

# Utprøving av WAIS-III i et klinisk utvalg

Denne artikkelen gir fem råd til brukere av WAIS-III-testen. Det anbefales å unngå delprøven Regning i vurdering av arbeidsminne. Resonnering og Tegneserier ser ut til å være overflødige til annet enn å beregne fullskala IQ. Generell Evne-Indeks (GEI) introduseres som et alternativ til fullskala IQ.

## TEKST

Jens Egeland

Ole Bosnes

Hans Johansen

PUBLISERT 5. juli 2009

## ABSTRACT:

### **Analysis of the Norwegian WAIS-III in a clinical sample**

This study is based on the analysis of 272 protocols from a mixed clinical sample. It identifies five possible pitfalls of the newly adapted Norwegian WAIS-III. The study concludes that: (1) Scores of subjects on «Similarities» with a below average IQ are underestimated on the «Similarities» subtest. (2) Allocation of «Arithmetic» within the factor structure was tested with a confirmatory factor analysis, which showed that a split loading between WM and the VC factor had the best «goodness of fit». (3) «Picture Arrangement» and «Comprehension» are redundant when computing POI and VCI. (4) «Coding» and «Digit Span» had the lowest loading on the g-factor. (5) WMI is sensitive, but not a specific measure of cognitive deficit. PSI seems a more specific measure of cognitive deficit.

Key words: WAIS-III, arithmetic, GAI, working Memory, processing speed

EMNER

WAIS-III

aritmetikk

GAI

arbeidsminne

prosesseringshastighet

Forfatterne ønsker å rette en takk til alle de psykologene i Psykiatrien i Vestfold HF som i en travel hverdag tok seg tid til å bistå prosjektgruppen med WAIS-III protokoller, og som på denne måten har bidratt til forhåpentligvis å kunne høyne kvaliteten i bruk av denne testen i Norge. Artikkelen er skrevet med støtte fra Helse Sør RHF til Jens Egeland.

Den tredje versjonen av Wechsler Adult Intelligence Scale (WAIS-III: Wechsler, 1997a) har nå vært brukt i Norge siden 2003. Testen er en norsk oversettelse og tilpasning av

den britiske versjonen av den opprinnelige amerikanske versjonen. De ulike revisjonene av Wechsler-testene har endret seg fra å være enhetlige intelligens tester til å være mer omfattende testbatterier av kognitiv funksjon i en bredere forstand (Tulsky, Saklofske, Wilkins & Weiss, 2001). Et overordnet IQ-mål og oppdelingen i en Verbal IQ og Utførings-IQ er beholdt, hovedsakelig av historiske grunner. Oppdelingen i fire faktorskårer som måler ulike aspekter ved kognitiv funksjon, fremheves i omtalen av WAIS-III som det bærende element i testen. Testen fremstår altså som et batteri som måler henholdsvis Verbal Forståelse, Perseptuell Organisering, Arbeidsminne og Prosesseringshastighet. Brukt sammen med hukommelsestesten Wechsler Memory Scale versjon 3 (WMS-III) kan klinikerer kartlegge relative styrker og svakheter hos klienten innenfor ytterligere to funksjonsområder, nemlig Verbal hukommelse og Visuell hukommelse (Tulsky & Price, 2003). Utviklingen av Wechsler-testene i retning av mer omfattende testbatterier av ulike funksjoner sees klart i den nyeste revisjonen av barnetesten WISC-IV, der den historiske todelingen i utføringstester og språktester nå er fjernet.

Den norske WAIS-III versjonen (Wechsler, Nyman & Nordvik, 2003) bruker amerikanske normer, mens man i barnetesten WISC-III valgte å lage en mindre svensk normering som også brukes i oversatt versjon i Norge (Ellertsen & Johnsen, 2003). Det er mye som tyder på at WISC-III-normeringen har vært utilstrekkelig. Særlig er det grunn til å hevde at språkprøvene er blitt for strenge, og at det øker risikoen for falskt positive påvisninger av spesifikke språkvansker (Egeland, Andreassen, Sundberg & Stensli, 2006). Et annet funn er at normene er særlig strenge for de eldste barna (Sundberg, Egeland, Andreassen & Stensli, 2006). En aktuell problemstilling for klinikere er om det også hefter systematiske feilkilder ved den nye WAIS-III-versjonen.

I den norske tilpasningen av WAIS-III har man nøydt seg med å foreta en validering av oversettelsen gjennom utprøving på et utvalg av 202 personer mellom 16 og 68 år. Basert på dette materialet ble det foretatt en konfirmatorisk faktoranalyse som bekreftet at den intenderte firefaktorløsningen hadde best statistisk samsvar mellom modell og observerte skårer også i den norske versjonen. Oversettelsen ble utprøvd med såkalte «overitems» for tre av verbalprøvene (Ordforståelse, Likheter og Resonnering). Det vil si at man i utprøvingen av testen hadde med flere ledd enn de som ble beholdt i den endelige versjonen. Leddene i delprøven Informasjon ble rangert etter leddanalyse av løsningsfrekvens i valideringsutvalget.

I klinikken vil WAIS-III hovedsakelig bli brukt overfor klienter der man mistenker en kognitiv vanske. Det er derfor særlig viktig at testen har gode psykometriske egenskaper også når den brukes på klienter som avviker fra gjennomsnittet. Den foreliggende undersøkelsen representerer en slik utprøving av testen i et blandet klinisk utvalg. I artikkelen vil vi fokusere på følgende fem mulige trusler mot validiteten av prøven:

### **1. Verbaltestenes vanskegrad**

Leddanalyser som viser løsningsfrekvens av de enkelte testledd, bør foretas innenfor hele den populasjonen testen skal brukes på. Ideelt sett skal løsningsfrekvensen utgjøre

en rett linje fallende fra 100 % som løser oppgaven til 0 %. Selv om man oppnår en slik fordeling i en normalgruppe der kjente kognitive vansker er eksklusjonskriterium, kan det være at kurven vil ha en annen form i et klinisk utvalg. I normering av tester er det viktig å undersøke friske normalpersoner. I valideringssammenheng er det også viktig å sjekke om kurven for løsningsfrekvens av oppgaver har den samme formen i et klinisk utvalg, selv om prosenten av oppgaver som løses, da vil kunne være lavere. En L-formet kurve, altså der løsningsfrekvensen faller bratt fra et høyt til et lavt nivå for senere å holde seg ganske jevn, kan føre til at klienter med lette vansker fremstår som for svake, mens klienter med mer omfattende vansker ikke blir tilsvarende undervurdert. En W-formet kurve vil svekke reliabiliteten av testen.

I den foreliggende undersøkelsen presenterer vi løsningsfrekvens av enkeltleddene i Likheter, Ordforståelse, Resonnering og Informasjon i et blandet klinisk utvalg som trolig er representativt for pasientpopulasjonen testen vil brukes overfor. Gjennom hyppig og langvarig bruk av WAIS-III har vi et klinisk inntrykk av at personer som skårer noe under gjennomsnittlig IQ, får uforholdmessig svake skårer på Likheter sammenlignet med de andre språklige prøvene. For å teste dette statistisk har vi subtrahert skalert skåre på Likheter fra gjennomsnittet av skalert skåre på de øvrige tre verbale prøvene. Vi har så delt utvalget i to grupper basert på hvorvidt total IQ er over eller under 90, og sammenlignet forskjellsskåren mellom Likheter og de andre språkprøvene i de to gruppene.

