

Klinisk utprøving av wais-iv

Leddanalyser, stopp-kriterier og arbeidsminnemåling

Analyser av resultater fra WAIS-IV fra et blandet klinisk utvalg viser at en blind tillit til normeringsmateriale kan svekke reliabilitet og validitet når det brukes overfor personer med lærevanskeproblematikk.

TEKST

Jens Egeland

PUBLISERT 5. november 2014

ABSTRACT:

Clinical validation of WAIS-IV. Item Analysis, Stop Criteria and Measurement of Working Memory.

Tests can display different psychometric properties in clinical samples compared to the normal standardization sample. However, it is particularly important that the test measures reliably and valid in the clinic. The present study examines three psychometric issues relating to the new Scandinavian version of the Wechsler Adult Intelligence Scale, version IV, in a mixed clinical sample of 226 subjects.

As some verbal items are new to the Scandinavian version, and original items are rearranged according to the performance of the relatively small Scandinavian pilot study, item analyses are performed to test whether verbal items become increasingly difficult as intended. Imperfect increase in difficulty level may have detrimental consequences to the validity of the stricter Stoprules in WAIS-IV. The rank order correlations between scores with different stop-criteria are examined. The third problem relates to the extension of working memory testing in the WAIS-IV by including a digit span sequence task. A Confirmatory Factor Analysis (CFA) is performed to test whether the New Digit Span test differentiates between storage and manipulation levels of working memory (WM). All subjects are also tested with the Spatial Span subtest from the Wechsler Memory Scale-III.

Results: With some few exceptions the item analysis showed the expected triangular form indicating increased difficulty level. Analyses of different stop criteria showed insufficient rank order correlations for Visual Puzzles, Information and Comprehension when applying the three zero-score stop rule prescribed in the handbook. The CFA showed best fit for a modality division of WM.

Conclusion: The study shows some insufficiencies that the clinician should be aware of when using the Scandinavian WAIS-IV version in the clinic.

Wechslers evnetest for voksne (Wechsler Adult Intelligence Scale: WAIS) (Wechsler, 2008) er den eneste psykologiske testen som har en bruksfrekvens på over halvparten av alle norske psykologer (Vaskinn & Egeland 2012). Resultatene fra denne prøven legger premissene for en rekke viktige avgjørelser diagnostisk og tiltaksmessig: Diagnostikk av psykisk utviklingshemning eller spesifikke lærevansker er eksempler på førstnevnte; tilgjengelighet for terapi og behov for spesialpedagogiske støttetiltak er eksempler på sistnevnte. I nevropsykologien brukes testen som en «benchmarking» av premorbid intellektuell funksjon eller som evidens for modalitetsspesifikke vansker eller fokale utfall etter hjerneskade. Den nåværende versjon fire av WAIS kom i norsk utgave i 2011, basert på et skandinavisk standardiseringsutvalg. Det er neppe noen test hvor den nasjonale utgaven har vært så grundig standardisert som WAIS. Allikevel er det grunn til å vurdere eventuelle svake sider ved testen og ved standardiseringsprosessen. De psykometriske egenskapene som gjør seg gjeldende i et normeringsutvalg, kan eksempelvis vise seg å være annerledes enn i et klinisk utvalg. I standardiseringen av WAIS-IV har man valgt å ikke bruke kliniske data i utprøvingen av de psykometriske egenskapene, så som faktorstruktur og reliabilitet.

Denne artikkelen er basert på en klinisk utprøving av WAIS-IV fra 226 fortløpende henviste pasienter til forfatterens deltids avtalepraksis i nevropsykologi. Følgende problemstillinger er undersøkt:

- 1) Er vanskegraden av ledd i de språklige prøvene gradvis økende også i et klinisk utvalg?
- 2) Svekker innføring av strengere stopp-kriterier reliabiliteten av de enkelte delprøvene?
- 3) Fører den nye sekvens-oppgaven i Tallspenn til en større sensitivitet for arbeidsminnevansker?

«En kunne tenke seg at personer med svak kunnskap om dette vil falle ut av delprøven selv om de kunne ha klart vanskeligere led»

Leddanalyser

I en evnetest skal vanskegraden av leddene øke gradvis, slik at en visuell fremstilling av løsningsfrekvens vil se ut som en rettvinklet trekant. I WAIS-III viste det seg at løsningsfrekvensen hos personer som samlet hadde en IQ under 90, så mer ut som en L, med den konsekvensen at de skåret signifikant under sitt eget gjennomsnitt på

Likheter, mens personer med middels god IQ skåret på samme nivå (Egeland, Bosnes & Johansen 2009). Selv om slike målefeil i enkeltdelprøver i liten grad påvirker fullskala IQ, vil det kunne ha betydning når man fortolker sterke og svake ferdigheter i diskrepansanalyser. Når Likheter i enkelttilfeller var over 3 skalerte skårer under de andre språklige testene i WAIS-III, ville effekten på indeksskåren kunne være over 5 skalapoeng, noe som i verste fall kunne bidra til en undervurdering av språklige ferdigheter sammenlignet med de visuelle-romlige evnene målt i Perseptuell resonneringsindeksen. Vi gjennomførte derfor leddanalyser for å se om tilsvarende forskjeller ses i WAIS-IV.

