

WISC-V ækvivalensstudier

I USA er et antal undersøgelser blevet gennemført med det formål at undersøge ækvivalensen mellem papir og blyant-versionen af WISC-V (herefter p/b), og den digitale version som administreres på Q-interactive. Disse undersøgelser har haft en række forskellige formål:

- At sikre at de to administreringsmåder genererer sammenlignelige resultater – det vil sige at samme opsætning af normer kan anvendes for såvel p/b-versionen som den digitale version.
- At sikre at de digitale versioner af delprøverne Kodning og Figursøgning er ligeværdige med p/b-versionerne, til trods for de forandringer af opgaver og grænseflader som blev gennemført i de digitale versioner.
- At sikre at administrering af WISC-V på Q-interactive producerer resultater der er ligeværdige med p/b-versionen i særlige grupper.

Nedenfor følger en kort sammenfatning af disse undersøgelser. Interesserede i undersøgelserne i deres helhed (inklusive referencer til andre forskningsundersøgelser) henvises til følgende rapporter som findes på 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*.

Undersøgelse 1: Ækvivalens mellem WISC-V p/b-version og WISC-V på Q-interactive

Ækvivalensstudier blev gennemført i forbindelse med det amerikanske standardiseringsarbejde, hvilket også omfattede administrering af WISC-V på Q-interactive. 350 børn i 175 par, matchet med hensyn til køn, alder, forældres uddannelsesniveau og etnicitet, blev udvalgt til analyse efter at dataindsamlingsarbejdet var afsluttet.

Analyserne viste at kun tre ud af de 18 delprøver (den amerikanske version af WISC-V indeholder flere delprøver end de ikke-engelsksprogede) udviste en signifikant effekt af format på 5 %-niveauet; Blokmønstre, Verbal ræsonnering og Regning. Ingen af effektstørrelserne overskred imidlertid det opsatte kriterium på $d = 0.2$. Mønstre af effektstørrelser var ikke konsekvent med dem som blev målt for tilsvarende delprøver i WISC-IV. De to delprøver med de største effektstørrelser i WISC-V, Blokmønstre og Verbal ræsonnering, udviste, i WISC-IV-undersøgelsen, effektstørrelser tæt på 0. De to delprøver som udviste større effektstørrelser end $d = 0.2$ i WISC-IV, havde moderate (Matricer) eller lave (Billedkategorier) effektstørrelser i WISC-V.

En serie af analyser blev gennemført for også at vurdere hvorvidt præstationsniveau eller baggrundsvARIABLES (køn, alder, forældres uddannelsesniveau og etnicitet) havde sammenhæng med effekter af administreringsformatet. Ud af sammenlagt 90 gennemførte analyser, faldt kun to ud med signifikante resultater; begge omhandlede delprøven Naming Speed. Denne delprøve er ikke inkluderet i WISC-V-versioner uden for USA.

Undersøgelse 2: Kodning og Figursøgning

I modsætning til de øvrige delprøver er delprøverne Kodning og Figursøgning på Q-interactive ikke parallelle med p/b-versionerne. Tilpasning af disse delprøver til et digitalt format har nødvendiggjort en ny grænseflade mellem opgaverne og barnet. Eksempelvis markeres svar nu ved at barnet trykker på iPad'en, stimuli præsenteres automatisk efter at barnet svaret, kravet til at barnet skal skrive er nu elimineret, og der findes ikke længere mulighed for selvkorrigerende af svar.

De digitale versioner af disse delprøver udviser ikke råscoreækvivalens i forhold til p/b-versionerne, hvorfor en statistisk konverteringsproces, skalaækvivalering (eng *equating*) er blevet gennemført, med det formål at tilpasse råscoren på de digitale versioner så de bliver ligeværdige med råscoren fra p/b-versionerne. Dette indebærer en transformering af råscorer til skalascorer som indikerer samme grad af præstation for begge administreringsformater, hvilket resulterer i, at samme opsætning af normer kan anvendes.