## **2. Faktorstrukturen: Problemet med deltesten Regning**

Etter hvert som WAIS har blitt et kognitivt testbatteri og ikke bare en evneprøve, er det viktig at testene som lader på de enkelte faktorene, er mest mulig «rene». Mens delprøver med høy g-faktor tidligere ble vektlagt som særlig relevante i en evnetest (Lezak, 1995), søker en nå å unngå tester som lader på flere faktorer. Ut fra dette resonnementet inngår ikke Tegneserier i noen av faktorindeksskårene. Det er imidlertid kjent at også Regning, som inngår i indeksen for Arbeidsminne (AMI), korrelerer høyt med verbaltestene (Arnau & Thompson, 2000), fordi de språklig formulerte regneoppgavene stiller store krav til verbal forståelse. Korrelasjonen mellom regning og språktestene kan også skyldes en tredjevariabel, i den forstand at gode språklige ferdigheter og tekniske regneferdigheter er særlig påvirket av skolegang.

**«Mye tyder på at WISC-III-normeringen har vært utilstrekkelig, og at språkprøvene er blitt for strenge. Det øker risikoen for falske påvisninger av spesifikke språkvansker»**

Det er altså mye som taler for at Regning måler noe annet enn Arbeidsminne. Dette kan komme enda klarere frem i et klinisk utvalg der klientene gjennomgående har noe svakere utdanning enn et normalutvalg. I denne studien har vi foretatt en konfirmatorisk faktoranalyse (KFA) hvor vi har variert plasseringen av Regning, enten i indeksen for Verbal Forståelse (VFI) eller i indeksen for Arbeidsminne (AMI). Vi har også testet en modell der Regning lader på begge faktorene samtidig, noe som blant annet Tulsy and Price (2003) og Arnau and Thompson (2000) fant ga best samsvar mellom data og faktorstruktur i den amerikanske normeringen. En mer omfattende KFA av den norske WAIS-III-versjonen er publisert annetsteds (Egeland, Bosnes & Johansen, 2008).

### **3. Er Resonnering og Tegneserier overflødige i faktorindeksene?**

Faktoranalysene basert på standardiseringsutvalgene til den amerikanske, canadiske, spanske og franske versjonen viser alle gjennomgående at delprøvene Resonnering og Tegneserier lader høyt på henholdsvis Verbal Forståelse og Perseptuell Organisering (Garcia, Ruiz & Abad, 2003; Gregoire, 2004; Saklofske, Hildebrand & Gorsuch, 2003; Wechsler, 1997a). Basert på det foreliggende kliniske materialet har Egeland et al. (2008) vist det samme for den norske versjonen. Like fullt er disse prøvene ikke tatt med når faktorindeks-skårene til disse faktorene skal beregnes. Dette er fordi de har vist seg å være overflødige i den forstand at de ikke endrer verdiene på faktorindeksskårene og ikke reduserer standardfeilen til gjennomsnittsverdiene. Det vil si at målene ikke mister reliabilitet om deltestene utelates. Samtidig representerer det en betydelig tidsgevinst for klinikerer ved at hun eller han kan slippe å gjennomføre disse to tidkrevende delprøvene. Vi avklarer hvorvidt dette er forsvarlig med den nye norske versjonen. Vår hypotese er at prøvene er unødvendige.

### **4. Et enklere evnemål: Utprøving av «General Evne Indeks»**

Et nytt mål kalt «Generell Evne Indeks» (General Ability Index: Tulsy et al., 2001; Iverson, Lange, Viljoen & Brink, 2006) har vært lansert som et alternativ til IQ-målet. GEI beregnes ut fra skårene på de seks delprøvene som utgjør Verbal Forståelses-Indeksen (VFI) og Perseptuell Organiserings-Indeksen (POI). Tilhengerne av GEI mener at det er unødvendig å inkorporere delprøvene Resonnering og Tegneserier i en totalskåre når den opprinnelige standardiseringsstudien av WAIS-III viste at de var overflødige i forhold til beregning av henholdsvis VFI og POI-skårene. For klinikerer, som stadig må balansere mellom praktiske og ideelle hensyn når det gjelder hvilke tester hun eller han gjennomfører, blir det vanskelig å forsvare bruken av to delprøver som bare inngår i totalskåren.

Psykologien har i mange årtier stridd med diskusjonen om hva intelligens *egentlig* er. Et synspunkt er at inkorporeringen av tempo- og oppmerksomhetstester utvanner IQ-begrepet, som opprinnelig vektla hensiktsmessig atferd og rasjonell tenkning (Wechsler, 1939). GEI vil kunne hevdes å ligge nærmere denne opprinnelige forståelsen. Alternativet til WAIS-III kan være å bruke kortversjonen Wechsler Abbreviated Scale of Intelligence (WASI: Psychological Corporation, 1999; Ørbeck & Sundet, 2007), som jo

følger dette resonnementet og ikke har noen spesifikke oppmerksomhets- eller tempoprøver. Tulsy et al. (2001) foretok en faktoranalyse av WAIS-III der det ble lagt inn et krav om kun én faktor, og fant da at de seks delprøvene som inngikk i GEI, korrelerte høyere med denne ene faktoren enn det koding og tallhukommelse gjorde. De andre delprøvene som inngår i AMI og PHI (Bokstav-Tall-Sekvensering og Symbolleting), inngår ikke i total IQ-beregningen. I denne undersøkelsen replikerer vi Tulsy og kollegers (2001) enfaktoranalyse. Finner vi at Koding og Tallhukommelse korrelerer lavere med enfaktoren enn de nevnte seks deltestene gjør, vil det styrke validiteten av GEI som et generelt evnemål.

## **5. Faktorprofiler**

Siden faktortenkningen har fått en fremtredende plass i forståelse og tolkning av WAIS-III, er det viktig å sjekke validiteten av indeksskårene. Den faktoren som i WAIS-II kalles «Arbeidsminne», kalles i WISC-III «Oppmerksomhet». En rekke kritikere har hevdet at denne skåren neppe er et valid mål på oppmerksomhet, arbeidsminne eller distraktibilitet. En undersøkelse av hvorvidt kliniske grupper med faktiske oppmerksomhetsvansker har spesifikke vansker på denne skåren, fikk nettopp tittelen «The third factor of the WISC-III: It's (probably) not freedom from distractibility» (Reinecke, Beebe & Stein, 1999). Egeland et al. (2006) fant at ADHD-gruppen ikke hadde svakere oppmerksomhetsskåre enn testresultater for øvrig i den norske WISC-III-versjonen. De individuelle klassifikasjonene viste at cirka hvert fjerde barn med ADHD hadde signifikant svakere oppmerksomhetsskåre sammenlignet med enten VFI eller POI. Forekomsten av signifikant svakere oppmerksomhetsskåre sammenlignet med VFI var imidlertid høyere blant barna med ikke-verbale lærevansker/autisme. Tilsvarende var også frekvensen av signifikant svakere oppmerksomhet enn POI-skåre større blant barna med spesifikke språklige lærevansker enn i ADHD-gruppen. Dette tyder på at oppmerksomhetsskåren i WISC-III har lav begrepsvaliditet, og det er derfor av interesse å se om vi finner det samme i WAIS-III. I den foreliggende studien har vi derfor undersøkt hvorvidt pasienter med ADHD har signifikant lavere AMI sammenlignet med GEI. For å teste spesifisiteten av et eventuelt spesifikt utfall i arbeidsminne har vi foretatt tilsvarende analyse også i andre kliniske grupper med nevrobiologiske forstyrrelser eller hjerneskade. Tilsvarende undersøkte vi hvorvidt PHI var svakere enn GEI i de samme gruppene.

