

Valg av evnetest har konsekvenser for vurdering av diagnosen psykisk utviklingshemning

PSY
KOL
OGI

I februarnummeret rapporterte Ole Bosnes om dårlig samsvar mellom resultater på WASI og WISC-III, og marte til forsiktighet. Vår undersøkelse av 81 barn tilsier at samsvaret blir dårligere jo eldre barna er.

TEKST

Lise Kjenseth

PUBLISERT 5. mai 2010

ABSTRACT:

Comparison of WISC-III- and WASI-scores in a clinical sample: choice of intelligence test has consequences for the diagnosis of mental retardation.

Analysis of WISC-III (Norwegian version published 2003, Swedish norms) and WASI (Norwegian version published 2007, U.S. norms) protocols from 81 children and adolescents referred for clinical assessment, showed that IQ and index scores for the WISC-III were significantly lower compared with the WASI scores. This was especially so for the older group of children (over 13 years). When comparing IQ, the full-scale IQ and performance score showed the weakest correlation. Using the WASI, 32 % of the children had a full-scale IQ less than 70. With the use of Swedish norms for the WISC-III, 81 % of the children had a full-scale IQ less than 70. It seems therefore, that the choice of intelligence test and corresponding norms has implications for the diagnosis of mental retardation.

Keywords: intelligence testing, abbreviated Wechsler scales, mental retardation, intellectual disabilities.

I Norge har Wechsler Intelligence Scale for Children-Third Edition (WISC-III) så langt vært den mest brukte evnetesten ved utredning og diagnostisering av barn og ungdom. Wechsler Abbreviated Scale of Intelligence (WASI) er en kortversjon av Wechslers intelligens tester for barn og voksne i aldersgruppen 6–89 år. Bruk av WASI baseres på amerikanske normer fra 1999, mens WISC-III vanligvis brukes med svenske normer fra 1999.

Når barn og unge henvises til vår avdeling, er de vanligvis tidligere testet med WISC-III av Pedagogisk-psykologisk tjeneste. I vårt arbeid har vi likevel ofte behov for retesting med et mindre tidkrevende, men samtidig pålitelig instrument, for å vurdere generelt intellektuelt nivå. WASI kan da være et alternativ.

I vår kliniske hverdag har vi imidlertid ofte erfart at resultatene på de to testene samsvarer dårlig. Spesielt gjelder dette for større barn og ungdommer med svakt

evnenivå. Også andre, som Bosnes (2010), rapporterer fra sin kliniske praksis om dårlig samsvar mellom resultater på WASI og WISC-III. Etter hvert ble vi opptatt av hvor stor forskjellen mellom skårene på WASI og WISC-III var, og hvor stor vi kan «tillate» den å være for i det hele tatt å snakke om «samsvar». Særlig var dette aktuelt ved diagnostisering av psykisk utviklingshemning (ICD-10, 2003).

Av WISC-IIIs totalt 13 deltester benyttes 10 for beregning av IQ: total-IQ (FSIQ), verbal-IQ (VIQ) og utførings-IQ (PIQ). WASI består kun av fire deltester som alle brukes for beregning av IQ: FSIQ, VIQ og PIQ. Ifølge Donders (2001) har klinisk nyttig informasjon som måles ved prosesseringshastighet (hurtighetsindeks), høy kriterievaliditet. Av den grunn ville man vente at prestasjoner på WASI ligger noe høyere enn på WISC-III. Dersom denne faktoren tas ut, som ved faktorskårene «verbal forståelse» og «perseptuell organisering» i WISC-III, vil en sammenligning av disse målene med VIQ og PIQ fra WASI være mer korrekt i slike grupper (Sundet, e-post, mai 2005; Ørbeck & Sundet, 2007). Ifølge den amerikanske manualen til WASI kan deltestene «Ordforråd», «Likheter» og «Terningmønster» betraktes som parallellversjoner til deltester med samme navn i WISC-III (se også Ørbeck & Sundet, 2007). I tillegg til IQ-mål kan det derfor være av særlig interesse å sammenligne disse.

Metode

Søknad om godkjenning av prosjektet ble sendt Regional komité for medisinsk og helsefaglig forskningsetikk Sør-Øst D. Svaret inneholdt følgende vedtak: «Prosjektet anses som kvalitetssikring av et etablert behandlingstilbud. Prosjektet faller derved utenfor retningslinjene for hva som skal behandles av komiteen.»

