équipes PEARSON ont évalué la conversion des résultats des items aux notes brutes des subtests pour éliminer toute erreur de calcul. (Cette étape est effectuée automatiquement avec Q-interactive.) Comme le calcul des notes brutes des subtests est une procédure ancillaire qui se produit après l'administration, nous n'estimons qu'il soit lié au mode de passation.

Pour l'analyse à double saisie, nous avons enregistré sur vidéo les administrations de l'ECRS (non-inhibition) et du subtest Mémoire logique afin de saisir le point de vue de l'examineur de l'administration, montrant ainsi le sujet, le livret de stimuli et le cahier de passation, mais masquant le formulaire d'observation de l'examineur. Chaque administration enregistrée sur vidéo a été cotée 10 fois par 10 évaluateurs différents. La moitié des cotations de chaque administration ont été effectuées en utilisant le format Q-interactive et l'autre moitié en utilisant le format papier-crayon. Chaque évaluateur a coté les dix administrations, cinq dans chaque format. Pour éviter toute corrélation entre le mode de passation et l'évaluateur, nous avons utilisé une table de nombres aléatoires afin d'identifier tout évaluateur qui coterait chaque administration au format Q-interactive. Nous avons ensuite effectué quelques ajustements afin que chaque évaluateur réalise cinq cotations pour chaque format.

Chaque évaluateur a effectué ses cotations indépendamment dans une pièce isolée. L'évaluateur a regardé la vidéo de l'administration et saisi ses réponses sur le formulaire d'observation papier ou sur la tablette via Q-interactive. Nous avons encouragé les évaluateurs à utiliser toutes les méthodes ou toutes les fonctionnalités de Q-interactive (par exemple, la capture audio) dont ils pourraient se servir dans leur pratique clinique. Ils ont coté chaque subtest ou tâche en temps réel. Ils n'étaient pas autorisés à s'arrêter et à redémarrer la vidéo pendant un subtest ou une tâche. Ils ne pouvaient pas non plus regarder la vidéo une deuxième fois. Les évaluateurs étaient seulement autorisés à interrompre la lecture de la vidéo entre deux subtests ou tâches.

Comme dans les précédentes études d'équivalence sur Q-interactive, des enregistrements vidéo ont été réalisés pour les administrations test-retest et pour les cotations à double saisie. Ces enregistrements remplissaient deux objectifs : premièrement, en cas de non-équivalence, les chercheurs pouvaient en étudier les causes possibles en examinant le comportement des sujets, des examinateurs et des évaluateurs. Deuxièmement, les vidéos fournissaient des informations sur la façon dont les individus interagissaient avec les matériels numérique et papier, ce qui pourrait s'avérer utile pour la conception future de tests. Les vidéos concernant les cotations à double saisie ont été prises en se plaçant derrière l'évaluateur afin de voir l'écran que l'évaluateur observait ainsi que la tablette numérique ou le formulaire d'observation sur lequel l'évaluateur saisisait ses réponses et entrait ses cotations.

Analyse

Test-retest. L'analyse test-retest était centrée sur le « score de variation » de chaque sujet (c'est-à-dire l'évolution de la note entre la première et à la deuxième administration). Si nous n'observions aucun effet du mode de passation, le score moyen de variation devait être le même pour le groupe d'ordre de passation papier-crayon / digital que pour celui d'ordre de passation digital / papier-crayon (exception faite de l'erreur d'échantillonnage et de l'erreur de mesure). Si nous observions un effet du mode de passation, le score moyen de variation des deux groupes d'ordre de passation devait correspondre au double ou à la moitié de la taille de l'effet : dans un groupe, l'effet multipliait par deux le score moyen de variation tandis que, dans l'autre groupe, il le divisait par deux. Nous avons calculé l'effet du mode d'administration en soustrayant le score moyen de variation pour l'ordre de passation digital / papier-crayon au score moyen de variation pour l'ordre de passation papier-crayon / digital, puis en le divisant par deux. Une valeur positive indiquait que le format digital obtenait des notes plus élevées que le format papier-crayon. Comme dans l'analyse à double saisie, l'effet du mode d'administration était exprimé en unités de note au subtest et la taille d'effet correspondait à l'effet du mode d'administration exprimé en unités d'écart-type.