## **Metode**

### **Utvalg**

I undersøkelsen har totalt 272 WAIS-III protokoller vært samlet inn fra to distriktspsykiatriske sentre og Psykiatrisk Fylkesavdeling i Psykiatrien i Vestfold HF, fra Sykehuset Namsos HF, Spesialsykehuset for Rehabilitering HF avdeling Stavern og Habiliteringssenteret i Sykehuset i Vestfold HF. Flertallet av pasientene er rekruttert fra forfatternes avtalepraksiser i henholdsvis Namsos og Tønsberg.

Kopier av komplette svarark fra 75 av WAIS-III-protokollene fra PIV HF har inngått i leddanalysene av de språklige prøvene. Norsk Samfunnsvitenskapelig Datatjeneste, som er personvernombud for Psykiatrien i Vestfold HF, har gitt tillatelse til innsamlingen av anonymiserte testprotokoller med diagnose påført av psykologen som gjennomførte testingen.

**«Undersøkelsen bekreftet mistanken om at Regning ikke primært er en testav arbeidsminne, men lader tungt på Verbal Forståelsesfaktoren»**

Tabell 1 viser kliniske og demografiske kjennetegn ved det samlede utvalget. ADHD-diagnosene fulgte DSM-IV-TR-kriteriene (American Psychiatric Association, 2000) med krav om dokumentert oppmerksomhetssvikt og, for ADHD-kombinert type, også dokumentert impulsivitet/hyperaktivitet i daglig funksjon med symptomstart før fylte 7 år. Hjerneskadegruppene hadde alle positive funn på billeddiagnostiske undersøkelser (CT eller MR). Schizofrenispektrum-diagnosene var satt gjennom konsensus i fagteam etter utredning. Rusrelaterte diagnoser var i hovedsak satt etter utredning ved ruspoliklinikk eller behandlingsopphold ved avdelinger for avhengighetsbehandling (i noen tilfeller også i henvisning fra fastlege/psykiater). Detaljene i diagnostiseringen av depresjon, personlighetsforstyrrelse, somatisk lidelse med psykiske symptomer og angstdiagnoser er i stor grad ukjent, da disse diagnosene var påført WAIS-protokollene ved oversending fra klinikere i psykisk helsevern til førsteforfatteren av denne artikkelen. Mens diagnosene så langt er uavhengig av utredningen med WAIS-III, var diagnostikken av de tre lærevanske-diagnosene ikke uavhengig av WAIS-III-resultatet. Typisk var imidlertid diagnosen generelle lærevansker satt etter gjentatt testing med ulike evneprøver, og språklige lærevansker satt etter utredning med et bredt nevropsykologisk testbatteri. Ikke-verbale lærevansker ble diagnostisert i tråd med Pelletier, Ahmad & Rourkes (2001) klassifikasjonsregler for sikker eller sannsynlig NLD (non-verbal learning disability).

**TABELL 1. DIAGNOSE OG DEMOGRAFISKE DATA**

	Antall (mann/ kvinne)	Alder (s.a.)	Utdanning			FIQ (s.a.)
			% grunnskole	% videregående	% høgskole	
ADHD-K	20 (16/4)	27 (9)	37	53	10	91 (15)

**TABELL 1. DIAGNOSE OG DEMOGRAFISKE DATA**

ADHD-U	11(7/4)	24 (6)	64	36	0	84 (8)
Ikke fokal hjernedysfunksjon	21(16/5)	39 (15)	35	45	20	87 (10)
Høyresidig hjerneskade	8(6/2)	46 (16)	25	25	50	91 (10)
Venstresidig hjerneskade	5 (3/2)	51 (8)	40	20	40	91 (10)
Språklige lærevansker	8 (6/2)	25 (8)	88	12	0	83 (6)
Ikke-verbale lærevansker	7 (4/3)	24 (7)	50	50	0	80 (6)
Generelle lærevansker	6 (2/4)	20 (5)	100	0	0	68 (7)
Schizofreni-spektrum	14 (9/5)	34 (10)	38	50	12	86 (13)
Depresjon	6 (4/2)	35 (17)	33	33	33	90 (18)
Rusavhengighet	9 (7/2)	24 (7)	50	50	0	83 (16)
Somatisk lidelse med psykiske symptomer	122 (69/53)	41 (15)	38	43	19	91 (13)
PF/angst	33 (16/17)	27 (10)	81	19	0	78 (12)
Normal funksjon	2 (1/1)	25 (8)	0	50	50	93 (14)
Total	272 (166/106)	35 (15)	46,6	38,6	14,8	87 (13)

Underutvalget på 75 personer som inngikk i leddanalysene, hadde en gjennomsnittlig fullskala IQ på 87 (s.a. 13). Aldersspennet i dette underutvalget var fra 16 til 62 år, med et gjennomsnitt på 30.8 år (s.a. 11.9). Skalerte skårer på de fire språktestene i denne

undergruppen var: Ordforståelse 8.1 (s.a.2.2) Likheter 7.5 (s.a. 2.9) Informasjon 8.6 (s.a. 2.5) og Resonnering 7.8 (s.a. 3.1). Den diagnostiske sammensetningen av leddanalyseutvalget var det samme som for det store utvalget, med unntak av at det var færre personer med somatiserende lidelser (kun 3 personer) og ingen personer med påvist venstresidig hjerneskade eller generelle lærevansker.

## Data-analyser

Alle analyser bortsett fra den konfirmatoriske faktoranalysen (KFA) ble foretatt med SPSS 12.0. Variansanalyse (ANOVA) ble brukt for å sammenligne forskjellen mellom skalert skåre på Likheter og gjennomsnittlig skalert skåre på de andre testene som inngår i VFI hos pasientgruppene med Total IQ henholdsvis over og under 90.

Den konfirmatoriske faktoranalysen ble foretatt med LISREL 8.3-programmet (Jöreskog & Sörbom, 1993) og ble basert på råskårene fra alle delprøver unntatt Tegneserier. Tegneserier ble utelatt fordi den verken inngår i noen av de fire faktorindeksskårene eller i den samlede IQ-beregningen. En rekke ulike statistiske mål kan beskrive samsvaret mellom modell og observerte skårer. I tråd med Gregoire (2004) og Keith (2005) har vi valgt å rapportere følgende mål: khi-kvadrat (<sup>2</sup>), 2 / d.f., GFI (Goodness of Fit Index), AGFI (Adjusted Goodness of Fit Index), CFI (Comparative Fit Index), RMSEA (Root mean square error of approximation).

Følgende modeller ble analysert: (1) Firefaktormodellen som er beskrevet i håndboken, (2) en firefaktormodell der regning ble plassert i Verbal Forståelses-faktoren og (3) en modell der Regning ble tillatt å lade både på VFI og AMI.

For å avklare hvorvidt Resonnering og Tegneserier styrker VFI og POI, har vi beregnet disse to faktorskårene basert både på tre delprøver slik håndboken foreskriver, og på alle fire delprøvene som lader på faktorene. De to variantene av hvert mål er så korrelert med to Pearson produkt-moment-korrelasjoner i tillegg til at standardfeilen til gjennomsnittene sammenlignes.

