



Pearson

# Equivalentie van een afname met Q-interactive™ en een afname op papier van cognitieve taken: WISC®-V

Q-interactive™ technisch rapport 8  
September 2014

Mark H. Daniel, PhD  
Dustin Wahlstrom, PhD  
Ou Zhang, PhD

**Vertaling en bewerking:**

Vertaalbureau Noorderlicht B.V.  
Joëlle Dek,  
*senior Product Developer*



# Q-interactive™

## technisch rapport 8

Mark H. Daniel, PhD

Dustin Wahlstrom, PhD

Ou Zhang, PhD

**September 2014**

**Vertaling en bewerking:**

Vertaalbureau Noorderlicht B.V.

Joëlle Dek, *senior Product Developer*

**Copyright: 2016 NCS Pearson, Inc.**

Alle rechten voorbehouden. Pearson, Q-interactive, en WISC zijn in de V.S. en/of andere landen handelsmerken van Pearson Education, Inc., of zijn aangesloten maatschappij(en).

# Inleiding

Q-interactive™ is een digitaal systeem van Pearson voor het individueel afnemen van tests. Het is ontworpen om assessments makkelijker en nauwkeuriger te maken, behandelaars eenvoudig toegang te geven tot een groot aantal tests en nieuwe manieren van testen te faciliteren die niet zonder hulp van een computer kunnen worden afgenomen of gescoord.

Q-interactive wordt gebruikt door middel van twee draadloos verbonden tablets; een van de testleider en een van de cliënt. Hierop kan de testleider afname-instructies lezen, responsen timen en deze vastleggen (inclusief audio-opnames). Ook kan via de tablet van de testleider de tablet van de cliënt bekeken en bestuurd worden. De tablet van de cliënt laat visuele stimuli zien en legt vast waar de cliënt het scherm heeft aangeraakt.

Het doel van het geschikt maken van testinstrumenten voor het Q-interactive platform was het behouden van de ruwe score-equivalentie tussen de standaard (papieren) en de digitale manier van afnemen en scoren. Als equivalentie wordt aangetoond dan kunnen de normen, betrouwbaarheid en validiteitsinformatie verzameld voor de papieren versie worden toegepast op Q-interactive-resultaten.

In dit onderzoek wordt de equivalentie van de scores van de digitale afname en de standaard afname van de *Wechsler Intelligence Scale for Children - Fifth Edition* (WISC-V; Wechsler, 2014) beoordeeld. Dit onderzoek is uitgevoerd in Amerika. In dit onderzoek zijn alle subtests van de Amerikaanse versie van de WISC-V onderzocht. Enkele van deze subtests (Information, Picture Concepts, Immediate Symbol Translation, Delayed Symbol Translation, Recognition Symbol Translation, Naming Speed Literacy en Naming Speed Quantity) zijn niet in de Nederlandstalige versie opgenomen.

In twee eerder uitgevoerde equivalentieonderzoeken leverden alle vijftien subtests van de *Wechsler Adult Intelligence Scale - Fourth Edition* (WAIS-IV; Wechsler, 2008) en dertien van de vijftien subtests van de *Wechsler Intelligence Scale for Children - Fourth Edition* (WISC-IV; Wechsler, 2003) vergelijkbare scores op bij de afname met Q-interactive en de standaard (papieren) afname. Op twee subtests van de WISC-IV (Matrix Redeneren en Plaatjes Concepten) waren de scores iets hoger bij de afname met Q-interactive.

Een derde onderzoek beoordeelde vier subtests (te weten de Trail Making Test, Design Fluency Test, Color Word Interference en

Verbal Fluency) van de *Delis-Kaplan Executive Function Scale* (D-KEFS; Delis, Kaplan, & Kramer, 2001) en de vrije reproductie-trials van de *California Verbal Learning Test - Second Edition* (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000), die allemaal de equivalentie aantoonde van de digitale en de papieren versies.

In het vierde onderzoek bleken drie subtests van de *NEPSY - Second Edition* (NEPSY-II; Korkman, Kirk, & Kemp, 2007) en twee subtests van de *Children's Memory Scale* (CMS; Cohen, 1997) equivalent te zijn. Het vijfde onderzoek evalueerde de subtests Oral Reading Fluency en Sentence Repetition van de *Wechsler Individual Achievement Test - Third Edition* (WIAT-III; Wechsler, 2009a), die beide voldeden aan het criterium voor equivalentie. In het onderzoek dat hierop volgde bleken alle subtests van de *Wechsler Memory Scale - Fourth Edition* (WMS-IV; Wechsler, 2009b) equivalent te zijn. Tenslotte werden de tests Zinnen Herhalen, Zinnen Formuleren, Aanwijzingen Volgen en Linguïstische Concepten, die onderdeel zijn van de *Clinical Evaluation of Language Fundamentals - Fifth Edition* (CELF-5; Semel, Wiig, & Secord, 2013), geëvalueerd.