I alt ble protokoller for 81 barn og unge (39 jenter og 42 gutter) henvist til avdeling for barne- og ungdomshabilitering ved Akershus Universitetssykehus (Ahus) undersøkt. Som ventet var barna noe eldre ved testing med WASI (gjennomsnittlig 12 år og 4 måneder) enn ved testing med WISC-III (11 år 3 md., med SD 2 år 4 md. for begge testene), da WASI alltid ble administrert sist.

Utvalget oppfattes som representativt for den barnegruppen som typisk henvises til nevropsykologisk utredning ved vår avdeling. Testresultatene på WASI og WISC-III (IQ-/indeksskårer og skalerte skårer) ble analysert med t-test for parede utvalg for sammenligning ved hjelp av gruppegjennomsnitt, differanseskårer og rangkorrelasjon. Korrelasjonsanalyse belyser i hvilken grad den innbyrdes rangordningen i gruppen er den samme ved test 1 og test 2. I tillegg ble testresultatene analysert med t-test for uavhengige utvalg, samt ved hjelp av enveis og toveis variansanalyser (ANOVA) for sammenligning av resultater på tvers av aldersgrupper (8–9 år, 10–12 år, 13 år og eldre) og kjønn. Beregninger ble gjort med statistikkprogrammet SPSS.

Resultater og diskusjon

Resultatene av undersøkelsen viste at alle IQ- og indeksmål for WISC-III med svenske normer var signifikant svakere enn de tilsvarende målene i WASI med amerikanske normer (se tabell 1). Differansen i fullskala-IQ mellom WASI og WISC-III var 17,1 poeng.

Enda større (19,8 poeng) var differansen mellom de to testene på utførings-skåren (PIQ). Dataene bekreftet vår antagelse om bedre overensstemmelse mellom faktorskårene «verbal forståelse» og «perseptuell organisering» fra WISC-III og tilsvarende IQ-mål fra WASI. Sammenligningen mellom WASI og WISC-III for utvalget sett under ett viste *svakest* samsvar på utføringsmålene, noe som igjen hadde innvirkning på totalskåren, selv om den innbyrdes rangordningen i barnegruppen var forholdsvis lik. Andre undersøkelser har ikke funnet så store forskjeller som i vår undersøkelse, selv om Bosnes' (2010) resultater fra et norsk utvalg var nærmere våre resultater enn amerikanske og canadiske studier. I WASI-manualen refereres sammenligningen av testresultatene for et ikke-klinisk utvalg på 176 barn/unge. Denne undersøkelsen ga bare 0,2 IQ-poengs forskjell, med korrelasjon $r = .87$. Scott, Austin og Reid (2007) fant en fullskala-IQ forskjell på kun 0,4 poeng i et klinisk utvalg av 25 canadiske barn, med $r = .81$. Bosnes (2010) fant i sitt kliniske utvalg på 55 barn en forskjell i total-IQ på 9,4 poeng.

TABELL 1. PRESTASJONER PÅ INTELLIGENSTESTENE WISC-III OG WASI (N=81). IQ-SKÅRER OG INDEKS/FAKTORSKÅRER MED GJENNOMSNI TT (M) OG STANDARDAVVIK (SD), SAMT KORRELASJON (R) OG IQ-SKÅRE-DIFFERANSE (= GJENNOMSNI TT LIG FORSKJELL) MELLOM DE TO TESTENE. *P 0,001

	WISC-III M (SD)	WASI M (SD)	r	Differanse WASI - WISC-III M (SD)
Total-IQ	59,1 (14,7)	76,2 (12,4)	. 82*	17,1* (8,5)
Verbal IQ	66,5 (14,3)	76,3 (12,9)	. 76*	9,8* (9,5)
Utførings-IQ	60,2 (16,2)	80,0 (13,3)	. 74*	19,8* (10,9)
Verbal forståelse	69,9 (13,4)	76,3 (12,9)	. 77*	6,4* (9,0)
Perseptuell organisering	65,6 (16,1)	80,0 (13,3)	. 75*	14,4* (10,6)

I Norge har det i de senere årene vært fremsatt flere synspunkter på – samt vært diskusjoner rundt – skåring, normering og tolkning ved bruk av WISC-III (Strand, 2005; Nilsen, 2005; Sundberg, Egeland, Andreassen & Stensli, 2006; Andreassen, 2006). Funn har vist at normene er for strenge for ungdommer med lavt evnenivå. Vi har derfor sett på hvordan aldersfaktoren påvirker *differansen* mellom testresultatene på WASI og WISC-III (se tabell 2). Her framgår det at forskjellen i IQ-/faktorskårer gjennomgående var størst for den eldste gruppen barn: signifikant ($p 0,001$) for alle verbale og ikke-verbale mål.