Konsekvenser av redusert kuttepunkt for reliabilitet

Kuttepunktet for når en deltest avsluttes, har variert opp og ned i ulike versjoner av WAIS. I WAIS-III ble kuttepunktene for de fleste verbalprøvene satt til seks nullpoengssvar. For å effektivisere testingen og unngå ubehaget med å måtte svare på mange ledd som er for vanskelige, ønsket man en reduksjon av antall feilsvar før delprøven ble avsluttet i versjon fire. En rangordens-korrelasjon på 0,98 eller høyere mellom et lavt og et høyt kuttepunkt var tilstrekkelig for å betrakte det lave kuttepunktet som reliabelt (Wechsler, 2008). Tilsvarende beregninger skal ha vært gjort også i den skandinaviske versjonen, men det fremkommer ikke av håndboken. Gjennom leddanalyser har leddene på de språklige prøvene blitt stokket noe om for å oppnå jevnt økende vanskegrad. På Informasjonsdelprøven er tre av leddene innledningsvis i testen geografirelatert. En kunne tenke seg at personer med svak kunnskap om dette vil falle ut av delprøven selv om de kunne ha klart vanskeligere ledd. I klinikken vil motivasjon og impuls kontroll kunne variere mer enn i et motivert standardiseringsutvalg, og at dette påvirker sannsynligheten for nullsvar allerede innledningsvis i oppgavene. Dette er imidlertid bare et problem hvis det påvirker rangkorrelasjonen. Vi beregner derfor rangkorrelasjon med kuttepunkt 3, 4 og 5 for delprøvene Visuelle Puslespill, Matriser, Ordforståelse, Informasjon og Likheter. Dernest beregner vi implikasjonene av eventuell svak rangkorrelasjon på enkeltprøver. Dette gjøres ved å se på antall personer som ville fått en annen skår ved et annet kuttepunkt. Dernest beregner vi om enkeltdelprøver systematisk får lavere skår enn andre tester innenfor samme faktorindeksskår. I et stort sammensatt utvalg er det ingen grunn til å forvente forskjeller her.

Måling av arbeidsminne

Arbeidsminne (AM) som konstruert har fått en stadig økende betydning i psykologien. Dysfunksjon i AM er dokumentert ved en rekke tilstander så som depresjon, schizofreni (Egeland et al. 2003), ADHD (Martinussen, Hayden, Hogg-Johnson & Tannock 2005) og kronisk smerte (Berryman, Stanton, Bowering, Tabor, McFarlane & Moseley 2013). Det er noe uenighet om hvilke dimensjoner som skal vektlegges i undersøkelse av arbeidsminnet. En dimensjon som det i økende grad refereres til, er lagring versus bearbeiding (Storage vs. Manipulation). Tallhukommelses-delprøven er bygd på denne distinksjonen: Tallhukommelse forlengs er et mål på nærmest automatisk

oppmerksomhetskapasitet svarende til den fonologiske sløyfen i Baddeley og Hitchs (1974) AM-modell eller til korttidsminne. Noen forbeholder AM-begrepet til oppgaver som i tillegg til enkel kortvarig lagring av informasjon også krever *arbeid*, altså at man må foreta en mental prosess simultant med lagringen. Tallhukommelse baklengs og Tallhukommelse i sekvens tilfredsstiller kriteriene til å være arbeidsminnetester i denne forstand. Et spørsmål er likevel om bearbeidingselementet er tungt nok til at man kan sies å måle noe systematisk forskjellig i Tallspenn forlengs og bearbeidingsoppgavene. Tallhukommelse i sekvens er en ny oppgave i WAIS-IV som er tatt med for nettopp å øke kravet til bearbeiding. I Baddeleys modell vil slik simultan bearbeiding stille krav til den sentrale styringsenheten ('the central executive'). Delprøven Bokstav-tall-serier regnes også som en AM-bearbeidingsoppgave, og er en valgfri test i WAIS-IV.

Når det gjelder enkel lagring, skilles det i Baddeleys modell mellom den fonologiske løkken og den visuelle skisseblokken. Diskusjonen i arbeidsminneforskning går på hva som er den mest potente dimensjonen – kompleksitetsnivå eller modalitetsforskjeller. I WAIS-IV undersøkes arbeidsminnet kun innenfor den auditive modaliteten. Hvis modalitetsforskjeller er marginale, vil det være gyldig å si noe generelt om arbeidsminnet basert kun på auditiv måling. I de fleste tilfeller ønsker vi nettopp å si noe generelt om hvorvidt en person har nedsatt arbeidsminne eller ikke. Hvis modalitetsforskjeller er den mest potente dimensjonen for å forklare varians i arbeidsminnetester, må man avgrense seg til å si noe om auditiv kapasitet, som kan være relevant nok i forhold til utredning av dysleksi, men hvor vi da ikke kan si noe om personens overordnede arbeidsminneressurser.