En sådan skalaækvivalering kan justeres for forskelle i sværhedsgrad mellem test, men ikke for forskelle i indhold. Et første nødvendigt skridt var derfor at sikre at de digitale versioner af delprøverne Kodning og Figursøgning udviste tilstrækkelig høj lighed med hensyn til konstrukt (indhold og målområde) og psyko-metriske egenskaber som tilsvarende p/b-versioner. En række tilpasninger og justeringer af de digitale versioner af Kodning og Figursøgning blev gennemført, hvorefter data blev indsamlet i et repræsentativt udvalg af børn ($N = 329$) i alderen 6:0–16:11 år.

Afhængig af delprøvens karakter er *reliabiliteten* blevet estimeret med test-retest metoden (split-half reliabilitet er ikke anvendelig til beregning af reliabilitet i delprøver eller for test som ikke kan deles i to ligeværdige halvdele). Analyser viste at de digitale versioner af Kodning og Figursøgning udviste test-retest reliabilitet svarende til p/b-formatet. De gennemsnitlige korrigerede reliabilitetskoefficienter for Kodning var .75 og for Figursøgning .80.

Den *interne struktur* for interkorrelationer mellem delprøver og indekser udviste lignende mønstre af korrelationer som fandtes i p/b-versionen.

Konfirmatorisk faktoranalyse (CFA) for to modeller blev gennemført; den første med Kodning og Figursøgning over for deres digitale modstykker (Model 1), den anden med kun p/b-versionerne af samtlige delprøver (Model 2). Statistiske tilpasningsmål viste meget god tilpasning for begge modeller, hvilket giver stærk støtte til at Kodning og Figursøgning måler lignende konstrukter uanset version eller administreringsmåde.

Analyse af ækvivalensen mellem p/b-versioner og digitale versioner viste en standardforskel mellem versionerne på $d = .10$ for Kodning og $d = -.13$ for Figursøgning. Disse effektstørrelser falder under det, af Pearson, opsatte kriterium med en effektstørrelse på maks $d = +/- .20$ for acceptabel ækvivalens. Til sammenligning indebærer en effektstørrelse på $d = .20$ godt $\frac{1}{2}$ skalascore på en skala med gennemsnittet 10 og en standardafvigelse på 3.

Undersøgelser af Kodning og Figursøgning i særlige grupper: For yderligere at afprøve validiteten og den kliniske anvendelighed sammenlignedes (ikke tilfældige) udvalg af forskellige kliniske grupper og matchede kontrolgrupper, når Kodning og Figursøgning, ligesom øvrige delprøver, blev administreret digitalt. Blandt de kliniske grupper var blandt andet særligt begavede børn, børn med Intellektuel funktionsnedsættelse af lettere grad og børn med specifikke indlæringsvanskeligheder. Resultatet fra sammenligningerne mellem disse kliniske grupper og matchede kontrolgrupper viste konsistente resultater i forhold til tidligere forskning i kliniske grupper.

Undersøgelse 3 og 4: WISC-V på Q-interactive i særlige grupper

To undersøgelser blev gennemført for at undersøge ækvivalensen mellem WISC-V p/b-administrering og administrering på Q-interactive i særlige grupper. Den første omfattede børn med Intellektuel funktionsnedsættelse af lettere grad og særligt begavede børn, den anden børn med ADHD og børn med autisme. Nedenfor findes korte sammenfatninger af resultaterne fra de forskellige grupper. Delprøverne Figursøgning og Kodning indgik ikke i disse undersøgelser, da disse delprøver ved tidspunktet for undersøgelsen ikke fandtes i digitale versioner. Derfor kan resultatet på HIK heller ikke præsenteres.

Særligt begavede børn

Totalt deltog 24 børn, som var blevet identificeret som særligt begavede, i undersøgelsen; 45.8 % piger, 54.2 % drenge. Gennemsnitsalderen var 11.5 år. En matchet kontrolgruppe blev udvalgt fra undersøgelse 1.

