



Q-interactive™

# Équivalence des tâches langagières de la CELF 5 entre les modes d'administration papier- crayon et via Q-interactive®™

## **7<sup>e</sup> rapport technique de Q-interactive**

**Mark H. Daniel, PhD,**

**Dustin Wahlstrom, PhD,**

**Xuechun Zhou, PhD,**

**Juin 2014**

## Introduction

Q-interactive<sup>®</sup>, application numérique de Pearson pour l'administration de tests en face à face via iPad, est conçue pour rendre l'évaluation plus pratique et plus précise, pour fournir aux professionnels 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 iPad synchronisés communiquant 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 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 plateforme 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 fidélité 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 septième étude d'équivalence de Q-interactive menée aux Etats-Unis. Dans cette étude, nous avons étudié l'équivalence des résultats entre l'administration standard et l'administration digitale de la *Batterie d'évaluation des fonctions langagières et de communication*<sup>®</sup>-cinquième édition (CELF<sup>®</sup> 5 ; Wiig, Semel & Secord, 2019).

Dans les deux premières études d'équivalence, les quinze subtests de l'*Échelle d'intelligence de Wechsler pour adultes*<sup>®</sup>-quatrième édition (WAIS<sup>®</sup>-IV ; Wechsler, 2008) ainsi que treize des quinze subtests de l'*Échelle d'intelligence de Wechsler pour enfants*<sup>®</sup>-quatrième édition (WISC<sup>®</sup>-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*<sup>™</sup> (D-KEFS<sup>™</sup> ; Delis, Kaplan & Kramer, 2001) ainsi que les essais de rappel libre du *Test d'apprentissage et de mémoire verbale*<sup>®</sup>-deuxième édition (CVLT<sup>®</sup>-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*<sup>®</sup>, deuxième édition (NEPSY<sup>®</sup>-II ; Korkman, Kirk, & Kemp, 2007) et deux subtests de l'*Échelle de mémoire pour enfants*<sup>™</sup> (CMS<sup>™</sup> ; 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*<sup>®</sup>-troisième édition (WIAT<sup>®</sup>-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.

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, du matériel physique nécessaire à la passation (par exemple la boîte des cubes de l'*Échelle d'intelligence de Wechsler pour enfants*®-cinquième édition WISC-V) et des cahiers de réponse (par exemple pour le subtest Barrage de l'*Échelle d'intelligence de Wechsler pour adultes*®-quatrième édition WAIS-IV) 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 fidélité de la version numérique.
- *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 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 chronomètre les performances, ou saisit et cote les réponses.
- *La précision avec laquelle l'application Q-interactive saisit et cote les réponses tactiles du sujet.* Avec la CELF 5, Q-interactive introduit la cotation automatique. 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. La fiabilité de cette technologie devait être évaluée, en particulier dans les cas de subtests où la réponse correspond à une séquence de réponses tactiles, ou bien lorsque le patient souhaite s'auto-corriger.
- *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 avec 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. Comme cela pouvait avoir pour conséquence de diminuer leurs scores, l'utilisation d'un clavier pour saisir les réponses verbales a été abandonnée.

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 décider comment 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 grande majorité des administrations ont été enregistrées sur vidéo, montrant ainsi les interactions entre l'examineur et le sujet par l'intermédiaire de leurs iPad. Ces vidéos permettent ainsi de vérifier que l'administration, l'enregistrement et la cotation sont correctement réalisés dans les deux formats (papier-crayon et numérique). 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.

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 de stimuli ou que leurs réponses tactiles soient saisies à l'écran plutôt qu'observées par un examinateur. 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 que lors d'une administration papier-crayon.

## Conception des études d'équivalence

Nous avons utilisé plusieurs plans expérimentaux pour les études d'équivalence sur Q-interactive.

L'étude sur la WAIS-IV a reposé sur un protocole à *groupes équivalents*, avec une répartition aléatoire ou non aléatoire des sujets dans les groupes. 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). Cette solution permet que l'interaction d'un sujet avec la tâche ne soit pas 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. 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 le rapport technique n°6 (MEM-IV).

