



Måleegenskaper ved den norske versjonen av Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V)

Per Håkan Brøndbo¹ (per.h.brondbo@uit.no)

Jens Egeland^{2,3} (jens.egeland@siv.no)

PsykTestBarn 2019, 1:5

ISSN 1893-9910

Mottatt dato 16. september 2019

Publisert dato 16. desember 2019

DOI 10.21337/0064

Som alle artikler i PsykTestBarn, kan denne fagfellevurderte artikkelen arkiveres og distribueres fritt for alle slags formål på følgende vilkår: korrekt referanse skal oppgis (se under), ingen kommersiell bruk og ingen bearbeidelse av tekst eller innhold.

Denne artikkelen skal siteres på følgende måte:

Brøndbo, P. H. & Egeland, J. (2019). Måleegenskaper ved den norske versjonen av Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V). PsykTestBarn, 1:5. doi: 10.21337/0064.

¹ UiT Norges arktiske universitet, RKBU Nord

² Sykehuset i Vestfold HF, Klinikk Psykisk Helse og Rusbehandling, ³ Universitetet i Oslo, Psykologisk Institutt

Sammendrag

Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) ble publisert i USA i 2014 og den norske versjonen kom i 2017. Norsk og internasjonal rettighetshaver er Pearson Assessment. Oversettelsesarbeidet, inkludert kulturell tilpasning, ble gjennomført av Pearson Assessment med hjelp fra klinisk ekspertise i Norge, Danmark og Sverige. WISC-V er en individuelt administrert test av intellektuelle evner hos barn og ungdom i alderen 6:0-16:11 år. Testen består av 15 deltester, ti primære og fem sekundære deltester. Testtiden er beregnet til 45–80 minutter, avhengig av antall deltester og barnets alder. Psykologer og fagpersoner med mastergrad i pedagogikk med fordypning i pedagogisk-psykologisk rådgivning eller spesialpedagogikk fra UiO er kvalifisert til å kjøpe testen. For personer med mastergrad i pedagogikk eller psykologi fra andre læresteder er det et eget sertifiseringsprogram.

Søket førte til 1119 norske, respektive 1865 svenske og danske treff. Ingen av treffene var relevante. Manualens skandinaviske normering er det eneste grunnlaget for vurdering av testens psykometriske egenskaper. Normer for aldersgrupper, validitet og reliabilitet, samt gjennomsnittsskårer for barn med lett intellektuell funksjonsnedsettelse og for særlig evnerike barn, ble rapportert i manualen.

Normeringsutvalget bestod av 660 barn fra 6:0 til 16:11. Resultatene ble presentert som skalerte skårer fra 1 til 19, med et gjennomsnitt på ti og et standardavvik på tre. Indeks-skårer ble konstruert med et gjennomsnitt på 100, et standardavvik på 15 og et skåreintervall på 45–155, respektive 40–160. Reliabilitetskoeffisienter (indre konsistens) på deltestnivå og indeksnivå var fra gode til utmerkede. Goodness-of-fit-analyse i forbindelse med konfirmatorisk faktoranalyse ga CFI 0,97-0,98 og RMSEA 0,05 for valgt femfaktormodell. Mønsteret av interkorrelasjoner mellom deltest, prosessskår og indeks viser store likheter med resultatene fra den amerikanske standardiseringen. Korrigerte rangkorrelasjonene for FSIQ og på indeksnivå mellom WISC-V og WISC-IV er høye (0,67-0,91). Kriterierelatert validitet som predikerer både intellektuell funksjonsnedsettelse og særlige evner var dårligere enn forventet og spesielt for gruppen særlig evnerike barn.

Det er behov for mer dokumentasjon av måleegenskapene i en klinisk skandinavisk kontekst. Basert på dokumentasjonen som finnes fremstår WISC-V i hovedsak som et reliabelt instrument for vurdering av evnenivå og kognitiv profil. Foreliggende data er utilstrekkelige til å vurdere hvorvidt WISC-V også er et valid instrument for vurdering av kognitiv profil. Klinisk profiltolkning bør dermed gjøres med forsiktighet.

Abstract

Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) was published in the USA in 2014, and a Norwegian version was published in 2017. The Norwegian and international owner is Pearson Assessment. Translation, including cultural adaptation, was performed by Pearson Assessment, with assistance from clinical expertise in Norway, Denmark and Sweden. WISC-V is an individually administered test of intellectual abilities for children and youth aged 6:0-16:11 years. The test consists of 15 subtests, where ten are primary and five are secondary tests. Estimated test time is 45-80 minutes, dependent on the number of subtests and the age of the child. Psychologists or professionals with a master's degree in education from the University of Oslo are qualified for purchasing the test, and there is a licensing program available for professionals with a master's degree in education from other universities as well as for master's in psychology.

The search found 1119 Norwegian, and 1865 Swedish and Danish hits. None of them relevant. The Scandinavian norms from the manual is the only available material for the evaluation of the psychometrics. Age-group norms, validity and reliability, in addition to mean scores for gifted children, and children with minor learning disabilities, were reported.