De eerste twee tests, waarvoor de testleider een uitgebreid mondeling antwoord moet registreren en scoren, bleken equivalent te zijn. De laatste twee subtests waren de eerste subtests die het automatisch scoren van het aanraken van de tablet door de cliënt toepasten. Dit leverde equivalente scores op voor Aanwijzingen Volgen, maar niet voor Linguïstische Concepten, waarbij bepaalde patronen van aanrakingen van de cliënt bij twee items onterecht als fout werden gescoord. Nadat deze programmeerfout was hersteld werd equivalentie aangetoond.

Bij alle equivalentieonderzoeken wordt aangenomen dat de digitale afname (met Q-interactive) invloed kan hebben op de testcores, onder andere om de volgende redenen.

**Interactie van de cliënt met de tablet.** Om de effecten van de interactie tussen cliënt en tablet te beperken, zijn er tot nu toe fysieke testvoorwerpen (bijv. de blokken van Blokpatronen) en gedrukte antwoordenboekjes (bijv. het responsformulier voor Figuur Zoeken) gebruikt bij de afname met Q-interactive. Het uiteindelijke

doel is om deze onderdelen te vervangen door interactieve digitale interfaces. Hiervoor moet echter de strategie voor het aantonen van equivalentie (wat de basis levert voor het vertrouwen op het bestaande psychometrische bewijs) worden gewijzigd naar het aantonen van de validiteit en betrouwbaarheid van de digitale versie. Dit wordt nu voor het eerst gedaan voor de subtests voor verwerkingssnelheid van de WISC-V. Dit is echter nog niet afgerond en wordt niet beschreven in dit rapport, dat zich uitsluitend richt op equivalentie.

**Interactie van de testleider met de tablet, met name tijdens het vastleggen en scoren van antwoorden.** De meeste verschillen tussen de papieren versie en de afname op Q-interactive kwamen tot op heden voor in de interface van de testleider. Het afnemen van een test op Q-interactive is anders dan de standaard afname omdat er binnen Q-interactive instrumenten en procedures bestaan die zijn ontworpen om de taak van de testleider te vereenvoudigen en te ondersteunen. Er is zeer veel moeite gedaan om te zorgen dat deze aanpassingen de nauwkeurigheid waarmee de testleider instructies en stimuli presenteert, de prestaties controleert en timet, en antwoorden vastlegt en scoort, niet aantasten.

**Hoe nauwkeurig het Q-interactive-systeem de aanrakingen van de tablet door de cliënt vastlegt en scoort.** De eerste test waarvoor automatisch scoren werd aangeboden binnen Q-interactive was de CELF-5. Bij eerdere uitvoeringen moest de testleider voor elk item een score invullen. Hierdoor behoudt de testleider de controle, maar is er geen baat van de mogelijkheden van het tabletsysteem om scores op aanrakingen te herkennen en toe te wijzen.

**Globale effecten van de digitale assessmentomgeving.** De globale effecten gaan verder dan alleen de interactie tussen de cliënt of testleider en de tablet. Zo werd een globaal effect waargenomen in een eerder onderzoek waarin de test een toetsenbord gebruikte om de mondelinge antwoorden van de cliënt vast te leggen. Het bleek dat cliënten langzamer gingen antwoorden of zelfs kortere antwoorden gingen geven om te voorkomen te moeten wachten

op langzaam typende testleiders die nog bezig waren hun antwoorden letterlijk uit te typen. Een ander soort globaal effect werd geobserveerd bij een aantal zeer jonge kinderen (2 of 3 jaar oud) die afgeleid raakten door de tablet, omdat zij die als speelgoed zagen; dit probleem wordt aangepakt door middel van meer onderzoek naar menselijke factoren op deze leeftijden.

Als een taak niet equivalent was bij de twee afnamewijzen, werd tijdens de onderzoeken van Q-interactive de oorzaak van het digitale effect onderzocht. Het is essentieel de oorzaak van het afnamewijze-effect te begrijpen om te kunnen besluiten hoe dit moet worden aangepakt. Als werd vastgesteld dat testleiders nauwkeuriger waren bij afname of scoren met Q-interactive, dan zorgt Q-interactive voor een vooruitgang in de assessmenttechnologie. Een gebrek aan equivalentie hoeft dan niet noodzakelijkerwijs een probleem te vormen. Men mag redelijkerwijs van een nieuwe technologie verwachten dat deze resultaten produceert die equivalent zijn aan de resultaten van testleiders die de standaard papieren afnamewijze correct gebruiken. De digitale afnamewijze laat geen fouten toe in afname of scoren die in de standaard afnamewijze wel voorkomen. Als het er daarentegen op lijkt dat een digitaal effect wordt veroorzaakt door een afname van de nauwkeurigheid van de testleider of de cliënt, dan is het van groot belang het Q-interactive-systeem aan te passen om de oorzaak van die fout weg te nemen. Alleen als dit niet mogelijk is zou er iets aan het effect gedaan kunnen worden door de normen aan te passen.