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 de 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 la MEM-IV (Rapport technique n°6) et dans la présente étude de la CELF 5.

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 standard (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 tout venant afin de maintenir l'accent sur l'estimation de la présence et de la taille d'effets avec le format numérique. 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 sur des sujets présentant des conditions cliniques particulières, la recherche initiale doit se concentrer principalement sur la question de savoir si le format numérique influe sur les résultats obtenus par des sujets tout venant.

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, niveau d'études des parents) 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). Ils ont également reçu une formation théorique et pratique suffisante dans les procédures d'administration et de cotation digitales 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.

A noter que d'après notre expérience, les professionnels ont besoin de pratiquer d'un certain temps pour se familiariser complètement à un nouveau mode d'administration.

# Étude d'équivalence de la CELF 5

## Méthode

### Mesures

La CELF 5 est un instrument complet utilisé pour évaluer les fonctions langagières et de communication de sujets âgés de 5 à 18 ans 11 mois. L'équipe de recherche R&D américaine a identifié quatre épreuves de la CELF 5 comme nécessitant d'être étudiées parce que leurs interfaces Q-interactive sur l'iPad du sujet ou de l'examinateur présentent des caractéristiques :

- Qui pourraient affecter la capacité de l'examinateur à administrer, saisir la réponse avec précision, ou qui pourraient avoir une incidence sur l'exactitude de la cotation automatique
- Et qui diffèrent des différents subtests étudiés dans de précédentes études et dont l'équivalence entre les supports papier et digital a déjà été démontrée

Ces quatre épreuves sont :

- *Identification de concepts* (5 – 8 ans) : le sujet entend une instruction orale portant sur des concepts de base et répond en touchant une ou plusieurs images sur la tablette, parfois dans un ordre précis.
- *Répétition de phrases* (5 – 21 ans) : le sujet entend une phrase et la répète ; l'examinateur transcrit ce que le sujet dit en modifiant la phrase imprimée.
- *Compréhension et exécution de consignes* (5 – 21 ans) : le sujet entend des directives orales de plus en plus longues et complexes, et réagit en touchant une ou plusieurs images sur la tablette, dans un ordre particulier.
- *Elaboration de phrases* (5 – 21 ans) : le sujet doit construire et prononcer une phrase à partir de mots imposés, tout en respectant les contraintes contextuelles imposées par les images.

Les épreuves *Répétition de phrases* et *Elaboration de phrases* sont incluses dans la présente étude parce qu'elles exigent de l'examinateur qu'il transcrive la réponse orale du sujet de façon précise et complète, afin qu'elle puisse être cotée plus tard si nécessaire. Les épreuves *Identification de concepts* et *Compréhension et exécution de consignes* sont choisies parce que certaines de leurs questions demandent une association ou une séquence de réponses tactiles, ce qui représente un défi pour le système de cotation automatique de Q-interactive.

### Sujets

L'échantillon comprenait 20 paires de sujets appariés sur le plan démographique, dont 14 faisaient partie de la tranche d'âge [5 – 8] à laquelle l'épreuve *Identification de concepts* est administrée. 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 du 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 (la passation est en ordre contrebalancée c'est-à-dire qu'elle est proposée d'abord sur papier-crayon puis sur Q-interactive ou bien l'inverse, mais tous les sujets passent le test sur les deux supports). Après la fin de la collecte des données, l'équipe de recherche a sélectionné en vue de l'analyse des paires de sujets (un membre pour chaque mode de passation) appariés par tranche d'âge, par sexe et par niveau d'études des parents.

Les sept examinateurs étaient qualifiés et expérimentés dans l'administration de tests de langage aux enfants et aux adultes. Les examinateurs ont reçu une formation sur place sur l'administration des épreuves de la CELF 5, au format papier-crayon ou au format digital sur Q-interactive. Ils ont mené plusieurs administrations d'entraînement ainsi qu'une administration d'essai qui a déterminé leur capacité à participer à l'étude. Les examinateurs n'étaient pas employés chez Pearson et ont été rémunérés pour leur participation.