Gruppen af særligt begavede børn præsterede signifikant højere end den matchede kontrolgruppe på samtlige primære og sekundære indekser. Gennemsnit for de primære indekser varierede mellem 121.3 (RSI) og 127.6 (VFI). Gennemsnittet for GFI var 127.1. Effektstørrelserne for samtlige primære indekser var store. Gennemsnit for de sekundære indekser varierede mellem 123.9 (KVI) og 127.1 (GFI). Effektstørrelserne for samtlige sekundære indekser var store. Gennemsnittet for GFI, for denne gruppe af børn, var, i standardiseringen af p/b-versionen identisk, 127.1. Videre analyser viste at 75 % af de børn som var identificeret som særligt begavede opnåede en GFI-værdi på 120 eller højere, sammenlignet med kun 13 % af børnene i kontrolgruppen, hvilket ligger på linje med tilsvarende undersøgelser gennemført under standardiseringen af p/b-versionen.

Gennemsnittet for samtlige delprøver, med undtagelse af Matricer, var signifikant højere i gruppen af særligt begavede børn sammenlignet med den matchede kontrolgruppe. Alle sammenligninger på delprøveniveau viste moderate til store effektstørrelser, højest for Ordforråd, Ligheder, Information og Regning, hvilket er på linje med resultater fra standardiseringen af p/b-versionen af WISC-V.

Børn med Intellectuel funktionsnedsættelse af lettere grad

Totalt deltog 22 børn, som var diagnosticeret med Intellectuel funktionsnedsættelse af lettere grad, i undersøgelsen; 36.4 % piger, 63.6 % drenge. Gennemsnitsalderen var 11.8 år. En matchet kontrolgruppe blev udvalgt fra undersøgelse 1.

Gruppen af børn med Intellectuel funktionsnedsættelse af lettere grad præsterede signifikant lavere end den matchede kontrolgruppe på samtlige primære og sekundære indekser. Gennemsnit for de primære indekser varierede mellem 62.3 (VFI) og 67.9 (RSI). Gennemsnittet for GFI var 63.7. Effektstørrelserne for samtlige primære indekser var store. Gennemsnit for de sekundære indekser varierede mellem 61.6 (AAI) og 64.7 (KVI). Effektstørrelserne for samtlige sekundære indekser var store. Gennemsnittet for GFI, for denne gruppe af børn, var, i standardiseringen af p/b-versionen, 63.5. Videre analyser viste at 91 % af de børn som var diagnosticeret med Intellectuel funktionsnedsættelse af lettere grad opnåede en GFI-værdi på 75 eller højere, sammenlignet med kun 5 % af børnene i kontrolgruppen.

Gennemsnittet for samtlige delprøver var lavere for gruppen af børn med Intellectuel funktionsnedsættelse af lettere grad sammenlignet med den matchede kontrolgruppe. De største effektstørrelser fandtes for delprøverne Tal-bogstav rangordning, Talspændvidde, Regning, Information, Figurvægte, Verbal ræsonnering, Visuel genkendelse og Information.

Børn med ADHD

Totalt deltog 25 børn, som var diagnosticeret med ADHD, i undersøgelsen; 36 % piger, 64 % drenge. Gennemsnitsalderen var 10.5 år. En matchet kontrolgruppe blev udvalgt fra undersøgelse 1.

Sammenlignet med gruppen af børn med ADHD, som indgik i standardiseringen af p/b-versionen, havde børnene i Q-interactive-gruppen en yngre gennemsnitsalder og højere uddannede forældre. Gruppernes etniske sammensætning adskilte sig også. En vurdering af symptomernes sværhedsgrad blev ikke foretaget mellem grupperne. Sammenlagt er det, på baggrund af dette, ikke overraskende at de to ADHD-grupper resultater adskiller sig. De målte forskelles retninger (fra deres matchede kontrolgrupper) er imidlertid de samme, og de gennemsnitlige resultater på samtlige indekser og delprøver, ligesom effektstørrelserne, er sammenlignelige. De forskelle, som kunne ses mellem den ADHD-gruppe som gennemførte p/b-versionen og den ADHD-gruppe som gennemførte Q-interactive-versionen, kan i hovedtræk henføres til gruppernes forskellige baggrundsvariable og graden af symptomer.