IQ pretenderer å være et mål på en generell evne som tilsvarende en g-faktor i en faktoranalyse. Vi foretok derfor en prinsippal komponentanalyse av de enkelte deltestene som inngår i IQ-beregningen, og begrenset faktorløsningen til kun én faktor, for å få en rangering av de ulike deltestenes ladning på g-faktoren.

Forskjellen mellom GEI og AMI samt GEI og PHI ble undersøkt med t-test for parvise sammenligninger.

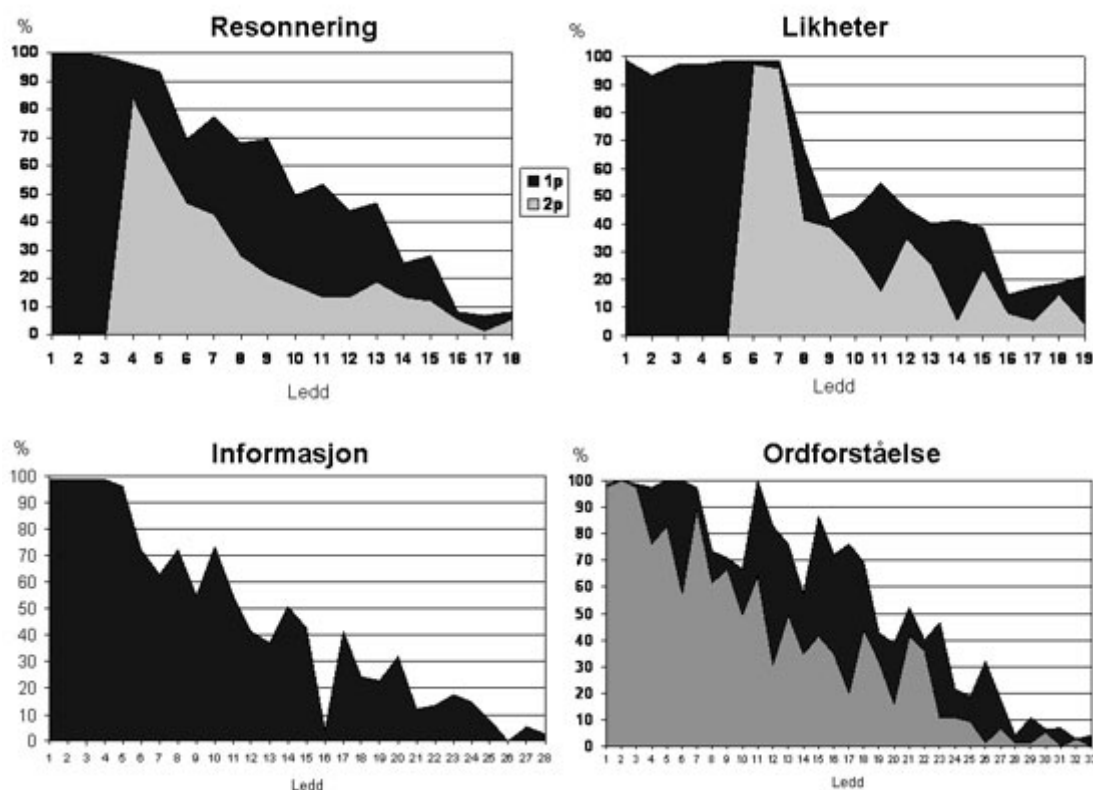
## Resultater

### 1. Verbaltestenes vanskelighetsgrad

Figur 1a–d viser løsningsfrekvens for de fire språkprøvene. Mørkt felt markerer antallet 1 poengsvar, mens det lysere feltet markerer 2 poengsvar. De tre første leddene på Resonnering og de fem første leddene på Likheter gis kun 1 poengsvar. Ideelt sett skal kurven falle jevnt fra 100 % løsningsfrekvens til nær 0 %. Med noen svingninger over og



under en slik ideell kurve ser det ut til at Resonnering, Ordforståelse og Informasjon i grove trekk har en slik fordeling. Noen enkeltledd fremstår imidlertid som mer vanskelige eller lette enn posisjonen i testen tilsier. For Informasjon foreskriver håndboken at Antarktisk skal telles med som et kontinent i ledd nummer 15, men ingen i utvalget klarte dette leddet. Når det gjelder Ordforståelse, er begrepet «starte» et lett tilgjengelig synonym for «begynne» for de aller fleste testpersonene. «Empati» og «delegere» hadde også en høyere svarfrekvens enn ventet i utvalget. På Likheter fremkom en L-formet løsningsfrekvenskurve. Nittiseks prosent av utvalget hadde to poengsvar på ledd 7, mens dette falt til 41 % med riktig to poengsvar på ledd 8. Deretter flatet kurven ut.



**Figur 1.** Løsningsfrekvens av ledd på de språklige delprøvene. NOTE: Prosentandel av forsøkspersoner (n=70) med 2-poengsskårer (lyst felt) og 1-poengsskårer (mørkt felt) kumulativt lagt til andelen med full skåre.

Skalert skåre på Likheter for de 116 personene i utvalget som hadde Total IQ <90, var 5.8 (s.a.2.0) mot 9.4 (s.a.2.7) for de 111 personene med IQ >90. Forskjellskåren mellom Likheter og de tre andre språkprøvene var henholdsvis -1.24 skalert skåre for gruppen med IQ <90 og -0.78 for de med IQ >90. Denne gruppeforskjellen i avvik mellom Likhetskåren og de andre VFI-delprøvene var signifikant på 0.05-nivået (altså gjorde gruppen med lavest IQ det relativt svakest på likheter).

## 2. Problemet med Regning

KFA-en viste at både den modellen der Regning ladet på VFI, og modellen der Regning var splittet på VFI og AMI, hadde lavere khi-kvadrat enn håndbokens fire-faktormodell.

Det vil si at forskjellen mellom modell og data var minst for disse modellene. Siden modellen med splittet ladning for Regning har en frihetsgrad mindre enn der Regning er plassert i VFI, blir khi-kvadratet dividert med frihetsgradene lavest for denne modellen. Ved samme forklaringsverdi for de ulike modellene foretrekkes den enkleste modellen, noe som gjenspeiles i mål som ser på både khi-kvadrat og frihetsgrader. De øvrige målene angir samsvar mellom modell og data, slik at høye verdier gir det beste resultat. GFI, AGFI og CFI over 0.9 regnes som akseptabelt. Jo lavere verdi av RMSEA, desto bedre samsvar. Verdier under 0.08 regnes som akseptabelt, mens verdier under 0.06 er uttrykk for et godt samsvar. Samlet viser KFA-en at håndbokens firefaktormodell gir et akseptabelt samsvar, men at begge de alternative modellene er bedre. Modellen med splittet ladning både på VFI og AMI gir best samsvar med data.