Norm data from 660 children 6:0 until 16:11 were reported. Results were presented as scaled scores ranged 1 to 19, with a mean of 10 and a standard deviation of 3. Index scores were constructed with a mean of 100, standard deviation of 15, and range 45-155, respectively 40-160. Internal consistency for subtests and indexes were good to excellent. Goodness-of-fit-analyses for confirmatory factor analyzes gave CFI 0,97-0,98 and RMSEA 0,05 for the chosen five-factor model. The pattern of inter-correlations between sub-tests, process scores and indexes were mostly similar to the US standardization. Corrected rank correlation for FSIQ and indexes between WISC-V and WISC-IV are high (0,67-0,91). Criterion validity for predicting minor learning disability and giftedness were poorer than expected, especially for identifying gifted children.

More documentation of the psychometrics in a clinical Scandinavian context is needed. Based on the available documentation, WISC-V mainly seems as a reliable instrument of intelligence assessment and assessment of cognitive profiles. It is not enough available data to evaluate if WISC-V also is a valid instrument, especially for assessment of cognitive profiles. Clinical interpretation of cognitive profiles should therefore be done with caution.

Innledning

Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) (Wechsler, 2014a), er den sist reviderte og oppdaterte versjonen av Wechsler Intelligence Scale for Children (Wechsler, 1949). WISC-V ble publisert i USA i 2014 og den norske versjonen ble publisert i 2017 (Wechsler, 2017). Norsk og internasjonal rettighetshaver er Pearson Assessment. Oversettelsesarbeidet, inkludert kulturell tilpasning, ble gjennomført av en prosjektgruppe som besto av prosjektleder fra Pearson Assessment og klinisk ekspertise fra Norge, Danmark og Sverige. Oversettelsesarbeidet har etterstrebet å beholde oppgavene i originalversjon der det har vært mulig, parallelt med å vurdere nødvendige endringer av språklig og kulturell art for at testen skal fungere etter intensjonen i en skandinavisk sammenheng.

WISC-V er en individuelt administrert test av intellektuelle evner hos barn og ungdom i alderen 6:0-16:11 år. En vesentlig revidering fra WISC-IV (Wechsler, 2009) er gjort på bakgrunn av forskning og klinisk erfaring. Skalaens faktorstruktur er endret fra en firefaktor- til en femfaktormodell. Nye deltester og prosessmål er inkludert for å utvide skalaens kliniske bruksområde. Deltester og prosessmålene gjør tilgjengelig informasjon om prosesser som bidrar til et barns prestasjoner. Testtiden er beregnet til 45–80 minutter, og avhenger av antall valgte deltester og barnets alder. Rettighetshaver beskriver følgende kompetansekrav for å kunne utføre testen: Enten må man være psykolog eller ha fullført masterprogrammet i spesialpedagogikk og masterprogrammet i pedagogikk med studieretning pedagogisk-psykologisk rådgivning fra UiO. For fagpersoner med mastergrad i pedagogikk fra andre læresteder eller mastergrad i psykologi finnes det et eget sertifiseringsprogram for å kunne bruke evnetester for barn (<https://www.pearsonassessment.no/produkter/utvikling-og-kognitive-evner/wisc-v.11.07.19>).

WISC-V er den første Wechsler-testen som er tilrettelagt for elektronisk administrering. To iPad'er benyttes under testen, en for testtaker og en for testadministrator.

WISC-V består av 15 deltester. Ti av dem betegnes som primære tester. Det er Terningmønster (Tm), Likheter (Li), Matriser (Mr), Tallhukommelse (Th), Koding (Kd), Ordforståelse (Of), Symbolleting, Figurvekter (Fv), Visuelle puslespill (Vp), Bildehukommelse (Bh). De tre sistnevnte testene er nye i norsk sammenheng. Alle Wechslers evnetester har blitt kritisert for at arbeidsminne bare har vært undersøkt i den auditive modaliteten (Egeland, 2015). Det fører til at man lett kan mistolke språklige vansker eller lese- og skrivevansker som indikasjon på et oppmerksomhetsproblem (Reinecke, Beebe, & Stein, 1999). Med Bh muliggjøres undersøkelse også av visuelt arbeidsminne.

