

WISC-V ekvivalensstudier

I USA har ett antal studier genomförts i syfte att undersöka ekvivalensen mellan papper och penna-versionen av WISC-V, och den digitala version som administreras på Q-interactive. Dessa studier har haft en rad olika syften:

- Att säkerställa att de två administreringsätten genererar jämförbara resultat – det vill säga att samma uppsättning normer kan användas för såväl papper och penna-versionen som den digitala versionen.
- Att säkerställa att de digitala versionerna av Kodning och Symbolletning är likvärdiga med de papper och penna-administrerade motsvarigheterna, trots de förändringar av uppgifter och gränssnitt som genomförts i de digitala versionerna.
- Att säkerställa att administrering av WISC-V på Q-interactive producerar likvärdiga resultat som papper och penna-versionen i särskilda grupper.

Nedan följer en kort sammanfattning av dessa studier. Den som vill ta del av studierna i sin helhet (inkluderande referenser till andra forskningsstudier) hänvisas till följande rapporter, vilka samtliga återfinns 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*.

Studie 1: Ekvivalens mellan WISC-V papper och penna-version och WISC-V Q-interactive

Ekvivalensstudien genomfördes i samband med det amerikanska standardiseringsarbetet, vilket även omfattade administrering av WISC-V på Q-interactive. 350 barn i 175 par, matchade med avseende på kön, ålder, föräldrars utbildningsnivå och etnicitet, valdes för analys efter att datainsamlingsarbetet avslutats.

Analyserna visade att endast 3 av de 18 deltesten (den amerikanska versionen av WISC-V har fler deltest än de icke engelskspråkiga) uppvisade en signifikant effekt av format 5 %-nivån; Blockmönster, Förståelse och Aritmetik. Ingen av effektstorlekarna överskred emellertid det uppsatta kriteriet på $d = 0.2$. Mönstret över effektstorlekar var inte konsekvent med de som uppmätts för motsvarande deltest i WISC-IV. De två deltesten med störst effektstorlekar i WISC-V, Blockmönster och Förståelse, uppvisade i WISC-IV-studien effektstorlekar nära 0. De två deltest som uppvisade större effektstorlekar än $d = 0.2$ i WISC-IV, hade måttliga (Matriser) eller låga (Bildkategorier) effektstorlekar i WISC-V.

En serie analyser genomfördes också för att utröna huruvida prestationsnivå eller bakgrundsvariabler (kön, ålder, föräldrars utbildningsnivå och etnicitet) hade samband med effekter av administreringsformat. Av totalt 90 genomförda analyser, föll endast två ut med signifikanta resultat; båda rörde deltestet Naming Speed. Detta deltest finns inte med i WISC-V-versioner utanför USA.

Studie 2: Kodning och Symbolletning

I motsats till övriga deltest, är Kodning och Symbolletning på Q-interactive inte parallella med papper och penna-versionerna. Anpassning av dessa deltest till ett digitalt format har inneburit ett nytt gränssnitt mellan uppgifterna och barnet. Exempelvis markeras nu svar genom att barnet trycker på iPad, stimuli presenteras automatiskt efter att barnet svarat, kravet på att barnet ska skriva har eliminerats, och det finns inte längre finns möjlighet till självkorrigering av svar.

De digitala versionerna av dessa deltest uppvisar inte råpoängsekvivalens i förhållande till pappersversionerna, varför en statistisk konverteringsprocess, skalekvivalering (*eng equating*) genomförts i syfte att anpassa råpoängen från de digitala versionerna så att de blir likvärdiga råpoängen från papper och penna-versionerna; det vill säga att transformering av råpoäng till skalpoäng indikerar samma grad av prestation för de både olika administreringsformaten, vilket i sin tur innebär att samma uppsättning normer kan användas.

En sådan skalekvivalering kan justera för skillnader i svårighetsgrad mellan test, men inte för skillnader i innehåll. Ett första nödvändigt steg var därför att säkerställa att de digitala versionerna av Kodning och Symbolletning uppvisade tillräckligt hög likhet avseende konstrukt (innehåll och mätområde) och psyko-metriska egenskaper som motsvarande papper och penna-versioner. En serie stegvisa anpassningar och justeringar av de digitala versionerna av Kodning och Symbolletning genomfördes, varefter data insamlades i ett representativt urval av barn ($N = 329$) i åldern 6:0–16:11 år.

Beroende på deltestens karaktär har *reliabiliteten* uppskattats med test-retest metoden (split-half reliabilitet är inte lämplig att beräkna för test som inte går att dela i två likvärdiga halvor). Analyser visade att de digitala versionerna av Kodning och Symbolletning uppvisade test-retest reliabilitet motsvarande den för pappersformatet. De genomsnittliga korrigerade reliabilitetskoefficienterna för Kodning var 0.75 och för Symbolletning 0.80.

Den *interna strukturen* för interkorrelationer mellan deltest och index uppvisade liknande mönster av korrelationer som återfanns i papper och penna-versionen.