In de eerdere equivalentieonderzoeken werden nagenoeg alle afnames op video opgenomen, zodat de interactie tussen de testleider, de cliënt en de tablet zichtbaar was. Op deze manier kon de nauwkeurigheid van de afname, alsmede de registratie en het scoren met zowel de digitale als de standaard afnamewijze, worden gecontroleerd als er een afnamewijze-effect (d.w.z. geen equivalentie) werd gevonden. Zo kon de oorzaak van het verschil worden vastgesteld en gecorrigeerd. Deze opnamen hadden het bijkomende voordeel dat zij lieten zien hoe testleiders en cliënten bij de verschillende afnamewijzen omgingen met het testmateriaal. Deze informatie kon gebruikt



worden om verbeteringen aan het ontwerp van de interface in te voeren. De afnames van de WISC-V werden in dit onderzoek niet op video opgenomen, aangezien het Q-interactive-team tijdens de eerste zeven onderzoeken (die twee jaar duurden en waarbij meer dan duizend afnames werden gedaan) zeer veel ervaring heeft opgedaan.

Samenvattend wijzen de equivalentieonderzoeken erop dat cliënten van vijf jaar en ouder (de jongste personen die werden getest) wanneer zij stimuli krijgen aangeboden op een tablet op vergelijkbare manier reageren als wanneer zij een gedrukt stimulusboek krijgen voorgelegd, of wanneer hun aanrakingen worden vastgelegd door het scherm in plaats van door observatie van de testleider. De enige uitzondering hierop werd gevonden bij de subtests Matrix Redeneren en Plaatjes Concepten van de WISC-IV. Het lijkt erop dat kinderen op subtests op het gebied van conceptueel redeneren met gedetailleerde visuele stimuli mogelijk beter presteren als de stimuli op een tablet worden getoond. (Het huidige onderzoek van de WISC-V biedt de mogelijkheid om de replicerbaarheid van deze bevindingen te onderzoeken.) Ook wijst al het bewijsmateriaal bij elkaar genomen erop dat als testleiders de verschillende interfaces van Q-interactive gebruiken, zij dezelfde resultaten krijgen als met het papieren materiaal.

# Onderzoeksopzet equivalentieonderzoek

Er werden verschillende experimentele opzetten gebruikt voor de equivalentieonderzoeken naar Q-interactive. In de meeste opzetten maakt elke proefpersoon een subtest slechts eenmalig, ofwel digitaal ofwel op papier (standaard). Met deze aanpak wordt voorkomen dat de manier waarop een proefpersoon de taak doet, verandert doordat hij/zij deze taak eerder heeft gedaan. Idealiter wordt geprobeerd alle effecten die de afnamewijze zou kunnen hebben op hoe de proefpersoon de taak voor het eerst doet op te sporen. Onderzoekopzetten waarin elke proefpersoon de test slechts een keer doet komen het dichtst in de buurt van een realistische testervaring.

De onderzoeken van de WAIS-IV en WISC-IV steunden hoofdzakelijk op een opzet met *equivalente groepen*, met of zonder random toewijzing van proefpersonen aan groepen. Deze opzet vergelijkt de prestaties van twee groepen, een groep die de test digitaal doet en een groep die de test op papier doet. De opzet met equivalente groepen wordt in detail beschreven in technisch rapport 1 en 2 van Q-interactive. Deze opzet, met random toewijzing, is ook gebruikt voor het onderzoek van de WISC-V.

Een andere opzet met een enkele afname, genaamd *dual-capture*, is geschikt als de digitale afnamewijze invloed heeft op hoe de testleider antwoorden vastlegt en scoort, maar er niet verwacht wordt dat dit invloed zal hebben op het gedrag van de proefpersoon. Elke proefpersoon uit een relatief kleine groep maakt de test slechts eenmalig, maar de afname wordt vanuit het perspectief van de testleider op video opgenomen. Op deze manier kan deze

worden bekeken door een aantal onderzoekers die de test scores met de papieren of de digitale afnamewijze. Een vergelijking van de gemiddelde scores op de twee afnamewijzen geeft aan of de afnamewijze invloed heeft op het proces van vastleggen en scoren van antwoorden. Details over deze opzet zijn te vinden in technisch rapport 3 (CVLT-II en D-KEFS), 5 (WIAT-III) en 6 (WMS-IV) op de Amerikaanse website en in de rapporten in het vertaalde rapport over de CVLT-II en D-KEFS op [helloq.nl](http://helloq.nl).