Det er fem sekundære tester: Informasjon (In), Tallbokstav-serier (Tb), Utstrykning (Ut), Resonnering (Rs) og Regning (Rg). Det er nødvendig å administrere syv av de ti primære testene for å beregne fullskala IQ (FSIQ). Administrering av alle ti primærttester er imidlertid nødvendig for å generere de fem indeksskårene som gir den kognitive profilen klinikerer i første omgang vil bruke for å vurdere sterke og svake egenskaper hos barnet. Den Perseptuelle resonneringsindeksen i WISC-IV er delt i Visuospatial indeks (VSI) som består av Vp og Tm, og Flytende resonneringsindeks (FRI) som består av Mr og Fv. Femfaktorløsningen er empirisk avledet basert på faktoranalyse. Det er også mulig å beregne fem teoretisk baserte sekundære indeksskårer. Generell evneindeks (GAI) fra tidligere versjoner benevnes nå som Grunnleggende evneindeks (GEI). GEI er utviklet som et mål på resonnering og abstraksjon som er ikke påvirkes av tempo eller oppmerksomhet (Esteve & Egeland, 2010). Tempo og oppmerksomhetsprøvene utgjør Basal Prosesseringsindeksen BPI. De seks prøvene som ikke er språklige, utgjør en ny Nonverbal indeks (NVI). Basert på de to auditive arbeids-hukommelsesprøvene kan man beregne en separat Auditiv arbeidshukommelsesindeks (AAI). Rg og Fv utgjør Mengderesonneringsindeksen (MRI). I tillegg finnes prosessskårer, som skal gi ytterligere informasjon om de prosesser som bidrar til et barns prestasjoner på enkelte deltester.

I den amerikanske manualen WISC-V Technical and Interpretive Manual (Wechsler, 2014b), beskrives omfattende studier av reliabilitet. Manualen viser til høy reliabilitet for indekser, deltester og prosesskårer, både i normutvalget ($N = 2200$) og i særskilte kliniske grupper, høy intertaterreliabilitet, samt høy test-retest-reliabilitet over en periode på flere uker. I manualen (Wechsler, 2014b) beskrives også analyser av indre konsistens, faktoranalyser, korrelasjoner med andre tester, samt normeringsstudier i særskilte kliniske grupper.

Flere valideringsstudier i bl.a. USA (Canivez, Watkins, & Dombrowski, 2016, 2017; Dombrowski, Canivez, Watkins, & Alexander Beaujean, 2015), Kanada (Watkins, Dombrowski, & Canivez, 2018), Storbritannia (Canivez, Watkins, & McGill, 2019), Spania (Fenollar-Cortés & Watkins, 2018), og Frankrike (Lecerf & Canivez, 2018) har mislyktes med å replisere støtte for den femfaktormodell som er valgt i manualen for WISC-V (Wechsler, 2014b). På bakgrunn av dette har forfatterne til valideringsstudiene konkludert at det er behov for forsiktighet ved tolkning av resultater utover Fullskala-IQ.

Metode

Bibliotekar Mari Elvsåshagen og Sølvi Biedilæ ved Regionsenter for barn og unges psykiske helse, helseregion Øst og Sør, søkte etter dokumentasjon på testens psykometriske egenskaper i databasene, PsycINFO, Medline, Embase, Cochrane Library, Oria (BIBSYS), Norart, SveMed+, PubMed, CRISStin.no, NORA.no, Forskningsdatabasen.dk og Swepub. Søkedato: 20.09.2018. Søkestrategien er tilgjengelig på <https://psyktestbarn.r-bup.no/no/artikler/wisc--wechsler-intelligence-scale-for-children>

Inklusjonskriterium var publikasjoner av studier som har undersøkt og rapportert minst ett av følgende i skandinaviske utvalg:

- normdata for testen
- reliabilitet: indre konsistens, test-retest, intertater og endringssensitivitet
- validitet: samsvar med liknende testskårer, samsvar med referansestandard eller annet kriterium, og/eller faktorstruktur

I tillegg, og *kun for norske versjoner av WISC-V*, publikasjoner som rapporterte gjennomsnittsskårer og/eller forekomster for henholdsvis generelle populasjoner og kliniske undergrupper.

Etter fjerning av dubletter gikk begge forfatterne gjennom alle identifiserte publikasjoners sammen- drag. Forfatterne foretok vurderingene uavhengig av hverandre. Alle publikasjoner som kunne virke relevante ble bestilt inn i fulltekst, og vurderingsprosessen ble gjentatt for disse.

Begge forfatterne vurderte normering, validitet og reliabilitet ved hjelp av en tilpasset versjon av Test review form and notes for reviewers (European Federation of Psychologists' Associations, 2013). Forfatterne foretok vurderingene uavhengig av hverandre.

Resultater

Litteratursøk

Det systematiske søket etter dokumentasjon for måleegenskapene førte til 1119 norske, respektive 1865 svenske og danske treff. Ingen av treffene var relevante. Mesteparten av treffene ble sortert bort grunnet publikasjonsdato før utgivelse av WISC-V og resterende basert på abstrakt. Ingen artikler ble innhentet i fulltekst.

Det er gjort en skandinavisk normering (svensk-dansk-norsk) som rapporteres i manualen for WISC-V (Wechsler, 2017). Data fra manualen er det eneste grunnlaget som foreligger for vår vurdering av testens psykometriske egenskaper. Normeringsutvalgene i Norge, Danmark og Sverige ($N = 660$) ble trukket for å gjenspeile normalbefolkningen i hvert land. Metoden som ble benyttet for å oppnå et slikt utvalg var ved å velge barn på bakgrunn av foreldrenes utdanningsnivå. Hvert land har bidratt med en tredjedel av utvalget. Det ble rapportert normer for aldersgrupper, validitet (begrepsvaliditet og kriterierelatert validitet) og reliabilitet (indre konsistens), samt gjennomsnittsskårer for en gruppe særlig evnerike barn og en gruppe barn med lett intellektuell funksjonsnedsettelse.