## Procédure

Sur les quatre épreuves de la CELF 5, les processus cognitifs utilisés lors d'une deuxième administration ont été jugés peu susceptibles d'être significativement affectés par le fait que le sujet ait passé l'épreuve peu de temps auparavant. Il est difficile de se rappeler le contenu précis des différents items et ces tâches ne se prêtent pas à des stratégies de résolution de problèmes. C'est pour ces raisons que nous avons choisi d'utiliser la méthode du test-retest (double présentation, en ordre contrebalancé).

Les deux épreuves *Répétition de phrases* et *Élaboration de phrases* auraient pu être étudiées avec la méthode de la double-saisie, mais comme seules quatre épreuves sont évaluées dans cette étude, nous avons jugé plus efficace d'utiliser la même méthode pour toutes les épreuves, soit la méthode du test-retest.

La formation et les administrations ont eu lieu entre février et avril 2014. Les épreuves ont été administrées dans l'ordre suivant :

- *Identification de concepts* (5 – 8 ans seulement)
- *Répétition de phrases*
- *Compréhension et exécution de consignes*
- *Élaboration de phrases*

Chaque sujet a passé les trois ou quatre épreuves, adaptées à son âge, dans un format puis dans l'autre, le tout au cours de la même session d'administration. Il n'a pas été dit aux sujets qu'ils passeraient le test une deuxième fois. Au cours des administrations au format papier-crayon, les examinateurs ont transcrits les réponses du sujet de manière habituelle sur le cahier de passation puis ont coté chacune d'elle. L'équipe de recherche a vérifié l'utilisation appropriée des règles d'administration (règles de départ, règles d'arrêt) mais n'a pas revérifié la cotation des items.



Comme dans les précédentes études d'équivalence Q-interactive, les administrations ont été filmées. Ces enregistrements ont servi deux objectifs. Tout d'abord, en cas de constatation de non-équivalence, ils permettent aux chercheurs d'étudier les causes possibles en examinant le comportement des examinateurs et des sujets. Par ailleurs, ils permettent de fournir de l'information sur la façon dont les examinateurs et les sujets interagissent avec le matériel numérique et papier, ce qui peut être utile pour la conception de futurs tests. Les vidéos ont été tournées d'en haut et sur le côté de l'examineur et du candidat et montraient les deux tablettes iPad.

L'analyse d'une étude d'équivalence avec la méthode du test-retest se concentre sur la différence de score pour chaque sujet (c'est-à-dire l'écart entre les scores de la première à la seconde administration). S'il n'y a pas d'effet du format, la différence de score attendue sera la même pour les sujets, que la séquence de passation soit papier-numérique ou numérique-papier.

S'il y a un effet de format, on s'attend à ce que la différence de score dans les deux groupes de séquences soit de deux fois la taille de l'effet, car dans un groupe de séquences, l'effet augmentera le score moyen de différence et dans l'autre groupe de séquences, l'effet de format le réduira.

L'effet du format est calculé en soustrayant la note moyenne de changement dans la séquence numérique-papier de la note moyenne de changement dans la séquence papier-numérique, le tout divisé par 2. Une valeur positive indique que le format numérique donne des notes plus élevées que le format papier. L'effet du format est exprimé en unités de note standard. La taille de l'effet exprime l'effet de format en unités d'écart type (3 dans le cas des notes standard).

La signification statistique de l'effet du format est calculée à l'aide d'un test t de paires d'échantillons : la différence de score d'un sujet dans le groupe numérique-papier est soustrait de la différence de score du sujet correspondant dans le groupe papier-numérique, et la moyenne et l'écart type de ces différences génèrent le score t. L'utilisation de paires de sujets appariés sur le plan démographique dans les deux types de séquences produit une puissance statistique élevée avec des échantillons de petite taille. En supposant une corrélation de retest de 0,8, un échantillon de 15 paires appariées est nécessaire pour obtenir une puissance de 0,8 pour détecter une valeur d'effet de 0,2 ( $\alpha = 0,05$ ).