**TABELL 2. MÅL PÅ SAMSVAR MELLOM TO FIREFAKTORMODELLER OG OBSERVERTE SKÅRER I 272 KLINISKE WAIS-III PROTOKOLLER**

<b>Modell</b>	<b>2 (d.f.)</b>	<b>2 /d.f.</b>	<b>GFI</b>	<b>AGFI</b>	<b>CFI</b>	<b>RMSEA</b>
Håndbokens 4-faktormodell	144.26 (59)	2.4	0.92	0.88	0.98	0.074
Alternativ plassering av Regning i VF <sup>1</sup>	114.54 (59)	1.9	0.94	0.90	0.99	0.059
Splittet ladning for Regning på AM <sup>2</sup> og VF	100.54 (58)	1.7	0.95	0.91	0.99	0.052

NOTE: <sup>1</sup>VF = Verbal Forståelses-faktoren <sup>2</sup>AM = Arbeidsminne-faktoren

### 3. Er Resonnering og Tegneserier unødvendige?

Tabell 3 viser POI og VFI beregnet slik håndboken foreskriver, samt tilsvarende verdier basert på fire delprøver (POI + Tegneserier og VFI + Resonnering). Det er 0.3 poeng i forskjell mellom de to alternative beregningene for hver av indeksskårene, hvilket ligger godt innenfor 95 % konfidensintervallene til gjennomsnittet for POI og VFI (avviket er kun henholdsvis 8 og 9.5 % av 95 % konfidensintervallet). Korrelasjonen mellom de to måtene å beregne POI og VFI (R i tabell 3) er like under 1.0.

**TABELL 3. SAMMENLIGNING AV POI1 OG VFI2 BASERT PÅ TRE ELLER FIRE DELPRØVER (N=272)**

	<b>M (s.a.)</b>	<b>s.f</b>	<b>95 % k.i</b>	<b>R</b>
POI	95,7 (15,5)	.94	93,9-97,6	
POI+tegneserier	96,0 (12,6)	.77	94,5-97,5	.97
VFI	88,9 (13,1)	.80	87,3-90,5	.98
VFI+resonnering	89,2 (11,9)	.72	87,8-90,3	

**TABELL 3. SAMMENLIGNING AV POI1 OG VFI2 BASERT PÅ TRE ELLER FIRE DELPRØVER (N=272)**

NOTE: 1 POI = Perseptuell organiserings-indeks; 2 VFI = Verbal Forståelses-indeks.

#### 4. Utprøving av GEI

G-faktoren som fremkom i den prinsipale komponentanalysen, hadde en Eigen-verdi på 5.8 og forklarte 54,1 % av variansen i skårene på de 11 prøvene som inngår i IQ beregningen. I tabell 4 er disse delprøvenes ladning på g rangert. Mens faktorladningen angir hvor mye testen korrelerer med g, angir kommunaliteten hvor stor andel av variansen i delprøven som deles av andre delprøver. Som forventet ladet Koding og Tallhukommelse lavest på g og hadde også lavest delt varians med andre delprøver.

**TABELL 4. DELTESTENE SOM INNGÅR I IQ-BEREGNINGENS LADNING PÅ EN FAKTOR (N=272)**

	<b>Faktorladning</b>	<b>Kommunalitet</b>
Ordforståelse	.82	.67
Likheter	.80	.64
Resonnering	.78	.61
Informasjon	.78	.61
Matriser	.76	.57
Regning	.75	.56
Terningmønster	.71	.50
Tegneserier	.67	.45
Billedutfylling	.66	.44
Tallhukommelse	.64	.41
Koding	.55	.30

NOTE: Eigen-verdi: 5.8. Forklart varians: 54.1 %.

#### 5. Faktorprofiler

Tabell 5 viser de ulike diagnosegruppene resultater på GEI, AMI og PHI. Alle gruppene unntatt pasientene med ikke-verbale lærevansker skårer signifikant svakere på AMI enn på GEI. PHI var signifikant svakere enn GEI for begge ADHD-gruppene, diffus og høyresidig hjerneskade, somatiske plager med psykiske symptomer,

schizofrenispektrumlidelser og for pasientene med psykiske lidelser i form av personlighetsforstyrrelse eller angst.

**TABELL 5. SAMMENLIGNING AV GEI1, AMI2 OG PHI3 I DE ULIKE DIAGNOSEGRUPPENE (N = 272)**

	<b>GEI</b>	<b>AMI</b>	<b>PHI</b>	<b>GEI-AMI</b> t pX		<b>GEI-PHI</b> t pX	
ADHD-K (n = 17)	92.0 (11.3)	81.8 (11.8)	84.0 (10.8)	5.6	<.001		
ADHD-U (n = 11)	99.9 (6.3)	80.4 (11.8)	80.0 (7.2)	4.0	.002	3.2	.010
Diffus hjerneskade (n = 20)	91.3 (9.9)	83.7 (11.8)	82.5 (11.6)	2.8	.012	3.8	.001
Høyresidig hjerneskade (n = 8)	97.9 (12.8)	85.1 (6.6)	77.9 (7.9)	2.6	.033	4.7	.020
Venstresidig hjerneskade (n = 5)	93.4 (10.8)	87.8 (10.2)	88.4 (11.0)	3.0	.041	0.7	i.s.
Språklige lærevansker (n = 8)	90.6 (6.4)	78.6 (8.3)	80.9 (7.1)	4.4	.003	2.7	.029
Ikke-verbale lærevansker (n = 6)	83.3 (5.1)	84.3 (3.9)	76.5 (8.1)	0.6	i.s.	2.5	i.s.
Generelle lærevansker (n = 5)	75.6 (6.2)	63.8 (7.3)	72.0	4.2	.014	0.6	i.s.
Shizofrenispektrum (n = 13)	88.9 (9.7)	81.5 (10.9)	80.9 (10.1)	4.2	.001	3.5	.004
Depresjon (n = 6)	93.7 (15.7)	85.7 (17.1)	87.3 (15.6)	3.8	.012	1.8	i.s.
Rus (n = 8)	88.9 (17.4)	79.5 (16.9)	85.0 (13.5)	2.9	.023	1.0	i.s.
Somatisk lidelse med psykiske symptomer (n = 116)	94.6 (11.7)	89.1 (13.8)	86.1 (13.0)	5.8	<.001	7.1	<.001
Personlighetsforst./angst (n = 32)	83.6 (10.2)	74.2 (10.4)	79.7 (11.9)	5.1	<.001	1.8	i.s.

**TABELL 5. SAMMENLIGNING AV GEI1, AMI2 OG PHI3 I DE ULIKE DIAGNOSEGRUPPENE (N = 272)**

Alle pasienter	91.4 (11.8)	84.2 (13.6)	83.4 (12.3)	11.8	<.001	10.8	<.001
----------------	----------------	----------------	----------------	------	-------	------	-------

## Diskusjon

Studien omhandler fem problemstillinger som vedrører psykometriske egenskaper ved den norske WAIS-III-versjonen. Før publisering ble testen utprøvd på en gruppe friske kontroller. Ut fra løsningsfrekvens har man da rangert leddene på de oversatte verbale deltestene og sjekket ut i hvilken grad nye oppgaver kunne brukes. Selv om testen har gode psykometriske egenskaper for personer som varierer nær middelveiden, er det en fare for at den kan miste disse egenskapene for personer lenger ute i variasjonsområdet. Derfor er det viktig også å undersøke hvordan den fungerer i et klinisk materiale, slik vi har gjort her. Det uforholdsmessige store fallet i løsningsfrekvens på delprøven Likheter for personer som samlet skårer noe under middelveiden, er et eksempel på en slik endring i psykometriske egenskaper. Selv om det altså er viktig å prøve ut testen i et klinisk utvalg, er det åpenbare begrensninger i hvilke konklusjoner som kan trekkes på bakgrunn av et slikt utvalg: Vi kan ikke utsi noe om normene, verken for Fullskala IQ-målet eller for GEI. Forsøket på å teste validiteten av faktorindeksene gjennom å se på profilen i kliniske grupper er problematisk, så lenge vi ikke kan dokumentere at friske kontroller har en flat profil som er forskjellig fra klientene. Med disse forbeholdene in mente viste undersøkelsen følgende:

### 1. Verbalprøvenes vanskegrad

Den visuelle presentasjonen av løsningsfrekvens av leddene i de språklige prøvene synes i hovedsak tilfredsstillende for Resonnering, Ordforståelse og Informasjon, selv om noen enkeltledd avviker fra en jevnt fallende kurve. Uten at det får betydelig konsekvens for resultatene på testen, vil vi peke på at ingen av personene i dette utvalget navngir Antarktis som et kontinent, og skal da etter håndboken få nullskåre. Dette betyr at testleddet ikke burde vært med. Det er et eksempel på hvor galt det kan gå når man oversetter verbale deltester og ikke tar hensyn til kulturelle forskjeller. Også Gyldendals Store Konversasjonsleksikon (Bull, Eskeland & Tandberg, 1972) opererer med fem verdensdeler. De åpner imidlertid opp for at «Antarktis kan regnes som en verdensdel, om enn uten innbyggere». Vi vil foreslå at Harcourt erstatter det nåværende leddet med det leddet som ble brukt i den forrige norske WAIS-versjonen. Løsningsfrekvensen for Likheter falt bratt på et tidlig nivå, noe som gjør at personer som har en Total IQ kun noe under middelveiden, allikevel vil få en svak skåre på denne prøven. I noen screeningsammenhenger kan det være aktuelt å velge ut kun én språklig delprøve for å få et grovt mål på språklig kognitiv funksjon. I vår undersøkelse var den skalerte skåren på Likheter noe lavere enn gjennomsnittet av de andre tre VF-prøvene både for de med IQ over og under 90 poeng. For personene i den svakeste

gruppen ville et grovt IQ-estimat basert på kun Likheter gjennomsnittlig undervurdere dem med 6.5 IQ-poeng sammenlignet med hvis de andre verbalprøvene ble brukt.

## **2. Problemet med Regning**

Undersøkelsen bekreftet mistanken om at Regning ikke primært er en test av arbeidsminne, men lader tungt på Verbal Forståelses-faktoren. Tidligere har flere studier på normalutvalg vist at Regning fremstår mer som en g-prøve enn som en egentlig arbeidsminneprøve (Arnau & Thompson, 2000; Bowden et al. 2006; Gregoire, 2004). Så langt vi kjenner til, har imidlertid kun to konfirmatoriske faktoranalyser systematisk utprøvd samsvar mellom modell og data for alternative plasseringer av testen. Arnau og Thompson (2000) fant bedre modellsamsvar når både Tegneserier og Regning fikk lade på Verbal Forståelses-faktoren i en 2. ordens KFA. Også Tulskey og Price (2003) fant i sin KFA på WAIS-III og WMS-III samlet at Regning ladet både på verbal forståelse og arbeidsminnefaktoren. Deres AMI besto imidlertid også av Spatialt Minnespenn fra WMS-III. Når vi finner at Regning lader mer på VFI enn på AMI i et klinisk materiale, kan dette skyldes at pasientene i utvalget har mindre utdanning enn gjennomsnittet i befolkningen, og at manglende *kunnskaper* mer enn manglende arbeidsminne begrenser prestasjonene. Regresjonsanalyser i Egeland et al. (2008) viste at utdanningsnivå forklarte mellom 61 og 47 prosent av variansen i prestasjonene på de fire språklige delprøvene. Derimot hadde utdanning mindre å si for de øvrige prøvene (varierende fra 29 til 12 % forklart varians). Regning grupperte seg imidlertid sammen med språkprøvene, idet utdanning forklarte 45 % av variansen.

### **«Er resonnering og Tegneserier unødvendige? Svaret er et ubetinget ja»**

Det er to implikasjoner av dette funnet. For det første er vi enige med Gregoire (2004), som foreslår at Regning erstattes med en renere arbeidsminnetest i neste revisjon av WAIS. Inntil det skjer, foreslår vi at klinikere viser tilbakeholdenhet i å tolke AMI-skåren som uttrykk for arbeidsminne. Der det er viktig å undersøke arbeidsminne, kan de enten komplettere undersøkelsen ved å bruke andre mål på denne funksjonen, eventuelt nøye seg med et estimat basert på Tallspenn og Bokstav-Tall-Sekvensering. Ulempen med det siste er at begge testene måler kapasiteten innenfor den auditive modaliteten. Personer med spesifikke auditive vansker kan derfor feilfortolkes til å ha en allmenn dysfunksjon i arbeidsminnet. Et alternativ som vi ikke selv har prøvd ut systematisk, er å bruke skåringsprogrammet for demografisk korrigerede faktorindeksskårer som utgiveren av WAIS-III har lagt ut til fri nedlasting på sin hjemmeside ([Harcourtassessment.com](http://Harcourtassessment.com)). Arbeidsminneindeksen her er basert på én auditiv og én visuell test, nemlig Bokstav-Tall-sekvensering fra WAIS-III og Spatialt spenn fra WMS-III.

## **3. Er Resonnering og Tegneserier unødvendige?**

Dette spørsmålet kan besvares med et ubetinget ja. I tråd med hva som er funnet i det amerikanske standardiseringsutvalget, finner vi i denne undersøkelsen at forskjellen mellom en VFI basert på tre eller fire tester er forsvinnende liten. Det samme gjelder forskjellen når man inkluderer Tegneserier i POI eller ikke.

#### **4. Utprøving av GEI**

Siden Tegneserier og Resonnering ikke er nødvendig for faktor-indeksene, er gevinsten ved å ta i bruk GEI desto større, siden disse to prøvene ellers bare har sin berettigelse i et samlet IQ- mål. I tråd med Tulsky et al. (2001) finner vi også at delprøvene Koding og Tallhukommelse lader lavest på det som fremstår som en g-faktor, når vi faktoranalyserer de delprøvene som inngår i IQ. Som nevnt tidligere ligger GEI innholdsmessig nær opp mot IQ- målet slik det fremstår i Wechsler Abbreviated Scale of Intelligence (Psychological Corporation, 1999; Ørbeck & Sundet, 2007) ved at heller ikke det målet har noen hastighets- eller oppmerksomhets-delprøver. Fordelen er at det samlede evnemålet (GEI eller IQ i WASI) trolig er mindre sensitivt i forhold til ervervet funksjonsnedsettelse. Det foreligger nå en rekke studier som viser temporær nedsettelse av arbeidsminne som følge av eksempelvis hormonendringer ved slanking (Kemps, Tiggemann & Marshall, 2005) eller støyeksponering nær flyplasser (Matheson, Stansfeld & Haines, 2003). En mer langvarig eller kronisk nedsettelse av arbeidsminne fremkommer etter hjernerystelser (Bernstein, 2002), ved depresjon (Landrø, Stiles & Sletvold, 2001), schizofreni (Goldman-Rakic, 1994), begynnende Alzheimers sykdom (Baddeley, Bressi, Della Sala, Logie & Spinnler, 1991) eller multippel sklerose (Parmenter, Shucard-Janet, Benedict & Shucard, 2006). Tilsvarende er redusert prosesseringshastighet et vanlig symptom ved lette hodeskader (Spikman, van Zomeren & Deelman, 1996) og depresjon (Egeland et al., 2003). Det synes hensiktsmessig å bruke uttrykket «kognitive utfall» om observasjonene ovenfor, men de fleste vil nok si at det er misvisende å si at personene med slike vansker er blitt mindre intelligente.