In de derde opzet, genaamd *test-hertest*, maakt elke proefpersoon de subtest twee keer, een keer met elke afnamewijze (in counter-balanced volgorde). Een test-hertestopzet is een zeer krachtige opzet, aangezien proefpersonen dienen als hun eigen controlegroep. Deze opzet is geschikt als de antwoordprocessen waarschijnlijk niet aanzienlijk zullen veranderen bij het hertesten, omdat de proefpersoon geen nieuwe oplossingen of nieuwe strategieën leert voor de aanpak van de taak of het oplossen van het probleem. De test-hertestopzet is gebruikt in het follow-uponderzoek van de subtests voor verwerkingssnelheid van de WAIS-IV (technisch rapport 1) en in de onderzoeken van de NEPSY-II (technisch rapport 4), WMS-IV (technisch rapport 6), en CELF-5 (technisch rapport 7).

Voor alle equivalentieonderzoeken is een effectgrootte van 0.2 of kleiner als standaard voor equivalentie aangehouden. De effectgrootte is het gemiddelde verschil tussen scores op Q-interactive en afname op papier, gedeeld door de standaardafwijking van de scores in de populatie. Een effectgrootte van 0.2 betekent dat er gemiddeld een verschil van iets meer dan een halve geschaalde scorepunt behaald kon worden, uitgaande van de veelgebruikte subtestschaal met een gemiddelde van 10 en een standaardafwijking van 3.

### **Selectie van deelnemers**

Voor de Q-interactive equivalentieonderzoeken (inclusief dit onderzoek) zijn steekproeven van niet-klinische proefpersonen gebruikt, zodat de aandacht kon worden gericht op het inschatten van de aanwezigheid en de omvang van eventuele effecten van de digitale afname. Omdat de mogelijke effecten van computerondersteunde

afname bij personen met bepaalde klinische aandoeningen niet bekend is, kan de deelname van proefpersonen met verschillende stoornissen aan de steekproef de interpreteerbaarheid van de resultaten in gevaar brengen. Het is uiteindelijk voor klinische toepassingen van Q-interactive van belang de interactie tussen afnamewijze en klinische aandoeningen te begrijpen; het eerste onderzoek richt zich echter op de belangrijkste vraag, namelijk of de digitale afname wel of geen invloed heeft op de scores verkregen bij niet-klinische proefpersonen.

Het is afhankelijk van de onderzoeksopzet hoeveel demografische representativiteit nodig is voor de steekproef. In de opzetten met equivalente groepen is het van belang dat de steekproeven die worden vergeleken een weergave vormen van de algemene populatie (geslacht, etnische achtergrond en sociaal-economische status [op basis van opleidingsniveau]) en dat de twee groepen demografisch vergelijkbaar zijn met elkaar. Bij *test-hertest* en *dual-capture*, die zich richten op vergelijkingen binnen proefpersonen, zijn de eigenschappen van de proefpersonen minder belangrijk; het is echter wel van belang dat er binnen de steekproef voldoende diversiteit in vaardigheidsniveaus en antwoordstijlen voorkomt om gevarieerde antwoorden te produceren, zodat de verschillende kenmerken van de digitale interface beoordeeld kunnen worden.

Testleiders die deelnamen aan de equivalentieonderzoeken waren opgeleid in de afnameprocedures voor de standaard papieren tests. Testleiders kregen voldoende training en oefening met de digitale afname en scoringsprocedures om de afname en vastlegging van antwoorden soepel te kunnen uitvoeren zonder veel aandacht te hoeven besteden aan de afnamewijze. De ervaring leert dat het een aanzienlijke hoeveelheid oefening vergt om een nieuwe afnamewijze goed onder de knie te krijgen.

# Equivalentieonderzoek WISC-V

# Methode

## Maten

De WISC-V is een uitgebreid instrument om intellectuele vaardigheden te beoordelen in de leeftijd van 6 tot en met 16 jaar.

Voor veel van de subtests is de afname en het scoren nagenoeg hetzelfde als voor de WISC-IV, die al geëvalueerd is wat betreft equivalentie van Q-interactive en de papieren versie. Maar aangezien verschillende subtests van de WISC-V nieuw zijn en een paar andere tests kleine veranderingen in afname of scoren hebben ondergaan (inclusief veranderingen aan de Q-interactive-interface voor de testleider), is de hele testbatterij opnieuw geëvalueerd wat betreft equivalentie.

## Deelnemers

Dit onderzoek werd uitgevoerd als onderdeel van de standaardisatie van de WISC-V. Bijna alle afnames vonden plaats in april en mei 2014, en een klein aantal afnames met Q-interactive vond in de maanden daarvoor plaats. De steekproef bestond uit niet-klinische kinderen van 6 tot en met 16 jaar. Mogelijke deelnemers werden gescreend op demografische eigenschappen en factoren voor uitsluiting, zoals perceptuele of motorische tekortkomingen of klinische aandoeningen. Volgens het steekproefplan waren ongeveer gelijke aantallen mannelijke en vrouwelijke proefpersonen nodig en ongeveer gelijke aantallen per leeftijd (jaar). Verder waren representatieve percentages van etnische achtergrond en niveaus van sociaal-economische status nodig (opleidingsniveau ouders).

