

## Ekvivalensstudier for WISC-V

I USA er det gjennomført en rekke studier med henblikk på å undersøke ekvivalensen mellom papirutgaven av WISC-V og den digitale versjonen som administreres med Q-interactive. Disse studiene har hatt en rekke ulike formål:

- Å sikre at de to administreringsmetodene genererer sammenlignbare resultater – det vil si at samme sett med normer kan brukes til både papirutgaven og den digitale versjonen.
- Å sikre at de digitale versjonene av Koding og Symbolleting er likeverdige med versjonene som administreres med papir og blyant, til tross for endringene i oppgaver og grensesnitt som er gjort i den digitale versjonen.
- Å sikre at administrering av WISC-V med Q-interactive genererer likeverdige resultater som papirutgaven for spesielle grupper (særlig begavede barn og barn med intellektuell funksjonsnedsettelse).

Nedenfor følger en kort oppsummering av disse studiene. Den som vil se på studiene i sin helhet (inkludert referanser til andre forskningsstudier), henvises til følgende rapporter, som alle er å finne på [www.helloq.com](http://www.helloq.com).

- Daniel, Wahlstrom, & Zhang (2014). Q-interactive Technical Report 8. *Equivalence of Q-interactive and Paper Administrations of Cognitive Tasks: WISC-V*.
- Raiford, Holdnack, Drozdick, & Zhang (2014). Q-interactive Technical Report 9. Q-interactive Special Group Studies: *The WISC-V and Children with Intellectual Giftedness and Intellectual Disability*.
- Raiford, Holdnack, Drozdick, & Zhang (2015). Q-interactive Technical Report 11. Q-interactive Special Group Studies: *The WISC-V and Children with Autism Spectrum Disorder and Accompanying Language Impairment or Attention-Deficit/Hyperactivity Disorder*.
- Raiford et al. (2016). Q-interactive Technical Report 12. *WISC-V Coding and Symbol Search in Digital Format: Reliability, Validity and Special Group Studies*.

### Studie 1: Ekvivalens mellom papirutgaven av WISC-V og WISC-V på Q-interactive

Ekvivalensstudien ble gjennomført i forbindelse med det amerikanske standardiseringsarbeidet, som også omfattet administrering av WISC-V med Q-interactive. 350 barn i 175 par, matchet med hensyn til kjønn, alder og foreldrenes utdanningsnivå, ble valgt for analyse etter at normeringen var avsluttet.

Resultatet viste at bare tre av de 18 deltestene (den amerikanske versjonen av WISC-V har flere deltester enn den norske) viste en signifikant effekt av administreringsformat på 5 %-nivået. Disse var Terningmønstre, Resonnering og Regning. Ingen av effektstørrelsene overskred imidlertid det oppsatte kriteriet  $d = 0.2$ . Mønsteret av effektstørrelser var ikke i samsvar med dem som var målt for tilsvarende deltester i WISC-IV. De to deltestene med størst effektstørrelse i WISC-V, Terningmønster og Resonnering, viste effektstørrelser nær 0 i WISC-IV-studien. De to deltestene som ga større effektstørrelser enn  $d = 0.2$  i WISC-IV, hadde moderate (Matriser) eller lave (Bildekategorier) effektstørrelser i WISC-V.

Det ble også gjennomført en serie med analyser for å utlede hvorvidt prestasjonsnivå eller bakgrunnsvariabler (kjønn, alder og foreldrenes utdanningsnivå) hadde sammenheng med effekter av administreringsformat. Av til sammen 90 gjennomførte analyser falt bare to ut med signifikante resultater, og begge gjaldt deltesten Benevnelsehastighet. Denne deltesten er ikke med i WISC-V-versjoner utenfor USA.

### Studie 2: Koding og Symbolleting

I motsetning til øvrige deltester er Koding og Symbolleting i Q-interactive ikke parallelle med papirutgavene. Tilpasning av disse deltestene til et digitalt format har ført til et nytt grensesnitt på oppgavene. For eksempel registreres svarene på Q-interactive ved at barnet trykker på iPad-en, nye stimuli presenteres automatisk etter at barnet har angitt sitt svar, kravet til at barnet skal skrive er eliminert, og det er ikke lenger mulighet for barnet å korrigere sine svar (med unntak av Demonstrasjons- og øvingsoppgavene).