Konfirmatorisk faktoranalys (CFA) för två modeller genomfördes; den första med Kodning och Symbolletning utbytta till digitala motsvarigheter (Modell 1), den andra med enbart pappersversionerna av samtliga (Modell 2). Statistiska anpassningsmått visade på samma utmärkta anpassning för båda modellerna vilket ger ett starkt stöd för att Kodning och Symbolletning mäter liknande konstrukt oavsett version/administreringssätt.

Analys av ekvivalensen mellan pappersversioner och digitala versioner visade på en standardskillnad mellan versionerna på $d = .10$ för Kodning och $d = -.13$ för Symbolletning. Dessa effektstorlekar faller under det av Pearson uppsatta kriteriet med en effektstorlek på maximalt $d = +/- .20$ för acceptabel ekvivalens. Som en jämförelse innebär en effektstorlek på $d = .20$ drygt en halv skalpoäng på en skala med medelvärdet 10 och standardavvikelsen 3.

Studier av Kodning och Symbolletning i särskilda grupper: För att ytterligare pröva validiteten och den kliniska användbarheten jämfördes (icke slumpmässiga) urval av olika kliniska grupper mot matchade kontrollgrupper, när Kodning och Symbolletning liksom övriga deltest gavs i digital version. Bland de kliniska grupperna återfanns bland andra särskilt begåvade barn, barn med lindrig intellektuell funktionsnedsättning och barn med specifika inlärningssvårigheter. Resultatet från jämförelserna mellan dessa kliniska grupper och matchade kontrollgrupper visar på konsistenta resultat i förhållande till tidigare forskning på kliniska grupper.

Studie 3 och 4: WISC-V på Q-interactive i särskilda grupper

Två studier har genomförts för att undersöka ekvivalensen mellan WISC-V papper och penna-administrering och administrering på Q-interactive i särskilda grupper. Den första omfattade barn med lindrig intellektuell funktionsnedsättning samt särskilt begåvade barn, den andra barn med ADHD samt barn med autism. Nedan återfinns korta sammanfattningar av resultat från dessa olika grupper. Symbolletning och Kodning ingick inte i dessa studier, då dessa deltest vid tidpunkten för studierna inte fanns i digitala versioner, varför resultat avseende HIK ej kan presenteras.

Särskilt begåvade barn

Totalt 24 barn som identifierats som särskilt begåvade deltog i studien; 45.8 % flickor, 54.2 % pojkar. Genomsnittsåldern var 11.5 år. En matchad kontrollgrupp drogs från studie 1.

Gruppen särskilt begåvade presterade signifikant högre än den matchade kontrollgruppen på samtliga primära och sekundära index. Medelvärdet för de primära indexpoängen varierade mellan 121.3 (FI) och 127.6 (VI). Medelvärdet för KRI var 127.1. Effektstorlekarna för samtliga primära index var stora. Medelvärdet för de sekundära indexen varierade mellan 123.9 (QI) till 127.1 (KRI). Effektstorlekarna för samtliga sekundära index var stora. Medelvärdet för GAI i denna grupp barn var i standardiseringen av papper och penna-versionen identiskt, 127.1. Vidare analyser visade att 75 % av de barn som identifierats som särskilt begåvade erhölet ett

KRI-värde på 120 poäng eller högre, jämfört med endast 13 % av barnen i kontrollgruppen, vilket ligger i linje med motsvarande studie genomförd under standardiseringen av papper och penna-versionen.

Medelvärdet för samtliga deltest, med undantag av Matriser, var signifikant högre i gruppen särskilt begåvade barn jämfört med den matchade kontrollgruppen. Alla jämförelser på deltestnivå visade på måttliga till stora effektstorlekar, högst för Ordförråd, Likheter, Information och Aritmetik, vilket är i linje med resultaten från standardiseringen av papper och penna-versionen av WISC-V.

Barn med lindrig intellektuell funktionsnedsättning

Totalt 22 barn som diagnosticerats med lindrig intellektuell funktionsnedsättning deltog i studien; 36.4 % flickor, 63.6 % pojkar. Genomsnittsåldern var 11.8 år. En matchad kontrollgrupp drogs från studie 1.

Gruppen barn med lindrig intellektuell funktionsnedsättning presterade signifikant lägre än den matchade kontrollgruppen på samtliga primära och sekundära index. Medelvärdet för de primära indexpoängen varierade mellan 62.3 (VI) och 67.9 (FI). Medelvärdet för KRI var 63.7. Effektstorlekarna för samtliga primära index var stora. Medelvärdet för de sekundära indexen varierade mellan 61.6 (AAI) och 64.7 (QI). Effektstorlekarna för samtliga sekundära index var stora. Medelvärdet för KRI i denna grupp barn var i standardiseringen av papper och penna-versionen 63.5. Vidare analyser visade att 91 % av de barn som diagnosticerats med lindrig intellektuell funktionsnedsättning erhöll ett KRI-värde på 75 poäng eller högre, jämfört med endast 5 % av barnen i kontrollgruppen.