Double saisie. L'analyse a comparé les notes moyennes à partir des cotations papier-crayon et Q-interactive de la même série d'administrations (sujets). La première étape consistait à calculer la note moyenne du format papier-crayon et la note moyenne du format digital pour chaque administration, sur la base des cinq cotations de chaque administration utilisant les deux formats. Nous avons ensuite calculé un score moyen global pour chaque mode de passation en établissant la moyenne de ces moyennes intra-administration sur l'ensemble des administrations. Ce calcul a indiqué un poids équivalent entre les formats de cotation papier-crayon et digital.

La taille d'effet pour chaque note correspondait à la différence entre les moyennes du format papier-crayon et du format digital, divisé par l'écart-type des notes à ce subtest pour l'ensemble de la population. Pour les subtests Dessins, Addition spatiale et Mémoire de symboles, qui sont cotés à l'aide de notes étalonnées, l'écart-type était de 3. Pour chacune des sous-scores de la partie Inhibition, qui indique des « notes pondérées », l'écart-type correspondait à la moyenne des moyennes quadratiques des écarts-types des groupes inter-âge dans l'échantillon témoin de la MEM-IV, pondérée de manière à refléter la distribution par âge dans cette étude. Nous avons effectué un test z pour évaluer la signification statistique, en utilisant comme erreur standard la moyenne quadratique des écarts intra-administration et intra-format par rapport à la moyenne.

Indépendamment du plan d'étude choisi, une valeur positive de l'effet du mode d'administration et de la taille d'effet indiquait que le format digital obtenait des notes plus élevées que le format papier-crayon.

Résultats

Test-retest. L'une des administrations de la partie Inhibition de l'ECRS n'était pas utilisable (pour des raisons indépendantes du mode de passation), si bien que l'analyse de cette partie était basée sur 14 des 15 paires appariées de sujets. Le calcul post-administration des notes brutes au subtest Dessins par les équipes de recherche PEARSON a identifié et permis de corriger un nombre considérable d'erreurs de calcul qui avaient été commises par les examinateurs.

Le tableau 3 indique les notes étalonnées moyennes (ou notes pondérées pour la partie Inhibition) pour chaque format d'administration au sein de chaque groupe d'ordre de passation. Pour les subtests qui utilisent des notes étalonnées, le sous-groupe ayant d'abord passé les administrations via Q-interactive obtenait des moyennes au premier essai qui étaient très proches des valeurs de la population générale. Dans le sous-groupe ayant d'abord passé les subtests au format papier-crayon, les moyennes au premier essai étaient variables et les écarts-types étaient systématiquement inférieurs aux valeurs de la population générale. Dans les deux sous-groupes, l'effet de pratique pour le subtest Addition spatiale était inférieur à celui des subtests Dessins ou Mémoire de symboles.

Tableau 3 Méthodes test-retest de l'étude A : Scores moyens (des écarts-types)^a par ordre et mode de passation

Subtest	Paires	Digital / Papier-crayon				Papier-crayon / Digital			
		Essai 1		Essai 2		Essai 1		Essai 2	
Dessins	15	9,80	(3,17)	12,87	(2,75)	7,87	(2,33)	10,47	(2,92)
Addition spatiale	15	9,93	(3,63)	10,07	(3,31)	11,07	(2,58)	11,87	(2,70)
Mémoire de symboles	15	9,87	(2,03)	12,00	(2,10)	8,67	(2,13)	10,47	(1,88)
Inhibition (ECRS)									
Temps	14	3,29	(1,07)	3,71	(0,47)	3,60	(0,74)	4,00	(0,00)
Erreurs d'omission	14	3,71	(1,07)	4,00	(0,00)	3,47	(1,41)	3,71	(1,07)
Erreurs de commission	14	6,79	(2,89)	7,36	(0,84)	6,53	(2,61)	8,00	(0,00)

^a Notes étalonnées pour les subtests Dessins, Addition spatiale et Mémoire de symboles ; notes pondérées pour la partie Inhibition