Tabell 1. Inkluderte studier

Referanse	Design	Populasjon	N	Mål	Rapporterte egenskaper, verdier
Wechsler (2017)	Tverrsnitt	Normalpopulasjon i Norge, Danmark og Sverige	660	IQ, indeksskårer og deltestskårer	Middelverdier, normer, validitet (type), reliabilitet (type)]

Middelverdier og/eller forekomster i kliniske undergrupper

Tabell 2a. Middelverdier (*M*) og standardavvik (*SD*) for ulike undergrupper

Referanse	Utvalg/gruppe	N/n	Fullskala IQ		Verbal forståelsesindeks		Flytende resonneringsindeks	
			M	SD	M	SD	M	SD
Wechsler (2017)	Særlig evnerike barn	22	120.1	8.5	113.8	13.5	121.0	12.4
Wechsler (2017)	Lett intellektuell funksjonsnedsettelse	22	63.0	12.3	66.8	16.1	70.7	10.3

Normer

Normdata ble samlet inn i perioden juni 2015 til mars 2016. Innsamlingen foregikk parallelt i Norge, Sverige og Danmark og ble publisert i WISC-V manual, del 1 Norsk versjon (Wechsler, 2017). Normeringsarbeidet ble gjennomført med hjelp av ca. 50 psykologer og psykologistudenter. Disse gjennomførte testingen av barna.

Vanskelighetsgraden av leddene i de fire språklige delprøvene Li, Of, In og Rs ble testet ut med såkalte «overitems» blant 30 barn i hver av tre aldersgrupper. Ti barn fra hvert land ble testet i hver gruppe. Basert på denne utprøvingen valgte man ut de ledd som fungerte best og rangerte dem utfra vanskelighetsgrad.

Selve normeringsutvalget bestod av 660 barn fordelt på elleve aldersintervaller fra 6:0 til 16:11, hvor hver aldersgruppe inkluderte 60 barn med jevn kjønnsfordeling og jevn fordeling mellom Norge, Sverige og Danmark. Andel barn i utvalget som anga

bruk av annet språk hjemme enn nasjonalspråket var 13 %. Av norske barn oppga 16 % bruk av annet språk enn norsk i hjemmet, hvilket ble sammenliknet med at 13 % av Norges befolkning er født i utlandet (Statistisk sentralbyrå, 2014). I manualen er det presentert bakgrunnsvariabler for normeringsutvalget, samt populasjonsdata hentet fra de respektive lands statistiske sentralbyrå. Det er gjennomført khi-kvadrat-tester for å undersøke eventuelle avvik mellom utvalg og utvalgsplan, uten å finne signifikante forskjeller.

For å sikre representativitet i nedre og øvre del av fordelingen er en liten andel barn med lett intellektuell funksjonsnedsettelse ($N = 12$), samt særlig evnerike barn ($N = 15$) tatt med i normeringsutvalget.

Resultatene fra normeringsstudien presenteres i manualen og forfatterne skriver at «normene er konstruert ved hjelp av «inferential norming» (Wechsler, 2017, s. 33). Resultatene ble presentert som skalerte skårer fra 1 til 19, med et gjennomsnitt

på ti og et standardavvik på tre. Indeks-skårer ble konstruert ved hjelp av summering av skalerte skårer, samt bearbeiding som resulterte i standard-skårer med et gjennomsnitt på 100, et standardavvik på 15, og et skåreintervall på 45–155 (VFI, VSI, FRI, AHI, PHI, MRI, AAI), respektive 40–160 (FSIQ, NVI, GEI, BPI).

I manualen er det normtabeller for konvertering av råskårer til skalerte skårer, konvertering av sum skalerte skårer til primære indekser og fullskala, sekundære indekser og prosessskårer, tabeller for å gjennomføre parvise sammenligninger for primære indekser og fullskala, sekundære indekser og prosessskårer, samt for ulike kombinasjoner av deltester presentert med kritiske verdier og frekvenser i normgruppene.

Reliabilitet

Indre konsistens ble undersøkt ved hjelp av split-half reliabilitet, med unntak av deltestene i PHI, da disse har tidsbegrensing og test-retest er en mer passende måte å undersøke reliabilitet på. Test-retest-koeffisienter for disse tre deltestene (Kd, Sl og Ut) er hentet fra den amerikanske standardiseringen.

Reliabilitetskoeffisientene på deltestnivå strekker seg fra gode til utmerkede (0,80 – 0,93), med Tm som lavest og Fv som høyest. Reliabilitetskoeffisientene for indeksene strekker seg fra 0,88 – 0,95, der PHI er lavest og FSIQ høyest. Det er også oppgitt reliabilitetskoeffisienter på deltestnivå for gruppene «særlig evnerike barn» og «barn med lett intellektuell funksjonsnedsettelse», som strekker seg fra 0,81– 0,97.