I undersøkelsen faktoranalyserer vi de tre tallspennoppgavene fra WAIS-IV, Bokstav-Tallserier sammen med Visuelt spenn forlengs og baklengs fra Wechsler Memory Scale III for å teste følgende faktormodeller:

- 1) AM som en faktor (ingen oppdeling i nivåer i WAIS-IV eller modaliteter i WAIS-IV/WMSIII)
- 2) Nivåmodellen: Tallspenn forlengs og Visuelt spenn forlengs måler enkel lagring, mens Tallspenn baklengs, Tallspenn i sekvens og Tall-bokstav-sekvensering måler bearbeidingsarbeidsminne.
- 3) Modalitetsmodellen: Alle tre tallspennsvariantene og Tall-Bokstav-sekvensering lader på en auditiv faktor mens Visuelt spenn forlengs og baklengs lader på en visuell faktor.

Metode

Utvalg

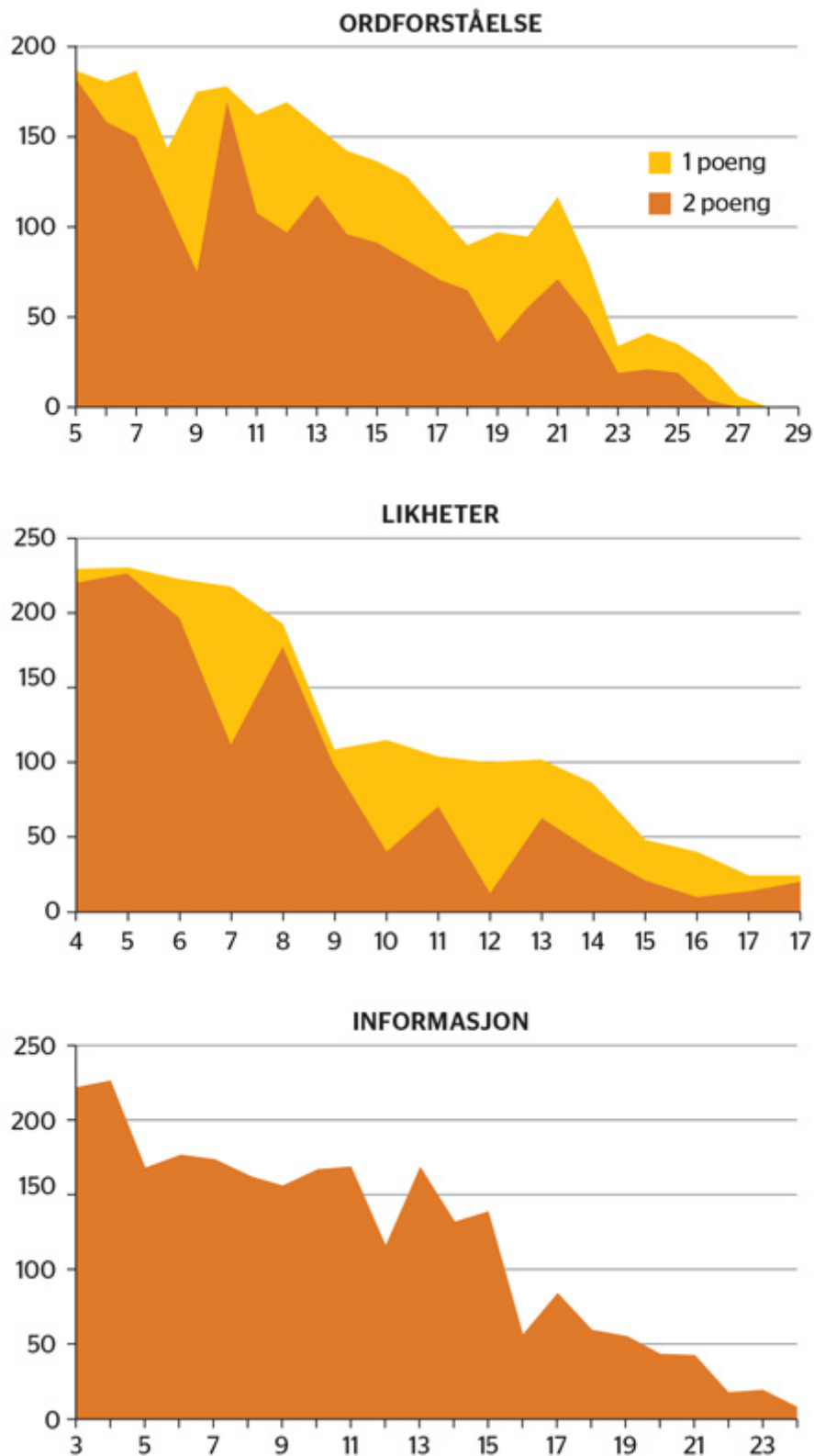
Registerdata fra 226 pasienter som kom til nevropsykologisk undersøkelse i forfatterens deltids avtalepraksis, er analysert med tillatelse fra Norsk samfunnsvitenskapelig datatjeneste. Alder varierte fra 16 til 82 år (gjennomsnitt: 38, s.a.14,3. Det var femtitte prosent menn. Gjennomsnittlig Generell evneindeks var 92 (s.a.15) og varierte fra 57 til

141. Diagnose etter utredning var ikke uavhengig av WAIS-IV skårer, slik at det ikke gir mening å referere diagnose i tabells form her. Når det gjelder årsak til henvisning, var 24 % henvist til ADHD-utredning. Uklare lærevansker var årsak til 20 % av henvisningene, mens ytterligere 15 % hadde en mulig kognitiv svikt av ukjent årsak. Ti prosent hadde kjent hjernedysfunksjon (hjerneslag eller traumatisk hodeskade). En komplett oversikt over alle henvisningsgrunner fremgår av Egeland (akseptert).