Le tableau 4 indique les résultats de l'analyse des effets du mode d'administration. Aucun des effets n'était statistiquement significatif. Les tailles d'effet pour les trois subtests complets (Dessins, Addition spatiale et Mémoire de symboles) étaient faibles (allant de -0,06 à 0,11) et restaient bien entre -0,20 et +0,20, fourchette correspondant au critère d'équivalence. Deux des sous-tâches de la partie Inhibition de l'*Examen cognitif rapide du sujet* (ECRS) n'ont montré aucun effet du mode d'administration, mais la troisième (Erreurs de la Commission) avait une taille d'effet de 0,35, dépassant le critère d'équivalence. Nous avons examiné tous les formulaires d'observation et tous les enregistrements vidéo pour vérifier qu'il n'y avait pas d'erreurs d'administration ou d'autres anomalies affectant la partie Inhibition.

Tableau 4 Méthodes test-retest de l'étude A : effets du mode d'administration

Score de variation							
Subtest	Digital / Papier-crayon		Papier-crayon / Digital		Effet du mode de passation t		Taille d'effet
	Moyenne <i>type</i>	Écart- <i>type</i>	Moyenne <i>type</i>	Écart- <i>type</i>			
Dessins	3,07	1,87	2,60	2,87	-0,24	-0,48	-0,08
Addition spatiale	0,13	1,51	0,80	2,04	0,34	1,13	0,11
Mémoire de symboles	2,13	1,92	1,80	1,15	-0,17	-0,54	-0,06
Inhibition de l'ECRS							
Temps	0,43	0,76	0,43	0,76	0,00	0,00	0,00
Erreurs d'omission	0,29	1,07	0,29	1,90	0,00	0,00	0,00
Erreurs de commission	0,57	3,03	1,57	2,68	0,50	0,89	0,35

Double saisie Cinq cotations étaient disponibles dans chaque format pour chaque administration, sauf pour la cotation papier-crayon de la tâche Dessin d'horloge de l'ECRS où seulement quatre cotations étaient utilisables. Les notes moyennes pour chaque format et les effets du mode d'administration sont indiqués dans le tableau 5. Une tâche de l'ECRS (Dessin d'horloge) a montré une taille d'effet qui dépassait le critère de 0,20, même si ce résultat n'était pas statistiquement significatif. Pour cette tâche, les notes étaient inférieures lors de l'utilisation de l'interface de cotation numérique. Un examen minutieux des cotations n'a décelé aucune erreur systématique ni aucune autre explication pour cet effet. Toutefois, le format n'a joué qu'un rôle négligeable sur les notes totales de l'ECRS, avec une taille d'effet de -0,06.

Tableau 5 Plan à double saisie de l'étude A : Différences entre les notes^a obtenues en utilisant les formats de saisie papier-crayon et Q-interactive

Subtest	Cotation <u>papier-crayon</u>		Cotation <u>Q-interactive</u>		Effet du mode de passation	z	Écart-type (Norme)	Taille d'effet
	Moyenne	Écart-type	Moyenne	Écart-type				
Orientation	6,50	1,58	6,50	1,58	0,00	0,00	1,36	0,00
Estimation du temps	2,62	1,36	2,50	1,24	-0,12	-1,00	0,80	-0,15
Contrôle mental :	3,38	0,95	3,32	0,95	-0,06	-1,00	0,76	-0,08
Contrôle mental :	5,51	2,45	5,44	2,40	-0,07	-0,30	2,09	-0,03
Dessin d'horloge	3,88	0,21	3,70	0,53	-0,18	-1,87	0,56	-0,32
Rappel incident	7,60	1,26	7,44	1,37	-0,16	-1,00	0,97	-0,17
Production verbale	5,80	0,51	5,63	0,63	-0,17	-2,68**	0,92	-0,19
Total ECRS					-0,26		4,29	-0,06
Mémoire logique	10,48	2,73	10,30	2,71	-0,18	-0,91	(3)	-0,06

** $p < 0,01$

^a Notes pondérées pour l'ECRS ; notes étalonnées pour Mémoire logique

Remarque : Le total de l'ECRS correspond à la somme des notes pondérées pour toutes les tâches, y compris Inhibition. $N = 50$ cotations dans chaque format, sauf pour la cotation papier-crayon du subtest Dessin d'horloge, pour lequel $N = 49$. La moyenne et l'écart type se fondent sur la distribution des moyennes à l'intérieur de l'administration.