Det er ikke gjennomført test-retest-studier, eller studier av intervurderingsreliabilitet. Her vises det til det amerikanske standardiseringsarbeidet.

Stoppkriterier

I WISC-V avsluttes de språklige prøvene etter tre null-poengssvar. I den amerikanske originalversjonen viser det seg at rangkorrelasjonen er nærmest uendret mellom tre null-poengssvar og hvis man administrerte enda flere ledd før

avslutning. Hvis rangeringen etter vanskelighetsgrad av ledd i de språklige prøvene er for upresis, vil dette imidlertid kunne medføre at barn etter tre null-poengssvar kunne ha greid presumptivt vanskeligere ledd. Vi kan ikke se at en test av reliabilitet av stoppkriteriene er gjort i den skandinaviske utgaven av WISC-V.

Validitet

I manualen presenteres interkorrelasjoner mellom deltest, prosesskår og indeks. Resultatene oppsummeres slik i manualen: «*Samtlige korrelasjoner er statistisk signifikante, og alle deltester korrelerer i noen grad med hverandre. Prosessskårer korrelerer høyt med sine tilsvarende deltester. Mønsteret av interkorrelasjoner viser store likheter med det som kom frem ved den amerikanske standardiseringen (se tabell 5.1 i kapittel 5 i WISC-V Technical and Interpretive Manual (Wechsler, 2014b), samt i WISC-IV og andre Wechsler-skalaer*» (Wechsler, 2017, s. 47).

Det ble gjennomført konfirmatorisk faktoranalyse. Resultatene indikerte at det kreves minst fire faktorer for å oppnå god tilpasning. Det ble valgt en femfaktormodell for den skandinaviske versjonen av WISC-V, da dette er i samsvar med den amerikanske originalversjonen. Likevel viser den skandinaviske faktoranalysen antydningvis bedre samsvar for firefaktorløsningen fra de fjerde revisjonene av WISC og WAIS. Mellom de ulike fire- og femfaktorløsningene er det ulik plassering av deltesten Regning. I firefaktorløsningene er det best samsvar når Regning kan lade både på arbeids-hukommelse og flytende resonnering og for en løsning der deltesten får lade på disse to og verbalfaktoren i tillegg. Selv om firefaktorløsningene var antydningvis bedre er også samsvaret mellom den valgte femfaktorløsningen og observerte skårer gode. Goodness-of-fit-analyse for valgt modell ga CFI 0,97–0,98 og RMSEA 0.05.

Det ble også gjennomført korrelasjonsanalyser med WISC-IV, der både WISC-V og WISC-IV ble administrert for 95 barn (31 norske, 31 svenske og 33 danske). Alle barna fikk WISC-V først, noe som trolig medførte en viss læringseffekt, og resulterte at de fikk høyere scorer på WISC-IV og en større gjennomsnittlig skåreforskjell. Utvalget hadde 4.8

høyere fullskala IQ på WISC-IV. Det er ni år mellom de to normeringene og tidligere har man i den vestlige verden funnet en økning i intelligens (Flynn-effekt) svarende til 3–5 IQ poeng hvert ti-år (Lynn, 2009). Det har imidlertid vært reist spørsmål om veksten i intelligens nå har stoppet opp (Sundet, Barlaug, & Torjussen, 2004).

I manualen presenteres gjennomsnittsverdier, standardavvik, ukorrigerede og korrigerede korrelasjoner, samt standardforskjeller for deltest-, prosess-, og indeksskårer mellom WISC-V og WISC-IV. De korrigerede rangkorrelasjonene for FSIQ og på indekssnivå er høye (0,67–0,91) og forfatterne konkluderer slik: «Samlet tyder omfanget av korrelasjoner mellom WISC-V og WISC-IV på at de to instrumentene måler likeartede konstrukter» (Wechsler, 2017, s. 57).

I manualen presenteres også resultater fra studier av gruppene særlig evnerike barn ($N = 22$, svenske og danske barn) og barn med lett intellektuell funksjonsnedsettelse ($N = 22$, svenske og danske barn). Særlig evnerike barn var barn som på forhånd fikk målt en FSIQ på minst 130. Barn med lett intellektuell funksjonsnedsettelse ble diagnostisert med diagnosekriteriene i DSM-IV (American Psychiatric Association, 2000). Dette ble gjort for å undersøke om kriterierelatert validitet predikerte både intellektuell funksjonsnedsettelse og særlige evner. Gjennomsnittsverdier, standardavvik og sammenligninger av gjennomsnittsverdier med en matchet kontrollgruppe ble presentert for begge gruppene av barn. For gruppen særlig evnerike barn er alle gjennomsnittsskårer på indekssnivå og de fleste på deltestnivå signifikant høyere enn for den matchede kontrollgruppen, men generelt er skårene lavere enn forventet basert på inklusjonskriteriene. Differansene på indekssnivå varierte mellom 7,6 skalapoeng i PHI til 17,3 for AHI. Forskjellen i FSIQ var 18,9. På deltestnivå skilte skåren på SI 1,18 i skalert skår, mens Th skilte 4,46 skalerte skårer. Av barna som var identifisert som særlig evnerike oppnådde 45,5 % en FSIQ på 120 eller høyere. For gruppen barn med lett intellektuell funksjonsnedsettelse er samtlige gjennomsnittsskårer signifikant lavere, både på indeks- og deltestnivå. Differansene på indekssnivå varierte mellom 23,2 for VSI til 32,4 for VFI. Forskjellen i FSIQ var 33,8. Forskjellen på deltest-nivå varierte fra 3,41