Medelvärdet för samtliga deltest var lägre i gruppen barn med lindrig intellektuell funktionsnedsättning jämfört med den matchade kontrollgruppen. De största effektstorlekarna återfanns för deltesten Bokstavs-siffer-serier, Sifferrepetition, Aritmetik, Information, Figurvikter, Förståelse, Bildminne och Information.

Barn med ADHD

Totalt 25 barn som diagnosticerats med ADHD deltog i studien; 36 % flickor, 64 % pojkar. Genomsnittsåldern var 10.5 år. En matchad kontrollgrupp drogs från studie 1.

Jämfört med gruppen barn med ADHD som ingick i standardiseringen av papper och penna-versionen, hade barnen i Q-interactive-gruppen en yngre genomsnittsålder och högre utbildade föräldrar. Gruppernas etniska sammansättning skiljde sig också åt. Någon bedömning av symptomens svårighetsgrad gjordes inte mellan grupperna. Sammantaget är det mot bakgrund av detta inte överraskande att de två ADHD-gruppernas resultat skiljer sig åt. De uppmätta skillnadernas (mot respektive matchad kontrollgrupp) riktningar är emellertid desamma, och de genomsnittliga resultaten på samtliga index och deltest, liksom effektstorlekarna, är likartade. De skillnader som påvisats mellan den ADHD-grupp som genomförde papper och penna-versionen och den ADHD-grupp som genomförde Q-interactive-versionen kan i huvudsak hänföras till gruppernas olika bakgrundsvariabler och grad av symptom.

Gruppen barn med ADHD presterade genomsnittligt på samtliga primära deltest, dock uppmättes en signifikant skillnad mot den matchade kontrollgruppen med avseende på FI. FI uppvisar en måttlig effektstorlek - för VI, VSI och AI är effektstorlekarna små. Beträffande de sekundära indexen presterar ADHD-gruppen signifikant lägre på AAI och KRI - effektstorlekarna är måttliga. Dessa resultat indikerar svårigheter med visuospatial bearbetning, fluid reasoning och auditivt arbetsminne.

Medelvärdet för två deltest var signifikant lägre i gruppen barn med ADHD jämfört med den matchade kontrollgruppen; Matriser och Bokstavs-siffer-serier. Effektstorleken för Matriser är hög. Måttliga effektstorlekar återfanns för Blockmönster, Aritmetik och Bokstavs-siffer-serier.

Barn med autism

Totalt 30 barn som diagnosticerats med autism med nedsatt språklig förmåga deltog i studien; 10 % flickor, 90 % pojkar. Genomsnittsåldern var 11.6 år. En matchad kontrollgrupp drogs från studie 1 refererad ovan.

Gruppen barn med autism med nedsatt språklig förmåga presterade signifikant lägre än kontrollgruppen på samtliga primära index; effektstorlekarna var genomgående stora. I samklang med tidigare forskning var FI och VSI relativt högre och uppvisade mindre effektstorlekar än VI. VI hade den näst största effektstorleken av de primära indexen efter AI. Avseende de sekundära indexen presterade gruppen barn med autism lägst på AAI, vilket också hade den största effektstorleken.

Gruppen barn med autism med nedsatt språklig förmåga presterade signifikant lägre på samtliga primära och sekundära deltest jämfört med kontrollgruppen. I linje med tidigare forskning uppnåddes det högsta

medelvärde inom Verbalt index för Likheter, och lägst för Förståelse. De största effektstorlekarna noterades för Förståelse, Bokstavs-siffer-serier, Ordförråd, Bildminne, Aritmetik och Sifferrepetition. De minsta effektstorlekarna uppmättes för Figurvikter, Matriser, Visuella pussel och Blockmönster.

Resultaten för både index och deltest är mycket likartade de som framkom för motsvarande studie vid standardiseringen av papper och penna-versionen.

Sammanfattning

De studier som hittills genomförts i syfte att säkerställa ekvivalensen mellan papper och penna-versionen av WISC-V och den digitala versionen på Q-interactive, har visat på små effekter av administreringsformat – både i normalpopulation och i särskilda grupper. I normalpopulation har inga effektstorlekar överskridande $d = .20$ uppmätts för något index eller deltest. Inga bakgrundsvariabler – eller prestationsnivå – har påvisats ha samband med effekt av administreringsformat. I de särskilda grupper som undersökts, genererar administrering av WISC-V på Q-interactive likartade prestationsnivåer och mönster som i de studier som genomfördes under standardiseringen av papper och penna-versionen. Den revidering som krävts för att digitalisera deltesten Kodning och Symbolletning har resulterat i reliabla och valida digitala deltest, och en skalekvivalering har genomförts som tillåter användning av papper och penna-normer för dessa deltest. Denna skalekvivalering sker automatiskt på i Q-interactive och genererar den korrekta skalpoängen för en viss råpoäng.

Ekvivalens mellan papper och penna-versioner och digitala versioner är viktigt att påvisa så länge ett test kommer att kunna administreras på båda sätten. Framtiden får utvisa huruvida pappersversioner av test blir obsoleta – redan nu pågår standardiseringsprojekt som kommer att utmyнна i att endast digitala versioner publiceras.

2016-10-20