Copyright © 2013 NCS Pearson, Inc. Tous droits réservés.

Q-interactive est une marque déposée aux États-Unis et/ou dans d'autres pays, de Pearson Education,

Étude B : Groupes équivalents

Méthode

Sujets

Un total de 99 sujets a passé les subtests Reproduction visuelle et Mots couplés au format Q-interactive, ainsi que les tests de covariables au format papier-crayon. Le groupe témoin ayant passé l'administration papier-crayon se composait des 186 membres de l'échantillon témoin de la MEM-IV, qui avaient également passé le *Test d'apprentissage et de mémoire verbale* (CVLT-II). Pour l'échantillon Q-interactive, les exigences d'inclusion, les contraintes démographiques et les procédures de recrutement étaient similaires à celles de l'étude A. Aucun de ces sujets n'avait un diagnostic clinique ou ne relevait d'une éducation spécialisée, même si trois d'entre eux étaient présentés comme doués / talentueux.

Les caractéristiques démographiques de l'échantillon Q-interactive sont présentées dans le tableau 6.

Tableau 6 Étude B : caractéristiques démographiques de l'échantillon Q-interactive

Caractéristique démographique		Sujets
Total		99
Âge	16-24	24
	25-43	48
	55-69	27
Sexe	Femmes	54
	Hommes	45
Origine ethnique	Afro-américain	12
	Asiatique	3
	Hispanique	21
	Blanc	59
	Autres	4
Niveau d'études	< 12 ans	4
	Bac	27
	Post-bac	33
	Licence	35

Procédure

Tous les sujets de l'échantillon Q-interactive ont passé les subtests dans l'ordre suivant :

- Reproduction visuelle I de la MEM-IV (au format Q-interactive)
- Mots couplés I de la MEM-IV (au format Q-interactive)
- Dessins I de la MEM-IV (au format Q-interactive)
- Essais immédiats et à court terme du CVLT-II (au format papier-crayon)

Tous les examinateurs étaient qualifiés et expérimentés dans l'administration de tests psychologiques. Ils ont été formés pour administrer la MEM-IV via Q-interactive et se sont entraînés à faire des passations avant le début de l'étude. Les essais ont eu lieu aux Etats Unis : à San Antonio (Texas), à Détroit (Michigan) et à Fort Collins (Colorado). Presque toutes les administrations ont été enregistrées sur support vidéo, afin que l'exactitude des deux administrations (papier-crayon et digital) puisse être évaluée en cas d'effet du mode d'administration et pour fournir des informations sur la façon dont les sujets et les examinateurs interagissaient avec l'interface Q-interactive. Tous les sujets et les examinateurs qui ne faisaient pas partie de PEARSON ont été indemnisés pour leur participation.

Les examinateurs ont effectué toutes les procédures habituelles de cotation et de saisie qui sont requises pour l'administration d'items. En outre, ils ont utilisé l'interface examinateur Q-interactive pour coter les réponses du subtest Reproduction visuelle après l'administration. Les notes totales aux tests administrés au format papier-crayon (Dessins et CVLT-II) ont été calculées par les équipes de recherche PEARSON qui a utilisé les réponses et les notes saisies dans le formulaire d'observation par les examinateurs.

Analyse

Nous avons utilisé les 186 cas de l'échantillon témoin de la MEM-IV pour développer des équations de régression multiple dans lesquelles les prédicteurs indépendants étaient la note étalonnée au subtest Dessins ou modèles subtest échelle, plusieurs notes de la CVLT-II (la note au test *t* de l'Essai A en rappel libre, la note au test *z* de l'Essai B en rappel libre et les notes aux tests *z* en rappel indicé et en rappel libre à court terme) ainsi que les caractéristiques démographiques (âge, sexe, origine ethnique et niveau d'études) et la variable dépendante était la note étalonnée soit au subtest Reproduction visuelle soit au subtest Mots couplés. Nous avons ensuite comparé les notes étalonnées observées aux subtests Reproduction visuelle et Mots couplés aux valeurs estimées à partir de ces équations de régression. Les résidus (les valeurs observées moins les valeurs estimées) représentent l'effet digital 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.