De digitale versjonene av deltestene oppviste ikke råskåreekvivalens i forhold til papirutgavene, og en

statistisk konverteringsprosess, skalaekvivalering (eng. *equating*) ble gjennomført for å tilpasse raskårene fra de digitale versjonene slik at de blir likeverdige som raskårene fra papirutgavene. Det vil si at omregning av raskårer til skalerte skårer indikerer samme grad av prestasjon for begge administreringsformatene, noe som i sin tur betyr at samme oppsett med normer kan brukes.

Slik skalaekvivalering kan justere for forskjeller i vanskelighetsgrad mellom testene, men ikke for forskjeller i innhold. Et første, nødvendig skritt var derfor å sikre at de digitale versjonene av Koding og Symbolleting oppviste tilstrekkelig stor likhet vedrørende konstrukt (innhold og måleområde) og psykometriske egenskaper som papirutgavene. En serie med trinnvise tilpasninger og justeringer i de digitale versjonene av Koding og Symbolleting ble gjennomført, hvorpå data ble samlet inn for et representativt utvalg barn ( $N = 329$ ) i alderen 6:0–16:11 år.

Avhengig av deltestenes karakter er *reliabiliteten* vurdert med test-retest-metoden (split-half-reliabilitet er ikke egnet til beregning for tester som ikke kan deles i to likeverdige halvdelers). Analysene viste at de digitale versjonene av Koding og Symbolleting oppviste test-retest-reliabilitet som tilsvarte den for papirformatet. De gjennomsnittlige korrigerede reliabilitetskoeffisientene for Koding var 0.75 og for Symbolleting 0.80.

Den *interne strukturen* for interkorrelasjoner mellom deltester og indekser oppviste et lignende mønster av korrelasjoner som i papirutgaven.

*Konfirmatorisk faktoranalyse* (CFA) for to modeller ble gjennomført, den første med Koding og Symbolleting byttet ut med tilsvarende digitale versjoner (modell 1), og den andre med bare papirutgavene av alle tester (modell 2). Statistiske tilpasningsmål viste samme utmerkede tilpasning for begge modeller, noe som gir sterk støtte for at Koding og Symbolleting måler lignende konstrukter uavhengig av versjon og administreringsformat.

*Analyse av ekvivalensen mellom papirutgaver og digitale versjoner* viste et standardavvik mellom versjonene på  $d = .10$  for Koding og  $d = -.13$  for Symbolleting. Disse effektstørrelsene faller under kriteriet Pearson har satt opp med en effektstørrelse på maksimum  $d = +/- .20$  for akseptabel ekvivalens. Til sammenligning betyr en effektstørrelse på  $d = .20$  vel en halv skalert skåre på en skala med gjennomsnittsverdien 10 og standardavviket 3.

*Studier av Koding og Symbolleting i spesielle grupper:* For å prøve ut validiteten og den kliniske nytteverdien ytterligere, ble et utvalg (ikke-tilfeldige) av ulike kliniske grupper sammenlignet med matchede kontrollgrupper når Koding og Symbolleting ble administrert i digital versjon, samt andre deltester. De kliniske gruppene var blant annet særlig begavede barn, barn med lett intellektuell funksjonsnedsettelse og barn med spesifikke lærevansker. Resultatet av sammenligningene mellom disse kliniske gruppene og matchede kontrollgrupper viser konsekvente resultater i forhold til tidligere forskning på kliniske grupper.

### Studie 3 og 4: WISC-V med Q-interactive i spesielle grupper

To studier er gjennomført for å undersøke ekvivalensen mellom administrering av WISC-V med papir og blyant og med Q-interactive i spesielle grupper. Den første omfattet barn med lett intellektuell funksjonsnedsettelse og særlig begavede barn, og den andre barn med ADHD og barn med autisme. Resultatene fra disse ulike gruppene er kort oppsummert nedenfor. Symbolleting og Koding var ikke inkludert i studiene ettersom disse deltestene ikke eksisterte i digitale versjoner når studiene ble gjennomført, og resultater for FSIQ kan derfor ikke presenteres.