for Bh til 6,36 for Th. Av barna som var identifisert som barn med lett intellektuell funksjonsnedsettelse oppnådde 73% en FSIQ på 75 eller lavere

Diskusjon og konklusjon

WISC-V, som ble utgitt i 2014, er en omfattende femte revisjon av den opprinnelige WISC. Den skandinaviske normeringen og den norske publikasjonen er fra 2017. Over tid har denne testen endret seg. Primært målte WISC generell evne der deltester opprinnelig ble valgt ut hvis de hadde høy ladning på den såkalte g-faktoren (korrelerte høyt med alle andre delprøver). Nå er testen et omfattende testbatteri som både gir et samlemål på generell evne og måler ulike kognitive delprosesser. Delprøver velges ut utfra spesifikk ladning på kognitive delfunksjoner. Over ti-årene har særlig amerikansk psykometri endret seg fra å vektlegge robuste samleskårer som sa noe om hvorvidt en elev var flink eller ikke, til å ville dekonstruere kognitiv funksjon i delprosesser som kan ligge under eller forklare de vansker barnet måtte ha i det daglige. Den grunnleggende tenkning som synes å underligge WISC-V synes betydelig påvirket av den såkalte «kvantitative prosess-tilnærmingen» (Ashendorf, Swenson, & Libon, 2013). Det gjør testen langt mer komplisert å bruke og fortolke, men potensialet for å forstå den underliggende kognitive strukturen er trolig mye bedre.

De to første WISC versjonene i Norge (WISC og WISC-r) ble standardisert med tung offentlig støtte, blant annet fra Forskningsrådet. Støtten ble gitt fordi evnetesting berører tunge offentlige interesser. I helsevesenet er det viktig med korrekte diagnoser. I utdanningsfeltet brukes evnetestene til å allokere hundrevis av millioner kroner til støttetiltak for elever med lærevansker. WISC-III var den første evnetesten som ble standardisert uten offentlig støtte, og må i ettertid vurderes som en billigversjon med store metodologiske svakheter (Egeland, 2006). Det hefter ingen tilsvarende vansker ved den fjerde versjonen. Fremtidige kliniske utprøvinger vil vise om det foreligger begrensninger i den foreliggende femte versjonen der Pearson Assessment, som eier testen, har

finansiert normeringsarbeidet som er gjort i Skandinavia.

Den femte versjonen av testen representerer et paradigmeskifte i testing ved at testen nå kan gjennomføres i elektronisk versjon. WISC-V er imidlertid en hybridversjon der den elektroniske versjonen er en tilpassing av den tradisjonelle versjonen. Tm har ingen elektronisk versjon. Det er ikke gjennomført reliabilitetstesting som sammenlikner den tradisjonelle versjonen og den elektroniske versjonen. Den neste Wechsler-testen som kommer på markedet (WAIS-V) vil bli konstruert som en elektronisk versjon fra starten av. Den femte versjonen muliggjør for første gang testing av arbeidsminne i både den visuelle og den språklige modaliteten, hvilket er et stort fremskritt. Det er en sammenheng mellom høy intelligens og god auditiv oppmerksomhetskapasitet (Th) mens visuell oppmerksomhetskapasitet derimot avviker minst fra gjennomsnittet hos de med intellektuelle vansker. Dette viser at de to deltestene på arbeidsminne ikke måler det samme og at begge bør være med.

Det systematiske litteratursøket vedrørende egenskaper ved testen ga ingen relevante treff utover manualen som ble utgitt i forbindelse med det skandinaviske normeringsarbeidet. At det ikke er vitenskapelige publikasjoner som anvender testen eller vurderer dens psykometriske egenskaper er ikke overraskende for en så nyutgitt test. Dokumentasjon av de psykometriske egenskapene til norske WISC-V er i denne drøftingen dermed begrenset til de resultatene som rapporteres i manualen, og er sammenfattet med hva man i litteraturen finner har vært psykometriske begrensninger ved skandinaviske normeringer av Wechsler-testene.