## Résultats

Le tableau 1 présente les caractéristiques démographiques des échantillons de chaque séquence. Le nombre de paires appariés qui ont pu être analysés était de 18 (la cible était de 20) pour les épreuves *Compréhension et exécution de consignes* et *Élaboration de phrases*, de 17 (sur 20) pour l'épreuve *Répétition de phrases* et de 13 (sur 14) pour l'épreuve *Identification de concepts*, en raison d'erreurs d'administration sans rapport avec le format Q-interactive. Chaque fois qu'une telle erreur s'est produite, la note du sujet dans les deux formats, ainsi que les notes du sujet apparié ont également été supprimées de l'analyse.

**Tableau 1 Caractéristiques démographiques de l'échantillon par groupe de séquences**

Caractéristiques démographiques	Trois autres épreuves		Identification de concepts	
	Numérique– Papier	Papier– Numérique	Numérique– Papier	Papier– Numérique
<b>Nombre de paires</b>	18	18	13	13
<b>Age (années)</b>	5–6	6	7	6
	7–8	8	7	7
	9–13	4	4	0
<b>Sexe</b>	Filles	9	9	6
	Garçons	9	9	7
<b>Niveau</b>	< 12 ans	1	4	1
<b>d'étude des parents</b>	Bac	6	2	3
	Post-Bac	3	6	1
	Licence	8	6	8

Les échantillons des groupes numérique-papier et papier-numérique sont parfaitement similaires dans les deux sexes, sont presque identiques en ce qui concerne l'âge et sont similaires en ce qui concerne le niveau d'étude des parents. Dans l'ensemble, l'échantillon reflète étroitement la population des enfants américains en ce qui concerne le sexe et le niveau d'étude des parents.

Le tableau 2 indique le nombre de paires appariées pour chaque mesure, ainsi que la moyenne et l'écart-type des notes standard pour les première et deuxième administrations dans chaque groupe de séquences (numérique-papier et papier-numérique). Les notes sont généralement plus élevées pour la deuxième administration que pour la première.

**Tableau 2 Moyennes (écart-types) des notes standard, par groupe de séquences et support de passation**

Épreuves	Paires	Séquence Numérique-Papier		Séquence Papier-Numérique	
		Numérique	Papier	Papier	Numérique
Identification de concepts	13	<del>10.23</del> (2.92) 10.38	10.54 (3.04)	10.92 (1.85)	<del>9.85</del> (2.12) 10.62
Répétition de phrases	17	11.35 (2.89)	12.47 (3.59)	11.12 (3.94)	11.41 (3.87)
Compréhension et exécution de consignes	18	11.00 (3.16)	11.11 (2.81)	10.33 (3.68)	10.50 (2.64)
Élaboration de phrases	18	11.22 (4.14)	11.17 (3.73)	12.39 (3.60)	12.89 (3.55)

L'ampleur et la signification statistique des effets du format sont présentées dans le tableau 3. Trois des valeurs de l'effet sont inférieures à 0,20 en valeur absolue et ne sont pas statistiquement significatives ; elles se situent dans les limites de tolérance pour considérer les formats comme équivalents. L'ampleur de l'effet de l'épreuve *Identification de concepts* est toutefois de -0,23 ( $p < 0,05$ ), ce qui indique que le rendement est inférieur avec l'administration sur Q-interactive.

**Tableau 3 Effets du format : différences de scores entre les administrations papier et numérique**

Épreuves	Score de changement				Effet du format	t	Taille d'effet
	Numérique-Papier		Papier-Numérique				
	Moyenne	Écart-type	Moyenne	Écart-type			
Identification de concepts	0.31	1.44	-1.08	2.36	-0.70	-2.05 -0.60	-0.23 -0.08
Répétition de phrases	1.12	1.69	0.29	1.16	-0.42	-1.78	-0.14
Compréhension et exécution de consignes	0.11	1.60	0.17	2.18	0.03	0.09	0.01
Élaboration de phrases	-0.06	1.73	0.50	2.28	0.28	0.86	0.09

Note : Voir ci-dessus pour la définition de l'effet du format. Un effet de format positif indique des notes plus élevées lors de l'administration numérique. Taille d'effet = effet du format / 3