Gruppen af børn med ADHD præsterede gennemsnitligt på samtlige primære delprøver, dog målt en signifikant forskel fra den matchede kontrolgruppe med hensyn til RSI. RSI udviste en moderat effektstørrelse – for VFI, VSI og AHI var effektstørrelserne små. På de sekundære indekser præsterede ADHD-gruppen signifikant lavere på AAI og GFI – effektstørrelserne var moderate. Resultaterne indikerer vanskeligheder med visuo-spatial bearbejdning, fluid reasoning og auditiv arbejdshukommelse.

Gennemsnittet for to delprøver, Matricer og Tal-bogstav rangordning, var signifikant lavere i gruppen af børn med ADHD sammenlignet med den matchede kontrolgruppe. Effektstørrelsen for Matricer var stor. Moderate effektstørrelser blev fundet for Blokmønstre, Regning og Tal-bogstav rangordning.

Børn med autisme

Totalt deltog 30 børn, som var diagnosticeret med autisme med nedsat sproglig evne, i undersøgelsen; 10 % piger, 90 % drenge. Gennemsnitsalderen var 11.6 år. En matchet kontrolgruppe blev udvalgt fra undersøgelse 1.

Gruppen af børn med autisme med nedsat sproglig evne præsterede signifikant lavere end kontrolgruppen på samtlige primære indekser og effektstørrelserne var gennemgående store. I overensstemmelse med tidligere forskning var RSI og VSI relativt høje og udviste mindre effektstørrelser end VFI. VFI havde den næststørste effektstørrelse af de primære indekser efter AHI. På de sekundære indekser præsterede gruppen af børn med autisme lavest på AAI, hvilket også havde den største effektstørrelse.

Gruppen af børn med autisme med nedsat sproglig evne præsterede signifikant lavere på samtlige primære og sekundære delprøver sammenlignet med kontrolgruppen. På linje med tidligere forskning blev det højes-

te gennemsnit opnået inden for det verbale indeks for delprøven Ligheder, og lavest for Verbal ræsonnering. De største effektstørrelser sås for Verbal ræsonnering, Tal-bogstav rangordning, Ordforråd, Visuel genkendelse, Regning og Talspændvidde. De mindste effektstørrelser blev målt for Figurvægte, Matricer, Visuelle puslespil og Blokmønstre.

Resultaterne for både indekser og delprøver var meget lig dem i tilsvarende undersøgelser ved standardiseringen af p/b-versionen.

Sammenfatning

De undersøgelser som hidtil er blevet gennemført med det formål at sikre ækvivalensen mellem p/b-versionen af WISC-V og den digitale version på Q-interactive, har vist små effekter af administreringsformat – både i normalpopulation og i særlige grupper. I normalpopulation har ingen effektstørrelser overskredet $d = .20$ i nogen indekser eller delprøver. Ingen baggrundsvariabler – eller præstationsniveau – har vist sig at have sammenhæng med effekt af administreringsformat. I de særlige grupper som blev undersøgt, genererer administrering af WISC-V på Q-interactive ensartede præstationsniveauer og mønstre, som i de undersøgelser der blev gennemført under standardiseringen af p/b-versionen. Den revidering, der var nødvendig for at digitalisere delprøverne Kodning og Figursøgning, har resulteret i reliable og valide digitale delprøver, og den skalaækvivalering som blev gennemført tillader anvendelse af p/b-normer for disse delprøver. Skalaækvivaleringen sker automatisk i Q-interactive og genererer den korrekte skalascore for en given råscore.

Ækvivalens mellem p/b-versioner og digitale versioner er vigtig så længe en test kan administreres på begge måder. Fremtiden må vise hvorvidt p/b-versioner af test bliver forældede – allerede nu foregår der et standardiseringsprojekt som vil udmunde i at kun digitale versioner publiceres.

20-10-2016