TABELL 2. SAMMENLIGNING AV GJENNOMSNI TT LIG FORSKJELL (DIFFERANSE) I IQ/INDEKS-SKÅRER MELLOM WASI OG WISC-III, GRUPPERT ETTER TESTALDER VED WISC-III (N = 81). **P 0,001 *P 0,05

TABELL 2. SAMMENLIGNING AV GJENNOMSNITTLIG FORSKJELL (DIFFERANSE) I IQ/INDEKS-SKÅRER MELLOM WASI OG WISC-III, GRUPPERT ETTER TESTALDER VED WISC-III (N = 81). **P 0,001 *P 0,05

	Alder ved testing med WISC-III	Total-IQ	Verbal-IQ	Utførings-IQ	Verbal forståelse	Perseptuell organisering
Differanse IQ-skåre WASI - WISC-III	8-9 år (n=30)	14,3**	6,9**	17,8**	3,6*	13,2**
	10-12år (n=32)	16,4**	9,1**	19,4**	6,0**	13,8**
	>13 år (n=19)	22,6**	15,7**	23,5**	11,4**	17,0**

Siden tre av deltestene på WASI kan betraktes som parallellversjoner til deltester i WISC-III, ble T-skårene for WASI omregnet til skalerte skårer (se egen tabell i WASI-manualen) for direkte sammenligning. Den største forskjellen på deltestnivå mellom WASI og WISC-III ble funnet på «Terningmønster» (se tabell 3).

TABELL 3. PRESTASJONER PÅ DELTESTNIVÅ FOR WASI OG WISC-III (N=81). GJENNOMSNITTLIG SKALERTE SKÅRER (M) OG STANDARDAVVIK (SD), SAMT DIFFERANSER (= GJENNOMSNITTLIG FORSKJELL) OG KORRELASJON (R) MELLOM DE TO TESTENE. * P 0,001

	Likheter	Ordforståelse	Terningmønster
WASI M (SD)	5,25 (2,89)	5,04 (2,91)	6,64 (2,77)
WISC-III M (SD)	5,43 (3,01)	5,35 (2,72)	4,69 (3,16)
Differanse M (SD)	-0,19 (2,48)	-0,31 (2,34)	1,95* (2,0)
r	.65*	.66*	.78*

På «Likheter» og «Ordforståelse» var WISC-III skårene på gruppenivå bedre enn WASI-skårene, i motsetning både til «Terningmønster» og IQ-/indeksskårer. Siden WASI alltid ble gjennomført sist, kan mulig lærings- eller retesteffekt bidra til forhøyet resultat og være en sannsynlig forklaring på at «Terningmønster»-skåren er signifikant høyere ved testing med WASI enn ved testing med WISC-III. Likevel bidrar ikke nødvendigvis tidligere erfaring med en test til bedre prestasjoner ved senere testing, men er blant annet avhengig av intervallet fra første til andre testing (Strauss, Sherman og Spreen; 2006). I vårt materiale ble det imidlertid ikke funnet noen slik sammenheng ($r = - .238$ for «Terningmønster»).

Vi fant en gjennomsnittlig skåreforskjell på fullskalanivå mellom WASI og WISC-III på 17 poeng, noe som innebærer at for mange testprotokoller var differansen enda større. En differanse i IQ-skårer på 15 poeng tilsvarer ett standardavvik. Ifølge Bracken (1988) er dette en signifikant forskjell mellom to tester som antas å måle det samme. Oppnådd

IQ-skåre skal likevel brukes med varsomhet, blant annet fordi testene har målefeil. Siden barna i vårt kliniske utvalg i utgangspunktet er svake evnemessig, må en test ha et tilstrekkelig antall enkle og graderte oppgaver for å sikre differensiert vurdering av evnenivået. For de mest vanlige psykometriske årsakene til signifikant forskjell mellom testresultater viser jeg ellers til drøfting hos Bracken (op.cit).