#### Særlig begavede barn

Til sammen 24 barn identifisert som særlig begavede, deltok i studien; 45.8 prosent jenter og 54.2 prosent gutter. Gjennomsnittsalderen var 11.5 år. En matchet kontrollgruppe ble hentet fra studie 1.

Gruppen med særlig begavede barn skåret signifikant høyere enn den matchede kontrollgruppen på alle primære og sekundære indekser. Gjennomsnittsverdien for de primære indeksskårene varierte mellom 121.3 (FRI) og 127.6 (VFI). Gjennomsnittsverdien for GEI var 127.1. Effektstørrelsene var store for alle primære indekser.

Gjennomsnittsverdien for de sekundære indeksene varierte mellom 123.9 (MRI) og 127.1 (GEI). Effektstørrelsene var store for alle sekundære indekser. Gjennomsnittsverdien for GEI i denne gruppen var identisk i standardiseringen av papirutgaven, 127.1. Videre analyser viste at 75 prosent av barna som var identifisert som særlig begavede, oppnådde en GEI-skåre på 120 eller høyere, sammenlignet med bare 13 prosent av barna i kontrollgruppen, noe som er i tråd med tilsvarende studie gjennomført ved standardiseringen av papirversjonen.

Gjennomsnittsverdien for alle deltester med unntak av Matriser var signifikant høyere i gruppen med særlig begavede barn enn i den matchede kontrollgruppen. Alle sammenligninger på deltestnivå påviste moderate til store effektstørrelser, størst for Ordforståelse, Likheter, Informasjon og Regning, noe som er i samsvar med resultatene av standardiseringen av papirversjonen av WISC-V.

## Barn med lett intellektuell funksjonsnedsettelse

Til sammen 22 barn diagnostisert med lett intellektuell funksjonsnedsettelse, deltok i studien; 36.4 prosent jenter og 63.6 prosent gutter. Gjennomsnittsalderen var 11.8 år. En matchet kontrollgruppe ble trukket fra studie 1.

Gruppen barn med lett intellektuell funksjonsnedsettelse skåret signifikant lavere enn barn i den matchede kontrollgruppen på samtlige primære og sekundære indekser. Gjennomsnittsverdien for de primære indeksskårene varierte fra 62.3 (VFI) til 67.9 (FRI). Gjennomsnittsverdien for GEI var 63.7. Effektstørrelsene var store for alle primære indekser. Gjennomsnittsverdien for de sekundære indeksene varierte mellom 61.6 (AAI) og 64.7 (MRI). Effektstørrelsene var store for alle sekundære indekser. Gjennomsnittsverdien for GEI i denne gruppen med barn var i standardiseringsversjonen av papirutgaven 63.5. Videre analyser viste at 91 prosent av barna som ble diagnostisert med lett intellektuell funksjonsnedsettelse oppnådde en GEI-skåre på 75 eller høyere, sammenlignet med bare 5 prosent av barna i kontrollgruppen.

Gjennomsnittsverdien for alle deltester var lavere i gruppen barn med lett intellektuell funksjonsnedsettelse enn i den matchede kontrollgruppen. De største effektstørrelsene var i deltestene Tall-bokstav-serier, Tallhukommelse, Regning, Informasjon, Figurvekter, Resonnering og Bildehukommelse.

## Barn med ADHD

Til sammen 25 barn diagnostisert med ADHD, deltok i studien; 36 prosent jenter og 64 prosent gutter. Gjennomsnittsalderen var 10.5 år. En matchet kontrollgruppe ble trukket fra studie 1.

Sammenlignet med gruppen barn med ADHD i standardiseringen av papirutgaven hadde barna i Q-interactive-gruppen lavere gjennomsnittsalder og foreldre med et høyere utdanningsnivå. Det ble ikke gjort noen vurdering av symptomenes alvorlighetsgrad i de to gruppene. På bakgrunn av dette ikke overraskende at de to ADHD-gruppens resultater er forskjellige. Retningene for de målte forskjellene (mot de matchede kontrollgruppene) er likevel de samme, og gjennomsnittresultatene for alle indeksene og deltestene er likartede, i likhet med effektstørrelsene. Forskjellene som ble påvist mellom ADHD-gruppen som gjennomførte papirutgaven, og ADHD-gruppen som brukte Q-interactive-versjonen, kan i hovedsak forklares av gruppenes ulike bakgrunnsvariabler og grad av symptomer.