Skandinavia er et lite testmarked og uten offentlig støtte vil normeringene måtte være basert på et mindre utvalg enn ved normering i USA. Rent psykometrisk trenger man like mange mennesker for å få reliable normer i et lite som i et stort marked. I USA ble testen normert på 2200 personer. EFPAs testreview-modell klassifiserer størrelsen på normeringsutvalg mellom 400 og 999 som «good» (European Federation of Psychologists' Associations, 2013). Det skandinaviske utvalget var

på 660 personer, hvorav 220 fra Norge dermed kan beskrives som godt. Testen foreligger ikke i samisk versjon.

Det ble brukt en pilotgruppe på 60 normalt fungerende barn for å sile ut og rangere ledd til de språklige testene. Dette utvalget er for lite til å kunne skille mellom inngangsledd som skal løses av nær sagt alle og vanskeligere ledd som bare løses av noen få prosent i utvalget. Hvis vanskelighetsgraden av lette ledd er noe tilfeldig som det blir når man har et lite utvalg, så kan det være at et stoppkriterium på tre påfølgende null-skårer for å avslutte testing, svekker reliabiliteten til testen. Personer med lærevansker kan ha vansker med noen av de enkle leddene som ikke har blitt fanget opp i det lille pilotutvalget. Hvis tre påfølgende ledd viser seg å være spesifikt vanskelige for personer med lærevansker, mens de ville greid senere ledd i testen, får de jo ikke anledning til det, når stoppkriteriet er tre null-poengssvar. Dette var et problem både i WAIS-III og WAIS-IV (Egeland, 2014; Egeland, Bosnes, & Johansen, 2009) der testutvikler brukte samme prosedyre som nå er brukt i WISC-V. Verken for WAIS-IV og WISC-V har man undersøkt om det strengere stoppkriteriet gir en mindre reliabel skår i de skandinaviske versjonene.

Femfaktorstrukturen som introduseres i WISC-V er teoretisk spennende og fikk empirisk støtte i den opprinnelige amerikanske utgaven (Wechsler, 2014b). Flere uavhengige valideringsstudier (Canivez et al., 2016, 2017; Canivez et al., 2019; Dombrowski et al., 2015; Fenollar-Cortés & Watkins, 2018; Lecerf & Canivez, 2018; Watkins et al., 2018) har imidlertid mislyktes med å replisere støtte for femfaktormodellen som den norske utgaven er basert på og forfatterne anbefaler at man viser forsiktighet med tolkning av resultater utover fullskala IQ. Også den konfirmatoriske faktoranalysen som ble gjort på det skandinaviske normmaterialet, viser at firefaktorløsninger er vel så gode. Vi skjønner at en norsk utgave av testen må ha samme tolkningsinstruksjoner som den amerikanske originalen. Spørsmålet er like fullt om endringen fra fire-faktorløsningene som de fjerde revisjonene av WISC og WAIS var basert på, til femfaktorløsningen i WISC-V, er tilstrekkelig begrunnet.

Det er behov for flere studier som dokumenterer og undersøker måleegenskapene når testen brukes i en klinisk kontekst i Skandinavia. Basert på den dokumentasjon som finnes, fremstår WISC-V i hovedsak som et reliabelt instrument for vurdering av evnenivå og kognitiv profil. Som ved tidligere Wechsler-tester, mener vi det er en svakhet at de språklige leddenes vanskelighetsgrad ikke har vært undersøkt i et klinisk utvalg og frykter at dette kan svekke reliabiliteten av testen i klinisk bruk. Foreliggende data er utilstrekkelig til å vurdere hvorvidt WISC-V også er et valid instrument, særlig for vurdering av kognitiv profil. Hvorvidt femfaktorløsningen er et fremskritt er fortsatt usikkert siden flere faktoranalyser ikke har replisert modellen. Femfaktorløsningen i den norske versjonen er heller ikke bedre enn firefaktorløsningene som mange klinikere kjenner fra tidligere versjoner. Inntil det foreligger mer dokumentasjon gjeldende validiteten av testen bør klinisk profiltolkning gjøres med forsiktighet.

Vurdér og oppgi mulige interessekonflikter

Per Håkan Brøndbo har vært klinisk ekspert i forbindelse med den skandinaviske normeringen av WISC-V. Jens Egeland har ikke hatt noen tilknytning til normeringsarbeidet eller til testeier Pearson Assessment.