Et vesentlig spørsmål blir likevel hvor store forskjeller som kan tillates når vi i praktisk klinisk arbeid angir at to testresultater skal «samsvare». Problemet kan illustreres når vi ser på andelen barn med IQ-skåre under 70 poeng etter testing med de to evnetestene. Som vist i tabell 4 gjelder dette hele 81,5 % av barna som ble testet med WISC-III, mens bare 32,1 % får IQ < 70 vurdert med WASI. Diagnostisering av psykisk utviklingshemning etter ICD-10-kriteriene bygger blant annet på testing av evnenivå. Hvilke konsekvenser bør våre funn få for valg av instrument ved diagnostisering av psykisk utviklingshemning? Selv om WASI-manualen advarer mot å benytte denne testen ved diagnostisering, så synes bruk av WISC-III på sin side å medføre risiko for overdiagnostisering av psykisk utviklingshemning (Grøsvik, 2008).

TABELL 4. FORDELING AV BARN ETTER EVNENIVÅ, IQ < 70 (N=81).

	IQ < 49	IQ = 50-69
WASI (N = 81)	0	26 (32,1 %)
WISC-III (N = 81)	23 (28,4 %)	43 (53,1 %)

De største differansene mellom WASI og WISC-III ble funnet hos de eldste barna. Våre resultater viser at det er nødvendig å være ekstra varsom i forhold til eldre barn og ungdommer når det gjelder vurdering av diagnosen psykisk utviklingshemning generelt, og gradering av evnenivå spesielt. Det vår undersøkelse ikke kan si noe om, er hvorvidt resultatene er knyttet til denne kliniske gruppens forholdsvis lave evnenivå, eller om funnene er uttrykk for en generell forskjell mellom WISC-III og WASI.

Konklusjon

Prestasjoner på de norske oversettelsene av evnetestene WISC-III (svensk normsett) og WASI (amerikansk normsett) hos et klinisk utvalg barn og unge henvist til barne- og ungdomshabiliteringen ble sammenlignet. Alle IQ- og indeksmål på WISC-III var signifikant svakere enn de tilsvarende målene i WASI. For utvalget som helhet var det best samsvar på verbalprøvene. Det svakeste samsvaret ble funnet mellom utføringsskårene. Spesielt for de eldste barna (over 13 år) ble forskjellen i skårer mellom de to testene stor. Videre har valg av evnetest store konsekvenser for vurdering av diagnosen psykisk utviklingshemning.

På bakgrunn av resultatene fra vår undersøkelse gir jeg min fulle tilslutning til Bosnes' håp om at WISC-IV har hatt en bedre utprøving/normering slik at vi nå får et pålitelig

verktøy til evnevurdering og diagnostisering i vår kliniske hverdag.

Lise Kjenseth

Ankerveien 82 E

0766 Oslo

Tlf. jobb 67 96 22 42

Mob. 91 77 69 98

E-post lise.kjenseth@ahus.no

Teksten sto på trykk første gang i Tidsskrift for Norsk psykologforening, Vol 47, nummer 5, 2010, side 421-423

TEKST

Lise Kjenseth

+ Vis referanser

Referanser

Andreassen, T. H. (2006). WISC-III og de eldste ungdommene: praktiske erfaringer. 506-507.

Bosnes, O. (2010). Norske versjoner av Wechsler-skalaene for barn: En advarsel fra klinisk praksis. 118-120.

Bracken, B. A. (1988). Ten Psychometric Reasons Why Similar Tests Produce Dissimilar Results. ,155-166.

Donders, J. (2001). Using a short form of the WISC-III: Sinful or smart? 99-103

Ellertsen, B. og Johnsen, I. M.B. (2003).

Oslo: Assesio Norge AS.

Grøsvik, Kjellaug (2008). Diagnostisering av utviklingshemning hos barn. I: Eknes, J. Bakken, T.L. Løkke, J. A. og Mæhle, I. (red.), Oslo: Universitetsforlaget.

5.opplag 2003. Oslo: Gyldendal Norsk Forlag.

Nilsen, D.Ø.E. (2005). WISC-III og normer. 532 - 533.

Scott, W. C., Austin, D. W. og Reid, D.S. (2007). Investigation of the WISC-III and WASI in Clinical Child Populations: A Framework for Further Exploration. , 249-254.

Strand, N. (2005). WISC-III skaper uro. 519-522.

Strauss, E., Sherman, E. og Spreen, O. (2006).

Third Edition. Oxford: Oxford University Press.

Sundberg, H., Egeland, J., Andreassen, T. H. og Stensli, O. (2006). Sammenligning av WISC-R og WISC-III-skårer i et klinisk utvalg: Svakheter i WISC-III-normer eller foreldete WISC-R-normer? 476-481.

Wechsler, D. (1991).
Psychological Corporation.

Ørbeck, B. og Sundet, K. (2007).
Tryck.

San Antonio: The

Stockholm: Katarina