Mogelijke deelnemers van wie de eigenschappen overeenkwamen met de demografische eisen van zowel de steekproef voor de papieren versie als die voor Q-interactive werden random toegewezen aan een van de afnamewijzen. Nadat de gegevens waren verzameld, selecteerde het onderzoeksteam paren (één uit elke afnamewijze) die overeenkwamen wat betreft leeftijdscategorie, geslacht, etnische achtergrond en opleidingsniveau van de ouders, om te gebruiken in de analyse.

De testleiders bevonden zich op verschillende locaties in het land. Alle testleiders kregen een training op afstand voor de WISC-V en de afname met Q-interactive, zodat zij bevoegd waren om de test op beide afnamewijzen af te nemen. Zij oefenden met de afname en kregen feedback over eventuele fouten tijdens de afname.

### Procedure

De testleiders legden de antwoordinformatie vast op de standaard wijze voor normering, waarbij onder andere het volledige letterlijke antwoord voor elk item van de verbale subtests werd uitgeschreven, en zij scoorden alle items. Het onderzoeksteam van Pearson controleerde de afnames op papier op juist gebruik van de afnameregels (zoals doorvragen, start- en afbreekregels), maar scoorde de items niet opnieuw. Alle totale ruwe scores van de subtests werden automatisch berekend, ofwel door medewerkers van Pearson die de itemscores uit de antwoordsleutel gebruikten, ofwel door het Q-interactive-systeem.

Het afnamewijze-effect werd voor elke subtest geschat door middel van multiële regressie met als onafhankelijke variabelen demografische gegevens (leeftijd, geslacht, etnische achtergrond en opleidingsniveau ouders), scores op bepaalde andere subtests of indexen van de WISC-V, en afnamewijze, en als afhankelijke variabele de geschaalde subtestscore. Hoewel het gebruik van proefpersoonparen gematcht op demografische gegevens met random toewijzing van afnamewijze al goede experimentele controle levert, verhoogt het gebruik van covariaten de statistische effectiviteit nog meer. Er werden WISC-V subtests in plaats van externe tests als covariaten gebruikt, omdat het onderzoek van de WISC-IV al had uitgewezen



dat deze subtests, op slechts een paar uitzonderingen na, nauwelijks worden beïnvloed door afnamewijze. De analyse werd in twee fasen uitgevoerd. De eerste fase gebruikte de volgende subtests of indexen die kleine afnamewijze-effecten hadden vertoond in het equivalentieonderzoek van de WISC-IV als predictorvariabelen: Blokpatronen, Rekenen, Cijferreeksen en de Verbaal Begrip Index. (Voor de analyse van elke verbale subtest werd de VBI vervangen door de subtest Overeenkomsten of Woordenschat.) De resultaten van deze fase werden gebruikt om de predictors verder te verfijnen en de analyse werd in fase twee herhaald.

Omdat de afnamewijze voor de papieren versie werd gecodeerd met een 0 en die van Q-interactive met een 1, was het niet gestandaardiseerde regressiegewicht voor afnamewijze een rechtstreekse schatting van het afnamewijze-effect in de normatieve score-eenheden voor die subtest. Dit werd omgezet naar een effectgrootte door te delen door de standaardafwijking van de normatieve score (3 of 15). Als we uitgaan van een multiële correlatie van .5 tussen de reeks predictoren voor demografische gegevens en subtestscore en de score op de doelsubtest, dan heeft een steekproef van in totaal 350 proefpersonen het onderscheidend vermogen van .56 om een effectgrootte van 0.20 (alfa = .05) waar te nemen.

Om vast te stellen of er een afnamewijze-effect zou kunnen zijn voor subgroepen van de populatie per leeftijd, geslacht, sociaal-economische status, etnische achtergrond of vaardigheidsniveau, werd een indicator voor afnamewijze-effect berekend voor elke Q-interactive-proefpersoon en werd de relatie van deze indicator tot elk van de individuele eigenschappen geanalyseerd. De indicator voor afnamewijze-effect was het verschil tussen de werkelijke subtestscore van de proefpersoon (met digitale afname) en de score die voor die proefpersoon werd voorspeld met de papieren afname op basis van de digitale variabelen en andere subtestscores van de WISC-V.