Referanser

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders : DSM-IV-TR*. Washington, DC: American Psychiatric Association.
- Ashendorf, L., Swenson, R. & Libon, D. (2013). *The Boston Process Approach to neuropsychological assessment: A practitioner's guide*. New York: Oxford University Press; US.
- Canivez, G. L., Watkins, M. W. & Dombrowski, S. C. (2016). Factor structure of the Wechsler Intelligence Scale for Children—Fifth Edition: Exploratory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 28*(8), 975-986. doi: <https://10.1037/pas0000238>
- Canivez, G. L., Watkins, M. W. & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children—Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 29*(4), 458-472. doi: <https://10.1037/pas0000358>
- Canivez, G. L., Watkins, M. W. & McGill, R. J. (2019). Construct validity of the Wechsler Intelligence Scale For Children – Fifth UK Edition: Exploratory and confirmatory factor analyses of the 16 primary and secondary subtests. *British Journal of Educational Psychology, 89*(2), 195-224. doi: <https://10.1111/bjep.12230>
- Dombrowski, S. C., Canivez, G. L., Watkins, M. W. & Alexander Beaujean, A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition with the 16 primary and secondary subtests. *Intelligence, 53*, 194-201. doi: <https://doi.org/10.1016/j.intell.2015.10.009>
- Egeland, J. (2006). Ny norsk WISC-III: Fare for overdiagnostisering av språklige vansker? *Tidsskrift for Norsk Psykologforening, 43*(7), 702-707. Lokalisert på <https://psykologtidsskriftet.no/fagartikkel/2006/07/ny-norsk-wisc-iii-fare-overdiagnostisering-av-spraklige-vansker>
- Egeland, J. (2014). Klinisk utprøving av WAIS-IV ; leddanalyser, stopp-kriterier og arbeidsminnemåling. *Tidsskrift for Norsk psykologforening, 51*(11), 911-919.
- Egeland, J. (2015). Measuring Working Memory With Digit Span and the Letter-Number Sequencing Subtests From the WAIS-IV: Too Low Manipulation Load and Risk for Underestimating Modality Effects. *Applied Neuropsychology: Adult, 22*(6), 445-451. doi: <https://10.1080/23279095.2014.992069>
- Egeland, J., Bosnes, O. & Johansen, H. (2009). Utprøving av WAIS-III i et klinisk utvalg. *Tidsskrift for Norsk psykologforening, 46*(7), 658-666.
- Esteve, O. F. & Egeland, J. (2010). Norsk version av WAIS-III: Ar det nya Generell Evne-Indeks ett anvandbart alternativ till Fullskala-IQ? *Tidsskrift for Norsk Psykologforening, 47*(3), 224-226. Lokalisert på <https://psykologtidsskriftet.no/fagbulletin>

- [/2010/03/norsk-version-av-wais-iii-ar-det-nya-generell-evne-indeks-ett-anvendbart](#)
European Federation of Psychologists' Associations. (2013). *EFPA Review model for the description and evaluation of psychological and educational tests: Test review form and notes for reviewers. Version 4.2.6*. Lokalisert på <http://www.efpa.eu/download/650d0d4ecd407a51139ca44ee704fda4>
- Fenollar-Cortés, J. & Watkins, M. W. (2018). Construct validity of the Spanish Version of the Wechsler Intelligence Scale for Children Fifth Edition (WISC-VSpain). *International Journal of School & Educational Psychology*, 1-15. doi: <https://10.1080/21683603.2017.1414006>
- Lecerf, T. & Canivez, G. L. (2018). Complementary exploratory and confirmatory factor analyses of the French WISC-V: Analyses based on the standardization sample. 793-808. Lokalisert på doi:<https://10.1037/pas0000526>
- Lynn, R. (2009). What has caused the Flynn effect? Secular increases in the Development Quotients of infants. *Intelligence*, 37(1), 16-24. doi: <https://doi.org/10.1016/j.intell.2008.07.008>
- Reinecke, M. A., Beebe, D. W. & Stein, M. A. (1999). The Third Factor of the WISC-III: It's (Probably) Not Freedom From Distractibility. *Journal of the American Academy of Child & Adolescent Psychiatry*, 38(3), 322-328. doi: <https://doi.org/10.1097/00004583-199903000-00020>
- Statistisk sentralbyrå. (2014). Oversikt over personer med ulik grad av innvandringsbakgrunn. Rapport nr. 2014/16. Lokalisert på <https://www.ssb.no/befolkning/artikler-og-publikasjoner/attachment/175380?ts=14a61e5cd90>
- Sundet, J. M., Barlaug, D. G. & Torjussen, T. M. (2004). The end of the Flynn effect?: A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32(4), 349-362. doi: <https://doi.org/10.1016/j.intell.2004.06.004>
- Watkins, M. W., Dombrowski, S. C. & Canivez, G. L. (2018). Reliability and factorial validity of the Canadian Wechsler Intelligence Scale for Children-Fifth Edition. *International Journal of School & Educational Psychology*, 6(4), 252-265. doi: <https://10.1080/21683603.2017.1342580>
- Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. San Antonio, TX, US: Psychological Corporation.
- Wechsler, D. (2009). *Wechsler Intelligence Scale for Children (Fourth Edition) Norsk versjon*. Stockholm: Pearson Assessment.
- Wechsler, D. (2014a). *Wechsler intelligence scale for children-Fifth Edition (WISC-V)*. Bloomington, MN: Pearson.
- Wechsler, D. (2014b). *WISC-V: Technical and interpretive manual*: NCS Pearson, Incorporated.
- Wechsler, D. (2017). *Wechsler intelligence scale for children-Fifth Edition (WISC-V) Manual, Del 1, Norsk versjon*. Stockholm Pearson Assessment