Afin de déterminer la cause de cette différence de scores pour l'épreuve *Identification de concepts*, l'équipe a comparé les réponses des personnes interrogées qui figuraient sur les enregistrements vidéo avec les scores qui avaient été automatiquement attribués par le système Q-interactive. Les vidéos ont montré quatre cas pour l'item 9 et sept cas pour l'item 14, où le candidat a répondu correctement mais où le système Q-interactive a attribué une note de 0. Lors de ces items, l'examineur demande au candidat de pointer une rangée d'images. La cotation a été programmée de telle sorte que le fait de toucher une image dans la rangée a été considéré comme une réponse correcte, mais que le fait de toucher plus d'une image dans la rangée, ou de toucher la même image plusieurs fois, soit considéré comme une erreur. Cette règle de notation diffère de la notation de l'administration papier, dans laquelle toute désignation par le candidat d'une seule rangée, quel que soit le nombre de touches, soit noté comme correcte. Des instructions ont été fournies aux

examineurs sur l'écran de saisie et pendant la formation pour qu'ils passent outre le système et attribuent le point au sujet pour toute réponse indiquant qu'il avait l'intention de pointer la bonne rangée. Toutefois, les examineurs n'ont manifestement pas été en mesure de le faire de façon fiable.

Suite à cette constatation, la programmation de Q-interactive a été modifiée en conséquence de sorte qu'un nombre quelconque de réponses tactiles d'une ou plusieurs images d'affilées soit considéré comme une réponse correcte par le logiciel. En supposant que ce changement se traduise par une notation correcte des réponses tactiles multiples (ce qui a été confirmé dans les tests bêta), l'effet du format sur l'épreuve *Identification de concept* est éliminé. Lorsque les notes appropriées sont attribuées aux items 9 et 14 dans les données de l'étude d'équivalence, les notes moyennes étalonnées pour l'administration numérique passent de 10,23 à 10,38 pour le groupe séquence numérique-papier et de 9,85 à 10,62 pour le groupe séquence papier-numérique. La taille de l'effet est alors de -0,08, ce qui n'est pas statistiquement significatif ( $t = -0,60$ ).

## Discussion

Deux des épreuves étudiées dans cette étude, *Répétition de phrases* et *Élaboration de phrases*, exigent que l'examineur transcrive avec précision la réponse orale du sujet. Pour ces épreuves, les interfaces Q-interactive de l'iPad examineur ont bien fonctionné.

Pour les deux autres épreuves, le but de l'étude était de voir dans quelle mesure la correction automatique des réponses du sujet serait efficace avec les éléments nécessitant des associations ou séquences de réponses tactiles. Il n'y a pas eu de différence de résultats pour l'épreuve *Compréhension et exécution de consignes* entre les administrations papier et numérique sur Q-interactive. Un léger effet de format sur *Identification de concepts* a été attribué à deux items pour lesquels la règle de notation automatique divergeait de la règle utilisée dans l'administration papier ; une fois cette différence éliminée, le critère d'équivalence était rempli.

En raison de la nature très efficace de la méthode test-retest, il n'y avait pas suffisamment de paires dans cette étude pour permettre une évaluation de l'influence des caractéristiques démographiques (âge, sexe et niveau d'études des parents) sur les effets du format.

Cette étude d'équivalence de la CELF 5 s'ajoute à l'ensemble des données probantes sur l'effet (ou l'absence d'effet) des caractéristiques de la conception de l'interface digitale, sur la façon dont les examineurs saisissent et cotent les réponses, particulièrement pour les jeunes enfants. Elle ajoute de nouvelles informations sur la précision de la notation automatique des réponses tactiles du sujet. Au fur et à mesure que ce corpus de connaissances s'enrichira, il pourra être généralisé à d'autres tests du même type et présentant les mêmes caractéristiques.

## Références bibliographiques

### Rapports techniques Q-interactive:

- 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.
- 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 scales for children®-fourth edition*. Bloomington, MN: Pearson.
- Wechsler, D. (2008). *Wechsler adult intelligence scales®-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.
- Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical evaluation of language fundamentals®-fifth edition*. Bloomington, MN: Pearson.