Resultater

Leddanalyser

Figur 1 viser leddanalysene for de tre Språklig forståelses-kjernedelprøvene. Det fremkommer noen avvikende ledd som enten er vanskeligere enn det foregående eller lettere enn det etterfølgende. I Ordforståelse gjelder dette særlig ledd 9, der de fleste mener at «gjøre ferdig» er synonymt med «avslutte» og der 2-poengssvarene «stoppe» eller «ende» er relativt lavfrekvente. I Likheter avviker ledd 7, 10 og 12 der 1-poengssvarene «kroppsdel», «natur» og «håp» er relativt høyfrekvente. Visuelt betraktet har Ordforståelse og Informasjon likevel en tilnærmet trekantform med gradvis økende vanskegrad. Likheter ligner noe mer på en L. Analyse av personer med Generell Evne-Indeks (GEI) på 85 eller lavere viser at Likheter har en skår på 0,77 skalert skår lavere (5,81, s.a. 2,0) enn gjennomsnittet av Ordforståelse og Informasjon (6,58, s.a.1,8) ($t = 3,22$, $p = 0.002$). Tilsvarende beregning gjort på personer med GEI over 85 viste en forskjell på 0,10. Denne forskjellen i grad av relativt svakere Likheter enn Ordforståelse/Informasjon mellom normalgruppen og lavskårgruppen var ikke signifikant.



Figur 1 Leddanalyser

Konsekvenser av redusert kuttepunkt

Korrelasjonene mellom standard kuttepunkt på 3 null-svar og henholdsvis 4 og 5 var som følger: Visuelle puslespill: 0,952 (kuttepunkt 4) og 0,929 (kuttepunkt 5). Matriser: 0,986 og 0,982; Informasjon: 0,951 og 0,911; Ordforståelse: 0,982 og 0,978; Likheter: 0,995 og 0,991. Begge korrelasjonene for Visuelle puslespill og Informasjon var altså lavere

enn kriteriet for et reliabelt lavt kuttepunkt. Dette gjaldt også korrelasjonen mellom kuttepunkt 3 og 5 i Ordforståelse.

For 41 % av utvalget ville et høyere kuttepunkt endre den skalerte skåren for Visuelle puslespill. Ti prosent av utvalget ville fått en skalert skår som var et standardavvik eller mer bedre ved kuttepunkt på 5. 1,7 % av utvalget ville fått en bedring på 5 skalerte skårer.

For Informasjon ville den skalerte skåren ha endret seg for 18 % av utvalget ved et høyere kuttepunkt. Tre prosent av utvalget ville fått en bedring på ett standardavvik eller mer. For Ordforståelse ville syv prosent fått endret skalert skår, men bare en prosent ville fått en bedring på ett standardavvik.

T-test for parede utvalg viser en signifikant forskjell mellom Visuelle puslespill (gjennomsnitt 8,0; s.a. 3,2) og henholdsvis Matriser (8,8, s.a. 3,2, $t = 3,63$, $p < 0.001$) og Terningmønster (8,9, s.a. 27, $t = 4,26$, $p < 0.001$) med standard kuttepunkt. Tillater man kuttepunkt 5 på Visuelle puslespill, øker gjennomsnittsskåren for Visuelle puslespill til 8,8,(s.a. 3,0) som er på samme nivå som Matriser og Terningmønster.

Innenfor Verbal forståelses-indeksskåren var det ingen signifikante forskjeller mellom de skalerte skårene for delprøvene som inngår i indeksen.

Tabell 1 Mål på samsvar mellom ulike faktormodeller og observerte skårer for tallspenn, bokstav-tall-sekvensering og visuelt spenn fra WAIS-IV og WMS-III (N=226).

MODELL	χ^2 (D.F.)	χ^2 / D.F.	GFI	AGFI	CFI	RMSEA
1 Enhetlig arbeidsminnemodell	32.43(9)	3.60	0.952	0.90	0.94	0.109
2 Nivåmodellen: TF+VF vs TB, TS, VB, BT	27.13(8)	3.39	0.96	0.90	0.94	0.104
3 Modalitetsmodellen: TF, TB, TS, BT vs VF,VB	6.55(8)	0.82	0.99	0.97	1.0	0.0