Gruppen barn med ADHD oppnådde gjennomsnittlige resultater på alle primære deltester, men én signifikant forskjell fra den matchede kontrollgruppen ble målt for FRI. FRI oppviser en moderat effektstørrelse, og for VFI, VSI og AHI er effektstørrelsene små. Vedrørende de sekundære indeksene presterer ADHD-gruppen signifikant lavere på AAI og GEI, og effektstørrelsene er moderate. Disse resultatene indikerer vansker med visuospatial bearbeiding, flytende resonnering og auditiv arbeidsminne.

Gjennomsnittsverdien for to deltester (Matriser og Tall-bokstav-serier) var signifikant lavere i gruppen barn med ADHD enn i den matchede kontrollgruppen. Effektstørrelsen for Matriser er stor. Moderate effektstørrelser viste seg for Terningmønster, Regning og Tall-bokstav-serier.

## Barn med autisme

Til sammen 30 barn diagnostisert med autisme med lett nedsatt språkevnne, deltok i studien; 10 prosent jenter og 90 prosent gutter. Gjennomsnittsalderen var 11.6 år. En matchet kontrollgruppe ble trukket fra studie 1 beskrevet ovenfor.

Gruppen barn med autisme med nedsatt språkevnne skåret signifikant lavere enn kontrollgruppen på samtlige primære indekser, og effektstørrelsene var gjennomgående store. I samsvar med tidligere forskning var FRI og VSI relativt høyere og oppviste mindre effektstørrelser enn VFI. VFI hadde den nest største effektstørrelsen av de primære indeksene, etter AHI. Vedrørende de sekundære indeksene skåret gruppen barn med autisme lavest på AAI, som også hadde den største effektstørrelsen.

Gruppen barn med autisme med nedsatt språkevnne skåret signifikant lavere på samtlige primære og sekundære deltester sammenlignet med kontrollgruppen. I samsvar med tidligere forskning ble den høyeste gjennomsnittsverdien i Verbal Forståelsesindeks oppnådd for Likheter, og den laveste for Resonnering. De største effektstørrelsene ble notert for Resonnering, Tall-bokstav-serier, Ordforståelse, Bildehukom-

melse, Regning og Tallhukommelse. De minste effektstørrelsene ble målt for Figurvekter, Matriser, Visuelle puslespill og Terningmønster.

Resultatene for både indekser og deltester er svært like de som forekom i tilsvarende studie ved standardiseringen av papirutgaven.

## Oppsummering

Studiene som så langt er gjennomført med henblikk på å sikre ekvivalens mellom papirutgaven av WISC-V og den digitale versjonen Q-interactive, har vist små effekter av administreringsformat, både i normalpopulasjonen og i spesielle grupper. I normalpopulasjonen er det ikke målt noen effektstørrelser som overskrider  $d = .20$  for noen indekser eller deltester. Det er ikke påvist at noen bakgrunnsvariabler eller prestasjonsnivåer har sammenheng med effekter av administreringsformatet. I de spesielle gruppene som er undersøkt, genererer administrering av WISC-V med Q-interactive like prestasjonsnivåer og mønstre som i de studiene som ble gjennomført ved standardiseringen av papirutgaven. Revideringen som har vært nødvendig for å digitalisere deltestene Koding og Symbolleting, har resultert i reliable og valide digitale deltester, og det er gjennomført en skalaekvivalering som tillater bruk av normene for papirutgaven til disse deltestene. Denne skalaekvivaleringen skjer automatisk i Q-interactive og genererer den korrekte skalerte skåren for en bestemt råskåre.

Det er viktig å demonstrere ekvivalens mellom papirversjoner og digitale versjoner når en test kan administreres på begge måter. Fremtiden vil vise hvorvidt papirutgavene av testene blir utdatert – allerede nå pågår standardiseringsprosjekter som vil føre til at bare digitale versjoner blir utgitt.