**TABEL 1** Demografische eigenschappen van de steekproeven

Afnamewijze			
Demografische eigenschappen		Papier	Q-interactive
Leeftijd (jaar)	Aantal proefpersonen	175	175
6		14	16
7		18	15
8		15	16
9		15	15
10		12	12
11		20	20
12		14	14
13		17	17
14		18	18
15		16	16
16		16	16
Gemiddelde		11.1	11.1
SD		3.2	3.2
Geslacht	Vrouwelijk	101	102
	Mannelijk	74	73
Etnische achtergrond	Afrikaans-Amerikaans	18	19
Latijns-Amerikaans		29	32
Blank		117	115
Overige		11	9
Ouder	<12 jaar onderwijs	17	12
Opleiding	<b>schooldiploma</b>	<b>HS</b>	<b>graduate</b>
		30	29
	Enige post-HS	54	55
	Diploma na 4 jaar	74	79

## Resultaten

In totaal werden 350 proefpersonen (175 gematchte paren) geselecteerd voor de analyse. Van de 113 deelnemende testleiders leverde 58% cases met beide afnamewijzen, 38% leverde alleen afnames met Q-interactive en 4% leverde alleen afnames met de papieren versie. Tabel 1 geeft de eigenschappen van de proefpersonen in de steekproeven die de afnames op papier en met Q-interactive deden.

**TABEL 2** Beschrijvende statistische gegevens voor subtests van de WISC-V, per afnamewijze

Subtest	Papier		Q-interactive	
	Gemiddelde	SD	Gemiddelde	SD
Rekenen	10.3	2.6	9.9	2.4
Blokpatronen	9.9	2.5	10.6	2.5
Begrijpen	9.9	2.6	9.5	2.6
Cijferreeksen	10.1	2.7	10.6	2.5
Gewichten	10.0	2.6	10.6	2.9
Informatie	10.2	3.0	10.2	2.6
Cijfers en Letters Nazeggen	10.3	2.5	10.7	2.4
Matrix Redeneren	9.9	2.5	10.6	2.9
Plaatjes Concepten	9.9	2.9	10.1	3.2
Plaatjesreeksen	10.3	2.5	10.7	2.7
Overeenkomsten	10.2	2.8	10.3	2.6
Figuur Samenstellen	9.8	2.6	10.0	2.7
Woordenschat	10.0	3.0	9.8	2.6
Immed. Symbol Translation	99.4	13.4	100.7	13.2
Delayed Symbol Translation	99.9	13.2	100.8	13.9
Recog. Symbol Translation	101.7	12.6	102.5	13.2
Naming Speed Literacy	100.7	13.9	103.0	14.2
Naming Speed Quantity	101.4	14.5	102.2	12.6

De twee steekproeven lijken voor elke demografische variabele erg op elkaar. De steekproef heeft in het algemeen een bijna gelijke vertegenwoordiging van elke leeftijd en heeft landelijk representatieve verhoudingen van etnische achtergrond. Vrouwelijke proefpersonen en kinderen van ouders met een academische opleiding zijn iets oververtegenwoordigd vergeleken met de algemene bevolking.

Tabel 2 geeft de gemiddelden en standaardafwijkingen van scores op de subtests van de WISC-V voor elke afnamewijze weer. Grote of systematische verschillen in scores tussen groepen zijn niet te verwachten, gezien de grote gelijkheid van de demografische eigenschappen van de twee versiegroepen en de random toewijzing van proefpersonen aan een afnamewijze.

De resultaten uit de eerste fase van de regressieanalyse leidden tot de selectie van een nieuwe reeks voorspellende variabelen met relatief lage afnamewijze-effecten voor gebruik in de tweede fase: Figuur Samenstellen, Plaatjes Concepten, Plaatjesreeksen, Cijfers en Letters Nazeggen en de Verbaal Begrip Index (voor analyse van de verbale subtests alleen Overeenkomsten of Woordenschat). Tabel 3 geeft de resultaten van de tweede reeks regressieanalyses, inclusief de multiële correlatie met de voorspellende variabelen, het niet gestandaardiseerde regressiegewicht voor afnamewijze, de *t*-waarde en statistische significantie geassocieerd met afnamewijze als predictor, en de effectgrootte van afnamewijze. Ter vergelijking worden de resultaten uit het equivalentieonderzoek van de WISC-IV gegeven.

De multiële correlaties liepen uiteen van .35 tot .71 (mediaan = .52), wat aangeeft dat de demografische en vaardigheidsvariabelen gemiddeld ongeveer een kwart van de variantie in de subtestscore verklaarden. Het doel van de predictorvariabelen was het vergroten van het onderscheidend vermogen van de analyses door de hoeveelheid variantie die verklaard moest worden te verkleinen; en dus zijn hoge multiële correlaties wenselijk, maar niet essentieel voor de validiteit van de analyses.