Arbeidsminne

Tabell 1 viser resultatet av den konfirmatoriske faktoranalysen. Målene som viser adekvat eller godt samsvar mellom modell og observasjoner er uthevet. Alle χ^2 -verdiene er signifikante, og det er ingen konvensjon med hensyn til hva som representerer et godt samsvar her. Den laveste verdien er den beste, men parsimonitetsprinsippet tilsier at den enkleste modellen er å foretrekke hvis de ellers forklarer like mye. Man deler derfor på frihetsgradene, og det resulterende målet ($\chi^2 / d.f$) bør være under 2 (Jöreskog & Sörbom 1993). I dette tilfellet er det kun modalitetsmodellen som har en slik lav verdi. RMSEA-verdier over 0,8 viser utilstrekkelig samsvar mellom modell og observerte skårer, mens verdier under 0.06 viser godt samsvar. RMSEA-verdiene viser klart utilstrekkelig samsvar for den enhetlige modellen og nivåmodellen og meget godt samsvar for modalitetsmodellen. Når det gjelder GFI, AGFI og CFI, viser verdier over 0,9

et godt samsvar. Samlet sett indikerer CFA- analysen at nivåmodellen og den enhetlige modellen ikke kan forklare de observerte skårene, mens modalitetsmodellen viser godt samsvar.

Diskusjon

Det er svært viktig at evnetester er valide og reliable, særlig innenfor det pasientsegmentet den primært skal brukes overfor, altså personer der det mistenkes en eller annen lærevanskeproblematikk. Normer må naturligvis genereres fra et normalutvalg, men kliniske utvalg er nødvendige for å sjekke reliabilitet og validitet. Historisk har vi flere eksempler på at Wechsler-testene oppfører seg annerledes i kliniske utvalg enn forutsatt i manualen, og at dette kan svekke gyldigheten av de kliniske vurderinger som gjøres på grunnlag av testen. I denne artikkelen har jeg plukket ut tre problemstillinger relatert til den siste versjonen, WAIS-IV. Nedenfor følger en kort drøfting med påfølgende anbefaling i forhold til hver av de tre punktene, før jeg avslutningsvis knytter noen overordnede kommentarer til testen.

Leddanalyser

Enkeltledd avviker en del, men i hovedsak hadde kurven over løsningsfrekvens i Ordforståelse og Informasjon den trekantformen man forventer. Likheter hadde noe mellom en trekantform og en L-form. Konsekvensen av dette er at personer som samlet skårer lavt på evnetesten, fikk en noe svakere skår på Likheter enn de fikk på de andre to prøvene som inngår i Verbal forståelses-indeksskåren. I WAIS-III var Likheter generelt mer vanskelig enn de andre VFIprøvene (Egeland, Bosnes & Johansen 2009). Gruppen med lavest IQ skåret da 1,24 skalert skår svakere på Likheter enn på de andre språklige prøvene. Det er altså grunn til å konkludere med at en metodisk svakhet i forrige versjon langt på vei er justert for i WAIS-IV.

Rangeringen av ledd ut fra vanskegrad er gjort på skandinavisk basis, og håndboken lister opp noen ledd som i standardiseringsutvalget falt noe forskjellig ut i de ulike skandinaviske land. I undersøkelsen her ser det ut til at det er andre ledd som avviker fra trekantlinjen. Noen er logiske, gitt at pasientutvalg ofte har noe svakere utdanning. De tre leddene som avviker på Informasjon (5, 12 og 16), kan gjerne kalles «dannelsesleddene» og gjenspeiler interesse for klassisk kunst og litteratur. At noen ledd avviker i vanskegrad, er i prinsippet ikke problematisk hvis de fleste personer testes med de fleste ledd. Ved liberale stoppkriterier vil dette være tilfellet. Når testen skal avsluttes etter tre nullskårer, kan det imidlertid true testens reliabilitet hvis vanskelige ledd kommer tidlig i testen.

Stoppkriterier

I denne undersøkelsen hadde Visuelle puslespill, Informasjon og Ordforståelse ikke tilfredsstillende reliabilitet ved et kuttepunkt på 3 nullpoengssvar. Det er særlig i Visuelle puslespill at dette kan ha konsekvenser for vurderingen av pasienten. I denne delprøven vil endringer på en råskår i flere tilfeller føre til en tilsvarende endring i

skalert skår. Fire av 10 som testes med delprøven, ville ha fått en bedre skalert skår hvis kuttepunktet var satt til 5, mens andelen som ville fått en annen skår, var mer beskjeden når det gjelder Informasjon og Ordforståelse. Nå kan en jo med rette hevde at ved et mer liberalt kuttepunkt ville normene ha sett noe annerledes ut. Klinikeren kan derfor ikke uten videre skåre ut fra et mer liberalt kuttepunkt. Min anbefaling er allikevel å fortsette testing utover det standardiserte kuttepunktet hvis pasienten gjør tre feil tidlig i Informasjon eller Visuelle puslespill. Fremkommer det da store avvik, må man i hvert fall beregne skårene sine med ut fra begge kuttepunkt og vurdere om det er grunnlag for å bruke den høyeste skåren. For et lite antall personer kan skårforskjellen på en enkelt delprøve slå ut med en samlet effekt på IQ skåren på 5–6 poeng, noe som klart er innenfor feilmarginene. Summen av feilmåling ut fra flere tester kan allikevel summere seg opp til å ha en betydning også for samlet GEI eller IQ. Hovedbekymringen er likevel knyttet til diskrepansanalyser. Effekten av feilvurdering av det egentlige nivået på en test kan i verste fall føre til feilslutninger når det gjelder spesifikke lærevansker.