#### **5. Faktorprofiler**

I studiene ovenfor er reduksjon i arbeidsminne og prosesseringshastighet dokumentert ved hjelp av andre mål enn de som inngår i AMI og PHI. Den femte og siste problemstillingen vi reiste om WAIS-III, gjaldt nettopp spørsmålet om det fremkommer signifikante intragruppe-forskjeller mellom faktorskårene som kan gi støtte til antagelsen av at AMI faktisk måler arbeidsminne og PHI måler prosesserings tempo. Vi nevnte i denne sammenhengen kritikken mot WISC-III der det viser seg at pasienter med ADHD hadde en jevn profil. Til tross for at de har et påvist oppmerksomhetsproblem, skåret de ikke svakere på oppmerksomhetsindeksen. Når det gjelder WAIS-III, finner vi derimot en slik sammenheng. Her skårer *alle* de kliniske gruppene med ett unntak systematisk svakere på AMI enn på GEI. I tråd med resonnetet under punkt 4 ovenfor fremstår således arbeidsminne som en kognitiv funksjon som lett forstyrres. Det er ikke urimelig å tenke seg at alle gruppene med svakere AMI har en genuin reduksjon i arbeidsminne. I en tilsvarende studie av

pasienter med nevropsykiatriske lidelser fant Iverson et al. (2006) også en systematisk gruppeforskjell mellom GEI og AMI. For å kunne dokumentere at reduksjonen skyldes arbeidsminnet og ikke at normene er systematisk noe for strenge i forhold til arbeidsminne relativt til GEI, hadde vi imidlertid ønsket å finne at effekten var mer differensiell i den forstand at den kun ble funnet i noen av gruppene. Eksempelvis er det ikke i tråd med forventningene å finne at en liten gruppe på seks personer som er diagnostisert å ha generelle lærevansker, fremstår med statistisk signifikant svakere AMI enn GEI. At de eneste som ikke skårer svakere på AMI enn på GEI, er pasientene med ikke-verbale lærevansker, indikerer at AMI ikke er et modalitetsuavhengig mål på arbeidsminne. Disse pasientene er jo forventet å ha en velfungerende auditiv kapasitet, men derimot å ha spesifikke visuelle vansker også innenfor arbeidsminne-domenet (Rourke, 1989). Sammenfattet har vi ved å kontrastere våre WAIS-III- funn med det Egeland et al. (2006) fant i forhold til WISC-III, påvist at mens oppmerksomhetsindeksen var lite sensitiv for oppmerksomhetsvansker i WISC-III, er den svært sensitiv, men derimot lite spesifikk i WAIS-III. Mens faren for å underdiagnostisere spesifikke oppmerksomhetsvansker er klart til stede i WISC-III, bør vi advare klinikerer mot mulige falskt positive i WAIS-III.

Nedsatt prosesseringshastighet målt ved PHI fremstår som et mer spesifikt utfall enn AMI-reduksjonen. Begge ADHD-gruppene, pasientene med diffus og høyresidig hjerneskade, schizofreni og somatiske lidelser med psykiske symptomer hadde alle statistisk signifikant svakere PHI enn GEI. Pasientene med generelle lærevansker, ikke-verbale lærevansker, venstresidig hjerneskade, rusproblematikk og depresjon hadde derimot ingen klar reduksjon. Når det gjelder depresjon, var det numerisk sett en forskjell, men den var ikke signifikant i den lille gruppen som deltok i undersøkelsen (n=6). Koding kan kanskje sies å være en test som ligner på lesing og skriving, mens symbolleting ligner mest på lesing. At pasientene med språklige lærevansker har redusert tempo, mens pasienter med ikke-verbale vansker ikke har det, indikerer at PHI heller ikke er modalitetsnøytral. Det hadde således kunnet styrke denne indeksskåren hvis den hadde inkorporert en mer romlig orientert prosesseringstest med tid som avhengig variabel. Klinikerer kan i så fall anbefales å bruke eksempelvis Trail Making Test A (Reitan, 1992) eller lignende delprøver fra D-KEFS (Delis, Kaplan & Kramer, 2001; Foss, 2005) som kompletterende mål.

## **Begrensninger**

Testdata i denne undersøkelsen er innhentet retrospektivt fra en rekke psykologer. Av dette følger det en risiko for at testene kan ha vært ulikt administrert og skåret. Det har ikke vært mulig å teste inter-dommer-reliabilitet. Klassifikasjonen i diagnostiske hovedgrupper gjenspeiler rådende diagnostisk praksis i psykisk helsevern, dvs. at det også her er en risiko for feildiagnostikk. Når det gjelder diagnostikken av lærevansker, er denne ikke fullt ut uavhengig av prøveresultatene, men diagnosene er imidlertid ikke satt ut fra WAIS-resultatene alene. Imidlertid innebærer dette en sammenblanding av avhengig og uavhengig variabel i undersøkelsen. Det må også understrekes at noen av diagnosegruppene er små. Dette øker på den ene siden risikoen for tilfeldige resultater



(Type 1-feil), men reduserer også muligheten for statistisk signifikante resultater relativt til de større diagnosegruppene (Type 2- feil).

## **Konklusjon**

På tross av de metodiske begrensningene som er nevnt ovenfor, vil vi likevel hevde at undersøkelsen gir oss ny og valid kunnskap som kan være til nytte for klinikerne som bruker evnetesten i utredning av kognitive vansker hos voksne. Følgende mener vi har direkte klinisk anvendbarhet:

- Korrigering for enkeltledd med særlig lav løsningsfrekvens (slik som spørsmålet om kontinenter i Informasjon).
- Unngå Likheter som en enkeltstående test av abstraksjonsevne.
- Tallspenn og Bokstav-Tall-Sekvensering gir trolig et renere mål på arbeidsminne enn arbeidsminneindeksen når den også er basert på delprøven Regning.
- Resonnering og Tegneserier er overflødige deltester hvis man benytter de andre deltestene som inngår i VFI og POI.
- Det nye evnemålet GEI er trolig et bedre alternativ til Total IQ hvis man ønsker et samlemål på intellektuell funksjon som ikke kontamineres av tilstandsavhengig svikt i prosesseringshastighet eller oppmerksomhetsfunksjoner.
- AMI er et sensitivt, men lite spesifikt mål i klientgrupper som er karakterisert ved redusert arbeidsminnefunksjon. PHI er mer spesifikt. Med forbehold om at utilstrekkelig reliabilitet av diagnostikken i vår studie kan ha påvirket resultatene, er det likevel grunnlag for å anbefale at begge indeksskåre kompletteres med andre visuelle eller romlige prøver på arbeidsminne og prosesseringshastighet.