**TABEL 3** Effectgrootte van afname met Q-interactive op elke subtest van de WISC-V

Subtest	R	Niet gestand. Regressie gewicht	t	Effect grootte	Effectgrootte WISC-IV
Rekenen	.55	-0.49	-2.11*	-0.16	0.10
Blokpatronen	.58	0.59	2.66**	0.20	0.02
Begrijpen	.55	-0.59	-2.51*	-0.20	0.00
Cijferreeksen	.54	0.25	1.04	0.08	0.13
Gewichten	.52	0.49	1.95	0.16	—
Informatie	.71	-0.15	-0.68	-0.05	0.07
Cijfers en Letters Nazeggen	.50	0.26	1.13	0.09	0.18
Matrix Redeneren	.48	0.51	1.99	0.17	0.27
Plaatjes Concepten	.40	0.07	0.22	0.02	0.21
Plaatjesreeksen	.42	0.21	0.83	0.07	—
Overeenkomsten	.66	0.11	0.50	0.04	0.02
Figuur Samenstellen	.52	0.11	0.46	0.04	—
Woordenschat	.66	-0.39	-1.69	-0.13	0.05
Immed. Symbol Translation	.57	0.52	0.44	0.03	—
Delayed Symbol Translation	.50	0.21	0.16	0.01	—
Recog. Symbol Translation	.44	-0.07	-0.06	0.00	—
Naming Speed Literacy	.41	1.73	1.24	0.12	—
Naming Speed Quantity	.35	-0.37	-0.27	-0.02	—

**NB:** Een positieve effectgrootte wijst op hogere scores met Q-interactive. Het niet gestandaardiseerde regressiegewicht valt binnen de normatieve scoreschaal van de subtest (SD van 3 voor de eerste 13 subtests en 15 voor de laatste vijf).

\* $p < .05$

\*\* $p < .01$

Drie van de achttien subtests lieten een statistisch significant afnamewijze-effect zien bij een significantieniveau van .05, maar geen van de effectgroottes overschreed het criterium van 0.20, wat gebruikelijk is voor de equivalentieonderzoeken van Q-interactive. Voor de tien subtests die de WISC-IV en WISC-V gemeen hebben, was het patroon van effectgroottes redelijk gelijk in de twee onderzoeken ( $r = .44$ ), maar er waren een aantal wezenlijke verschillen. De twee subtests met de grootste effectgroottes op de WISC-V, Blokpatronen en Begrijpen, hadden bij het onderzoek van de WISC-IV effectgroottes van nagenoeg 0. De twee subtests met effectgroottes groter dan 0.20 in het onderzoek van de WISC-IV hadden daarentegen gematigde (Matrix Redeneren, 0.17) tot kleine (Plaatjes Concepten, 0.02) effectgroottes in het onderzoek van de WISC-V. Bovendien waren de effectgroottes in het onderzoek van de WISC-V beter in balans (vijf negatief en dertien positief), terwijl alle effectgroottes van de WISC-IV (exclusief verwerkingsnelheid) positief waren.

De mogelijke aanwezigheid van differentiële afnamewijze-effecten voor proefpersonen met verschillende vaardigheidsniveaus of verschillende demografische eigenschappen werd in een reeks analyses onderzocht. Er werd een indicator voor afnamewijze-effect berekend voor elke proefpersoon waarbij met Q-interactive een subtest werd afgenomen. Deze indicator was het verschil tussen de werkelijke subtestscore van de proefpersoon en zijn/haar voorspelde score bij de papieren afname, waarbij als predictors de reeks demografische variabelen en subtestscores uit de analyses uit fase twee werden gebruikt. Voor de continue predictorvariabelen (vaardigheid, leeftijd en sociaal-economische status) was de maat voor relatie de correlatiecoëfficiënt tussen elke variabele en de indicator van het afnamewijze-effect; voor geslacht was het de  $t$ -waarde; en voor etnische achtergrond was het de  $F$ -waarde uit de variantieanalyse. In deze analyse was de vaardigheidsscore de voorspelde score voor papieren afname. De resultaten staan in tabel 4.

Het is op basis van toeval te verwachten dat er tussen 90 statistische tests vier of vijf statistisch significante resultaten voorkomen (op het niveau .05), maar slechts twee van de resultaten in tabel

**TABEL 4** Relatie van afnamewijzeffect met vaardigheidsniveau en demografische gegevens

Subtest	Correlatie			• Geslacht ( <i>t</i> ) <sup>b</sup>	Etnische achtergrond ( <i>F</i> )
	Vaardigheid <sup>a</sup>	Leeftijd	SES		
Rekenen	-.01	.03	-.05	-0.40	0.06
Blokpatronen	.02	.04	.03	0.08	0.67
Begrijpen	.00	-.05	-.02	0.63	0.88
Cijferreeksen	-.03	.06	-.01	-0.80	0.33
Gewichten	.03	-.03	-.04	-0.46	0.42
Informatie	-.10	-.07	-.04	-0.21	0.78
Cijfers en Letters Nazeggen	-.14	.03	-.06	0.83	1.08
Matrix Redeneren	.02	-.01	-.01	0.55	0.17
Plaatjes Concepten	.04	.04	-.02	-0.06	0.59
Plaatjesreeksen	.04	.05	.09	-0.20	0.66
Overeenkomsten	.04	.07	-.04	-0.60	0.47
Figuur Samenstellen	.09	-.07	.11	0.24	0.91
Woordenschat	-.11	-.01	-.07	0.37	0.69
Immed. Symbol Translation	-.02	-.06	-.01	-0.70	1.26
Delayed Symbol Translation	-.02	-.10	-.07	-0.35	2.17
Recog. Symbol Translation	-.02	-.07	-.01	-0.13	0.98
Naming Speed Literacy	-.10	.03	-.08	-1.80	0.10
Naming Speed Quantity	-.06	.15*	-.12	-2.00*	0.64

*N* = 175. Voor deze analyse is het afnamewijze-effect het verschil tussen de werkelijke subtestscore op Q-interactive en de score voorspeld uit demografische gegevens en andere subtests voor afname op papier.