Arbeidsminne

Det er nå godt dokumentert at arbeidsminne kan inndeles både i forhold til modalitet (visuelt/ romlig vs. auditivt) og etter nivå eller kompleksitet (korttidsminne eller enkel lagring vs. bearbeiding). Denne studien viser at arbeidsminnemålene fra Tallspenn og Bokstav-tallserier ikke synes å måle ulike aspekter av arbeidsminne godt nok til at nivådistinksjonen har noe for seg. Når disse delprøvene inngår i konfirmatorisk faktoranalyse sammen med Visuelt spenn fra WMS-III, fremkommer derimot et godt samsvar mellom modalitetsmodellen og observerte skårer. Det er derfor problematisk at WAIS-IV kun måler arbeidsminne i den auditive modaliteten. Vi vil da ikke vite om utfall indikerer spesifikke språklige vansker, dysleksi eller er et uttrykk for mer generelle arbeidsminnevansker. En rekke studier viser at personer med dysleksi skårer svakt på auditive spennprøver og bedre på tester av visuelt spenn (Lasonen, Leppämäki & Hokkanen 2009; Melby-Lervåg, Lyster & Hulme 2012). I forhold til ADHD synes det holdepunkt for at manipulasjons/ bearbeidings-arbeidsminne er mest affisert (Martinussen et al. 2005). Dernest finner man størst utfall i den visuelle modaliteten og minst utfall i auditivt korttidsminne, som altså er det vi primært måler i WAIS. Fokuset på arbeidsminnevansker ved ADHD kan lede klinikeren til å feilslutte at det er en risiko for ADHD ved lav skår på WAIS, der tidligere studier jo egentlig har vist en flat profil på Wechsler- testene (Egeland, Sundberg, Andreassen & Stensli 2006; Lundervold & Sørensen 2008).

Resultater som i likhet med de foreliggende vektlegger betydningen av modalitet i Wechsler- testene, har vært funnet før. I en konfirmatorisk faktoranalyse av hele WAIS-IV og WMS-IV fant Holdnack, Zhou, Larrabee, Millis og Salthouse (2011) at to modeller passet like godt: En syvfaktormodell med separate auditive og visuelle AM-faktorer og en femfaktor modell der AM fremsto som kun én faktor. En KFA av WMS-III fant en AM-faktor bestående av kun Visuelt spenn og Bokstav-Tall-serier (Millis, Malina, Bowers & Ricker 1999).

Usikkerheten som er knyttet til hvordan man skal fortolke Tallspenn og Bokstav-Tallserier, forsterkes av at førstnevnte sammen med Regning utgjør Arbeidsminneindeksen i WAIS-IV. Flere studier av faktorstrukturen i WAIS-III satte spørsmålstegn ved gyldigheten av Regning som del av denne faktoren (Gregoire 2004; Egeland, Bosnes & Johansen 2008). I hvert fall i kliniske utvalg fremsto Regning som en verbal kunnskapstest. Utdanning forklarte 20 % av variansen i Regning, noe som var på nivå med de andre verbalprøvene, men vesentlig forskjellig fra de andre arbeidsminnetestene (Egeland 2008). Evnen til å korttidslagre og simultant bearbeide informasjonen i regneoppgavene var muligens avgjørende for om personer i normalutvalg løste oppgavene, men i kliniske utvalg satte manglende læring av de fire regneartene grenser for testen som mål på arbeidsminnekapasitet. I standardiseringsfaktoranalysen av WAIS-III testet man aldri hvorvidt Regning skulle plasseres i en annen faktor. Det er gjort i WAIS-IV. Både i den opprinnelige versjonen og i den skandinaviske finner man da at modellene som tillater Regning å lade på Verbal forståelses-faktoren gir best samsvar mellom modell og observasjoner. Implikasjonene av denne nye faktorstrukturen i WAIS-IV innarbeides imidlertid ikke i tolkningskapitlene i håndboken, som synes å se bort fra kunnskapselementet som underligger Regning. Arbeidsminnemåling med WAIS-IV synes altså å være problematisk både hvis en fortolker Tallspenn og Bokstav-Tall-serier alene, og når en inkorporerer Regning i en samleindeks. Mitt råd vil være at klinikerer bør vise tilbakeholdenhet med å fortolke AMI som uttrykk for et mulig arbeidsminneproblem eller oppmerksomhetsvanske, særlig hvis klienten har svak utdanning. Videre anbefales det å skaffe seg et mål på visuelt spenn for å kunne kvalitetssikre at svak skår på Tallspenn eller Bokstav-tall-serier er uttrykk for en generelt svekket kapasitet, og ikke kun reduksjon i den auditive modaliteten. Aktuelle prøver som da kan brukes, er Spatialt spenn fra Wechsler Memory Scale eller Knoxblokker (Bornstein 1983). Når det gjelder førstnevnte, fant Bosnes og Troland (2012) at Spatialt spenn fra WMS-III lå på samme nivå som total IQ i deres befolkningsutvalg fra Helseundersøkelsen i Nord-Trøndelag. Tallhukommelse (og dermed også arbeidsminneindeksen) lå imidlertid vesentlig lavere enn i Spatialt spenn og IQ målt med WAIS-III. Dette ledet forfatterne til å advare mot overdiagnostikk av oppmerksomhetsvansker hvis man vektla WAIS alene.