La taille d'effet pour chaque note correspond à la différence entre les moyennes du format papier-crayon et du format digital, divisée par l'écart-type des notes de référence pour la population générale (par exemple, 0,3 pour une note étalonnée de la MEM-IV). Nous avons effectué un test *t* pour évaluer la signification statistique de l'hypothèse selon laquelle la différence entre les moyennes n'est pas égale à 0. Une valeur positive indique que le format digital obtient des notes plus élevées que le format papier-crayon. L'effet du mode d'administration est exprimé en unités de note étalonnée et la taille d'effet correspond à l'effet du mode d'administration exprimé en unités d'écart-type.

Copyright © 2013 NCS Pearson, Inc. Tous droits réservés.

Q-interactive est une marque déposée aux États-Unis et/ou dans d'autres pays, de Pearson Education,

Résultats

Quatre cas du subtest Reproduction visuelle n'ont pu être utilisés en raison d'erreurs d'administration qui n'étaient pas liées au format Q-interactive. Le tableau 7 présente les statistiques descriptives des quatre subtests dans l'échantillon Q-interactive. Les moyennes et les écarts-types des notes étalonnées aux subtests de la MEM-IV étaient proches de la moyenne de la population générale (respectivement 10 et 3) et les notes du CVLT-II étaient également proches de leurs valeurs correspondantes dans la population, indiquant que l'échantillon est correctement représentatif de la population générale.

Les corrélations multiples des notes aux tests de covariables et des données démographiques avec les notes étalonnées aux subtests Reproduction visuelle et Mots couplés étaient respectivement de 0,58 et de 0,61 pour l'échantillon témoin de la MEM-IV et de 0,46 et de 0,63 pour l'échantillon Q-interactive.

La magnitude et la signification statistique des effets du mode d'administration sont indiquées dans le tableau 7. Les tailles d'effet pour les subtests Reproduction visuelle (-0,13) et Mots couplés (-0,03) administrés via Q-interactive restaient bien dans la fourchette de la norme d'équivalence (entre -0,20 et +0,20) et n'étaient pas statistiquement significatives.

Tableau 7 Étude B : Statistiques descriptives et tailles d'effet

Test / Subtest	Résidus ^a					t	Taille d'effet
	N	Moyenne	Écart-type	Moyenne	Écart-type		
<i>Tests de covariables administrés au</i>							
Dessins (note étalonnée) du CVLT-II ^b	99	9,3	2,6				
Liste A en rappel libre (t)	99	55,9	11,3				
Liste B en rappel libre (z)	99	-0,2	1,0				
Rappel libre à court terme (z)	99	0,3	1,2				
Rappel indicé à court terme (z)	99	0,1	1,0				
<i>Subtests administrés via Q-interactive :</i>							
Reproduction visuelle (note étalonnée)	94	9,9	2,9	-0,40	2,56	-1,55	-0,13
Mots couplés (note étalonnée)	99	10,1	2,7	-0,09	2,07	-0,45	-0,03

^avaleur observée moins valeur estimée

^bscore moyen au test $t = 50$, écart-type = 10 ; score moyen au test $z = 0$, écart-type = 1

Discussion

Aucun des subtests de la MEM-IV n'a montré un effet du mode de passation via Q-interactive dont l'effet de taille était de 0,20, indiquant qu'une administration via Q-interactive produit des résultats équivalents à une administration papier-crayon. Deux tâches optionnelles appartenant à l'*Examen cognitif rapide du sujet* (ECRS), à savoir Dessin d'horloge et Erreurs de commission dans la partie Inhibition, ont montré respectivement des tailles d'effet de -0,32 et 0,35. Toutefois, ces effets n'étaient pas statistiquement significatifs et l'ensemble des notes à l'ECRS avait une taille d'effet négligeable de -0,06. Ces résultats viennent enrichir le corpus de recherches indiquant que, lorsque les tests initialement conçus pour une administration papier-crayon sont scrupuleusement adaptés pour une administration via la plate-forme Q-interactive, les résultats sont comparables, quel que soit le mode de passation utilisé.

Références bibliographiques

Cohen, M. (1997). *Children's memory scale*™. 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.

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 memory scale*®—fourth edition. Bloomington, MN: Pearson.

Wechsler, D. (2009b). *Wechsler individual achievement test*®—third edition. Bloomington, MN: Pearson.