Jens Egeland

Forskningsenheten

Psykiatrien i Vestfold HF

Pb 2267

3103 Tønsberg

E-post [jens.egeland@piv.no](mailto:jens.egeland@piv.no)

*Teksten sto på trykk første gang i Tidsskrift for Norsk psykologforening, Vol 46, nummer 7, 2009, side 658-666*

### **TEKST**

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

KONTAKT: [jens.egeland@siv.no](mailto:jens.egeland@siv.no)

**Ole Bosnes**

**Hans Johansen**

## Referanser

- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders*. Fourth edition. Text revision. Washington D.C. American Psychiatric Association.
- Arnau, R. & Thompson, B. (2000). Second-order confirmatory factor analysis of the WAIS-III. *Assessment*, 7, 237-246.
- Baddeley, A. D., Bressi, S., Della Sala, S., Logie, R. & Spinnler, H. (1991). The decline of working memory in Alzheimer's Disease. *Brain*, 114, 2521-2542.
- Bernstein, D.M. (2002). Information processing difficulty long after self-reported concussions. *Journal of the International Neuropsychological Society*, 8, 673-682.
- Bowden, S.C., Weiss, L.G., Holdnack, J.A. & Lloyd, D. (2006). Age-related invariance of abilities measured with the Wechsler Adult Intelligence Scale-III. *Psychological Assessment*, 18, 334-339.
- Bull, F., Eskeland, A. & Tandberg, E. (red.) (1972). *Gyldendals store konversasjonsleksikon*, 3. utgave. Oslo: Gyldendal.
- Burton, D.B., Ryan, J.J., Axelrod, B.N. & Schellenberger, T. (2002). A confirmatory factor analysis of the WAIS-III in a clinical sample with crossvalidation in the standardization sample. *Archives of Clinical Neuropsychology*, 17, 371-387.
- Delis, D. C., Kaplan, E. & Kramer, J. H. (2001). *Delis-Kaplan Executive function System (D-KEFS)*. The Psychological Corporation, San Antonio, Texas.
- Egeland, J., Rund, B.R., Sundet, K., Landrø, N.I., Asbjørnsen, A., Lund, A., Roness, A., Stordal, K.I. & Hugdahl, K. (2003). Attention profile in schizophrenia compared to depression: Differential effects of processing speed, selective attention and vigilance. *Acta Psychiatrica Scandinavica*, 108, 276-284.
- Egeland, J., Andreassen, T.H., Sundberg, H., & Stensli, O. (2006). Ny norsk WISC-III: Fare for overdiagnostisering av språklige vansker? *Tidsskrift for Norsk Psykologforening*. 43, 702-707.
- Egeland, J., Bosnes, O., & Johansen, H. (2008). Factor Structure of the Norwegian version of WAIS-III in a clinical sample: The Arithmetic problem. *Assessment*. Doi: 1177/1 073 191 108 324 464
- Ellertsen, B. & Johnsen, I. M. B. (2003). *WISC-III, Wechsler Intelligence Scale for Children- Third Edition, Norsk versjon, Administrasjonsveiledning*. Oslo: Assessio Norge AS.
- Foss, T. (2005). *Delis-Kaplan Executive Function System. Norsk Manuals supplement*. Stockholm: Harcourt Assessment.
- Garcia, L.F., Ruiz, M.A. & Abad, F.J. (2003). Factor structure of the Spanish WAIS-III (spanish). *Psicothema*, 15, 155-160.
- Goldman-Rakic, P.S. (1994). Working memory dysfunction in schizophrenia. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 6, 348-357.
- Gregoire, J. (2004). Factor structure of the French version of the Wechsler Adult Intelligence Scale-III. *Educational and Psychological Measurement*, 64, 463-474.
- Harcourtassessment.com: WAIS-III-WMS-III Factor Analysis.
- Iverson, G.L., Lange, R.T., Viljoen, H., & Brink, J. (2006). WAIS-III General Ability Index in neuropsychiatry and forensic psychiatry inpatient samples. *Archives of Clinical Neuropsychology*, 21, 77-82.
- Jöreskog, K. & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Lincolnwood, Illinois: Scientific Software International.
- Keith, T.Z. (2005). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D.P. Flanagan & P.L. Harrison (eds.) *Contemporary intellectual assessment*. New York: Guildford Press.

- Kemps, E., Tiggemann, M. & Marshall, K. (2005). Relationship between dieting to loose weight and the functioning of the central executive. *Appetite*, 45, 287-294.
- Landrø, N.I., Stiles, T.C. & Sletvold, H. (2001). Neuropsychological function in nonpsychotic unipolar major depression. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology*, 14, 233-240.
- Lezak, M.D. (1995). *Neuropsychological Assessment*. Third Edition. New York: Oxford University Press.
- Matheson, M.P., Stansfeld, S.A. & Haines, M.M. (2003). The effects of chronic air-craft noise exposure on children's cognition and health: 3 field studies. *Noise & Health*, 5, 31-40.
- Parmenter, B.A., Shucard-Janet, J.L., Benedict, R.H. & Shucard, D.V. (2006). Working memory deficits in Multiple Sclerosis: comparison between the n-back task and Paced Auditory Serial Addition Test. *Journal of the International Neuropsychological Association*, 12, 677-687.
- Pelletier, P.M., Ahmad, S.A. & Rourke, B.P. (2001). Classification rules for basic phonological processing disabilities and nonverbal learning disabilities: Formulation and external validity. *Child Neuropsychology*, 7, 84-98.
- Reinecke, M., Beebe, D. & Stein, M. (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, 322-328.
- Reitan, R. M. (1992). *Trail Making Test: Manual for administration and scoring*. Tucson, Arizona: Reitan Neuropsychological Laboratory.
- Rourke, B. (1989). *Non-verbal learning disabilities: the syndrome and the model*. New York: Guilford Press.
- Saklofske, D.H., Hildebrand, D.K. & Gorsuch, R.L. (2002). Replication of the factor structure of the Wechsler Adult Intelligence Scale -Third Edition with a Canadian Sample. *Psychological Assessment*, 12, 436-439.
- Spikman, J.M., van Zomeren, A.H. & Deelman, B.G. (1996). Deficits of attention after closed-head injury: Slowness only? *Journal of Clinical and Experimental Neuropsychology*, 18, 755-767.
- Sundberg, H., Egeland, J., Andreassen, T.H. & Stensli, O. (2006). En sammenligning av WISC-r og WISC-III protokoller for et barnpsykiatrisk utvalg. *Tidsskrift for Norsk Psykologforening*, 43, 476-481.
- The Psychological Corporation (1999). *The Wechsler Abbreviated Scale of Intelligence Manual*. San Antonio, TX: Harcourt Brace & Company.
- Tulsky, D.S. & Price, L.R. (2003). The joint WAIS-III and WMS-III factor structure: Development and cross-validation of a six-factor model of cognitive functioning. *Psychological Assessment*, 15, 149-162.
- Tulsky, D.S., Saklofske, D.H., Wilkins, C., & Weiss, L.G. (2001). Development of a General Ability Index for the Wechsler Adult Intelligence Scale- third edition. *Psychological Assessment*, 13, 566-571.
- Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
- Wechsler, D. (1997a). *The Wechsler Adult Intelligence Scale (3rd ed.)* San Antonio, TX: Psychological Corporation.
- Wechsler, D., Nyman, H. & Nordvik, H. (2003). *WAIS-III: Wechsler Adult Intelligence Scale: Manual. Norsk Versjon*. Stockholm: Psykologiförlaget.
- Ørbeck, B. & Sundet, K.S. (2007) *WASI Norsk versjon, Manual supplement*. Stockholm: Harcourt Assessment.