«Historisk har vi flere eksempler på at Wechslertestene oppfører seg annerledes i kliniske utvalg enn forutsatt i manualen, og at dette kan svekke gyldigheten av de kliniske vurderingene som gjøres på grunnlag av testen»

Når det gjelder Knox-blokker som en alternativ test av visuell arbeidsminnekapasitet, må en være klar over at testen har gamle normer, noe som kan svekke sikkerheten i

slutninger om relative visuelle versus auditive vansker. Der det er særlig viktig å mene noe om det, anbefales klinikerer også å undersøke manipulasjonselementet i arbeidsminne med ytterligere krevende tester, så som Paced Auditory Serial Addition Test, samt å undersøke kontrollert oppmerksomhet med tester som setter krav til fokusering av oppmerksomheten gjennom hemming av irrelevant informasjon. Mye ny forskning på arbeidsminne har vært opptatt av at arbeidsminnekapasiteten henger sammen med evne til inhibisjon. Dette måles ikke i noen av WAIS-testene, i den forstand at all informasjon som presenteres, er relevant for oppgaven. Tester som Color Word Interference Test stiller nettopp krav til hemming og er dermed sensitiv for det som gjerne kalles Eksekutiv oppmerksomhet (Egeland 2010). Tallspenn og i noe omfang også Bokstav-Tall-serier, siden den faller sammen med de enklere prøvene i KFA-en, måler trolig mer robust automatisk oppmerksomhetskapasitet.

Generell diskusjon

Denne kliniske utprøvingen av WAIS-IV viser noen utilstrekkeligheter som klinikerer kan ha nytte av å vite om, og som særlig er relevante der man foretar diskrepansanalyser av sterke og svake sider. Samtidig er det klart at få tester er så grundig standardisert både i Skandinavia og USA som nettopp Wechsler-testene. Den høye bruksfrekvensen, samt at det ofte fattes viktige avgjørelser på bakgrunn av testresultatene, gjør at kravene til reliabilitet er særlig store. I tillegg vil klinikere ofte gjøre seg nytte av den omfattende forskningen som omhandler kjennetegn på kliniske grupper eller profiler for spesielle kognitive vansker. Det er da viktig at ikke forhold knyttet til lokale standardiseringer ugyldiggjør slutninger man ellers ville trekke ut fra internasjonal forskningslitteratur. Konkret er det i denne studien påvist at noen verballedd avviker fra optimal gradvis økende vanskegrad, og klinikerer kan ha nytte av å identifisere disse. De mindre avvikene i leddanalysen kan ha bidratt til nedsatt reliabilitet av et strengt stoppkriterium, og klinikerer anbefales å fortsette undersøkelsen utover stoppkriteriet på Visuelle puslespill og Informasjon. At WAIS-IV ikke måler arbeidsminne i flere modaliteter, kan sies å være en designmessig utilstrekkelighet som ikke har sammenheng med den skandinaviske standardiseringen. Klinikerer må fortolke arbeidsminnemålene ut fra begrensningen som ligger i at man måler kun den auditive modaliteten, og være oppmerksom på at modalitet ser ut til å være den viktigste dimensjonen når arbeidsminnet måles med enkle prøver som ikke viser seg å differensiere mellom enkel lagring og manipulasjon.

«Når testen skal avsluttes etter tre nullskårer, kan det imidlertid true testens reliabilitet hvis vanskelige ledd kommer tidlig i testen»

Det er noen begrensninger i undersøkelsen. Det er et hensiktsmessighetsutvalg som ikke er trukket ut fra forhåndsvalgte kriterier og som er dårlig diagnostisk beskrevet. En fordel med dette er imidlertid at det er personer som er rekruttert fortløpende og dermed representerer et tverrsnitt av personer som testes med WAIS i en nevropsykologisk spesialistpraksis. De vil variere både i forhold til grad av kognitive vansker og modalitetsutfall. Kun 20 % var henvist for utredning av lærevansker, og ti prosent hadde traumatisk hodeskade eller hjerneslag. Det er blant disse 30 prosentene en kan forvente å finne noen med høyresidige utfall eller non-verbale lærevansker på den ene siden eller venstresidige utfall eller språklige vansker på den andre siden. Pasienter med slike utfall vil presumptivt forventes å utvise modalitetsforskjeller, men den lille andelen av pasienter med slike vansker er ikke tilstrekkelig til å forklare at modalitetsforskjeller fremstår som den mest potente dimensjonen i utvalget som helhet.