<sup>a</sup> Voorspelde subtestscore voor afname op papier.

<sup>b</sup> Een positieve waarde van *t* betekent dat het afnamewijze-effect groter was voor vrouwelijke proefpersonen.

\**p* < .05.

4 waren statistisch significant. Beide resultaten werden gevonden voor de subtest Naming Speed, waarbij de tendens bestond dat oudere proefpersonen meer baat hadden bij de afname met Q-interactive dan jongere proefpersonen, en mannelijke meer dan vrouwelijke. Dit zijn subtests die in de Nederlandstalige versie van de WISC-V niet zijn opgenomen. Over het algemeen ondersteunen de bevindingen in tabel 4 de vergelijkbare bevinding in het onderzoek van de WISC-IV, namelijk dat er geen differentiële effecten zijn van afname met Q-interactive voor vaardigheidsniveau, leeftijd, sociaal-economische status, geslacht en etnische achtergrond.

### Discussie

Dit onderzoek was een herhaling van het equivalentieonderzoek van tien van de subtests van de WISC-IV en tegelijkertijd het eerste onderzoek van acht nieuwe subtests. In overeenstemming met het overweldigende patroon van resultaten in de zeven eerdere equivalentieonderzoeken vielen alle afnamewijze-effecten binnen het vastgestelde criterium voor equivalentie voor Q-interactive (d.w.z. 0.20 of minder). Twee subtests die effectgroottes groter dan 0.20 hadden bij de WISC-IV (Matrix Redeneren en Plaatjes Concepten) hadden in dit onderzoek kleinere effectgroottes. Bij de twee subtests die de grootste effectgroottes hadden in het onderzoek van de WISC-V (+0.20 voor Blokpatronen en -0.20 voor Begrijpen) komt nauwelijks tot geen interactie van de proefpersoon met de tablet kijken: voor Blokpatronen toont de tablet van de proefpersoon simpelweg het patroon dat moet worden gemaakt en voor Begrijpen wordt de tablet van de proefpersoon helemaal niet gebruikt. Verder hadden deze subtests in het onderzoek van de WISC-IV effectgroottes van respectievelijk slechts 0.02 en 0.00.

Gaandeweg blijkt uit de equivalentieonderzoeken van Q-interactive dat in sommige onderzoeken wel kleine effecten worden waargenomen maar dat deze effecten, tenzij een oorzaak kan worden gevonden (door bijvoorbeeld de video-opnames van de afnames te bestuderen), over het algemeen niet replicerbaar of systematisch zijn en waarschijnlijk hoofdzakelijk worden veroorzaakt doordat er geen perfecte onderzoeksopzet bestaat om dit soort kwesties zonder fouten te evalueren. Desalniettemin stelden



deze onderzoeken verschillende malen werkelijke gebreken vast die vóór de uitgave werden gecorrigeerd maar die anders misschien niet waren opgemerkt (bijv. de fout in de oorspronkelijke timingfunctie voor de subtests voor verwerkingssnelheid van de WAIS-IV en het subtiele probleem met de beeldkwaliteit voor Onvolledige Tekeningen van de WAIS-IV). De combinatie van empirische resultaten en de nauwkeurige bestudering van de afname- en scoringsprocedures geeft een betrouwbaarheidsniveau voor equivalentie dat met slechts een van de twee methoden niet verkregen had kunnen worden.

Het huidige onderzoek toonde aan dat er nagenoeg geen statistisch significante verschillen waren in afnamewijze-effect tussen de subgroepen wat betreft leeftijd, geslacht, etnische achtergrond, sociaal-economische status of vaardigheidsniveau.

Dit werd in de eerdere equivalentieonderzoeken van Q-interactive ook steeds weer gevonden. Deze reeks resultaten geeft aan dat de algemene bevinding dat er geen afnamewijze-effect gevonden wordt in brede zin van toepassing is op de algemene niet-klinische populatie.

## Referenties

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

**Copyright: 2016 NCS Pearson, Inc.**

Alle rechten voorbehouden. Pearson, Q-interactive, en WISC zijn in de V.S. en/of andere landen handelsmerken van Pearson Education, Inc., of zijn aangesloten maatschappij(en).