En annen begrensning i undersøkelsen er knyttet til at skåringen av de språklige leddene er gjort av forfatteren. Det ville styrke undersøkelsen hvis data fra flere klinikker hadde inngått i undersøkelsen. Dette er jo en registerstudie, og det hadde vært mulig å hente inn materiale fra andre kolleger, men da uten at man ville få til den systematiske utprøvingen av ulike kuttepunkter. For enkle ledd på Informasjon er det svært lite fortolkning (bare ett rett svar på hovedstaden i Italia!), men i Ordforståelse og Likheter kan man tenke seg at interdommer-reliabiliteten er noe svakere. Avvikene i løsningsfrekvens i dette materialet skyldes imidlertid ikke vanskelig skårbare ideosynkratiske svar, men derimot en høy frekvens av prototypiske 1-poengssvar angitt i skåringsmanualen.

Samlet sett er det etter min vurdering trolig at funnene i denne undersøkelsen kan generaliseres til andre kliniske sammenhenger der man på indikasjon anvender WAIS-IV.

Teksten sto på trykk første gang i Tidsskrift for Norsk psykologforening, Vol 51, nummer 11, 2014, side 911-919

TEKST

Jens Egeland, Klinikk for Psykisk Helse og Rusbehandling, Sykehuset i Vestfold, Psykologisk institutt, Universitetet i Oslo

KONTAKT: jens.egeland@siv.no

+ **Vis referanser**

Referanser

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. I G. Bower (red.), The psychology of learning and motivation (Vol. 8, s. 47-90). San Diego, CA: Academic Press.

Berryman, C, Stanton, T. R., Bowering, K. J., Tabor, A., McFarlane, A. & Moseley, G. L. (2013). Evidence for Working Memory deficits in chronic pain: A systematic review and meta-analysis. Pain, 154, s. 1181-1196.

- Bornstein, R. A. (1983). Construct validity of the Knox cube test as a neuropsychological measure. *Journal of Clinical Neuropsychology*, 5, s. 105-114.
- Bosnes, O. & Troland, K. (2012). Wechsler Memory Scale III og Wechsler Adult Intelligence Scale III utprøvd i et utvalg av HUNT 3 populasjonen. *Tidsskrift for Norsk Psykologforening*, 49, s. 462-467.
- Egeland, J. (akseptert). Measuring Working Memory with Digit Span and Letter- Number span from the WAIS-IV: Risk of underestimating modality effects. *Applied Neuropsychology: Adult*
- Egeland, J. (2008). Status m.h.t. de norske WISC-III og WAIS-III testene. Forslag til retningslinjer for klinisk bruk. Foredrag Norsk Nevropsykologisk Forenings årsmøtekonferanse, november, 2008.
- Egeland, J. (2010). Undersøkelse av oppmerksomhet med vekt på CPT (2010). I Hestad, K. & Egeland, J. (red.). *Klinisk Nevropsykologi*. Trondheim: Tapir Akademisk Forlag.
- Egeland, J., Bosnes, O. & Johansen, H. (2008). Factor Structure of the Norwegian Version of WAIS-III in a Clinical Sample: The Arithmetic Problem. *Assessment*, 16, s. 292-300.
- Egeland, J., Bosnes, O. & Johansen, H. (2009) Utprøving av WAIS-III i et klinisk utvalg. *Tidsskrift for Norsk Psykologforening*, 46, s. 658-667.
- Egeland, J., Sundberg, H., Andreassen, T-H. & Stensli, O. (2006). Reliability and validity of Freedom from Distractibility and Processing Speed Factors in the Norwegian WISC-III-version. *Nordic Psychology*, 58, s. 136-149.
- Gregoire, J. (2004). Factor structure of the French version of the Wechsler Adult Intelligence Scale-III. *Educational and Psychological Measurement*, 64, s. 463-474.
- Holdnack, J. A., Zhou, X., Larrabee, G. J., Millis, S. R. & Salthouse, T. A. (2011). Confirmatory factor analysis of the WAIS-IV/WMS-IV. *Assessment*, s. 178-191.
- Jöreskog, K. & Sörbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Scientific Software International: Lincolnwood, Illinois.
- Lasonen, M., Leppämäki, S. & Hokkanen, L. (2009). Adult Dyslexia and Attention Deficit Disorder in Finland - project DyADD. WAIS-III cognitive profiles. *Journal of Learning Disabilities*, 42, s. 511-527.
- Lundervold, A. & Sørensen, L. (2008). Kognitiv funksjon hos barn med ADHD, eksemplifisert ved WISC-III resultater. *Tidsskrift for Norsk Psykologforening*, 45, s. 1143-1148.
- Martinussen, R., Hayden, J. Hogg-Johnson, S. & Tannock, R. (2005). A meta-analysis of Working Memory impairments in children with Attention Deficit/ Hyperactivity Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44, s. 377-384.
- Melby-Lervåg, M., Lyster, S-A. H. & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, 138, s.322-352.
- Millis, S. R., Alina, A. C., Bowers, D. A. & Ricker, J. H. (1999). Confirmatory factor analysis of the Wechsler Memory Scale- III. *Journal of Clinical and Experimental Neuropsychology*, 21, s. 87-93.
- Vaskinn, A., Egeland, J. (2012). Testbruksundersøkelsen: En oversikt over tester brukt av norske psykologer. . *Tidsskrift for Norsk Psykologforening*, 49, s. 658-665.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale (4. Ed.): technical and interpretative manual*. San Antonio, TX, Harcourt Assessment; 2008.