

**Validering av normer for oversatte tester av
oppmerksomhet og hukommelse i et norsk normalutvalg**

Jens Egeland , Kjetil Sundet , Nils Inge Landrø , Bjørn Rishovd Rund , Arve
Asbjørnsen , Kenneth Hugdahl , Anders Lund , Atle Roness og Kirsten Stordal

Validering av normer for oversatte tester av oppmerksomhet og hukommelse i et norsk normalutvalg

Mange tester som brukes i klinisk praksis er ikke normert for norske forhold. Dette gir risiko for feilvurderinger. I artikkelen sammenliknes normer for en rekke tester med de faktiske prestasjonene i et norsk utvalg.

Innledning

Tester av hukommelse og oppmerksomhet benyttes stadig mer i planlegging av rehabilitering ved skade eller sykdom i sentralnervesystemet, i differensialdiagnostiske sammenhenger og i forsikringsoppgjør etter ulykker. Det utvikles kontinuerlig nye tester på det internasjonale markedet. Ny kunnskap om forholdet mellom hjerne og atferd har ført til utvikling av nye test-paradigmer. Mens man tidligere vanligvis bare benyttet prestasjonsorienterte tester som Wechsler-skalaene, synes det å være en dreining mot mer prosessorienterte tester (Delis, Kramer, Freeland & Kaplan, 1998). Man nøyer seg ikke med kun å kartlegge om klienten husker godt eller dårlig, men er interessert i hvilket element i prosessen som eventuelt svikter. Ofte vil det være aktuelt å bruke flere tester for å få et bilde av ulike kognitive del-prosesser. Klinikeren som ønsker å være metodisk oppdatert kan ikke avvende fremtidige norske normeringer, men må inntil videre ta i bruk tester som ikke er normert for norske forhold.

At testene ikke er normert på norske utvalg, øker faren for feilklassifikasjoner. I denne artikkelen sammenlignes det originale normgrunnlaget på en rekke tester som er i klinisk bruk i Norge med de faktiske prestasjonene for 50 friske personer mellom 20 og 49 år. Antall personer er for lite til å kunne brukes som en normeringsgruppe, men kan likevel gi et grovt bilde av i hvilken grad normer utviklet i engelskspråklige land er gyldige i Norge.

Heldigvis har de få publikasjonene som har sett på gyldigheten av norske testoversettelser, vist et rimelig godt samsvar med de engelskspråklige originalversjonene. I den norske normering av fonetiske (FAS) og semantiske (COWAT) ordflyt-oppgaver fant Tjemsland, Walbækken og Egeland (2001) et godt samsvar mellom norske og engelskspråklige normer. Tilsvarende fant Brager-Larsen, Sundet, Engvik, Ørbeck og Nes (2001) at interkorrelasjoner mellom delprøver samt IQ-verdier for en mindre gruppe kontroller og pasienter med kognitive vansker, indikerte godt samsvar for den norske oversettelsen av Wechsler Abbreviated Scale of Intelligence (WASI). Også Bosnes og Ellertsen

(2003) fant at den norske oversettelsen av California Verbal Learning Test (CVLT) tok vare på egenskapene ved den opprinnelige amerikanske testen. De mente at normene var for strenge, men dette var en kritikk som også har blitt rettet mot den amerikanske versjonen.

Metode

Deltagerne i undersøkelsen utgjøres av kontrollgruppen i «Bergen–Oslo-prosjektet» (Egeland, Rund et al., 2003; Egeland, Sundet et al., 2003; Egeland et al., 2004). Tjuefem kvinner og tjuefem menn mellom 20 og 49 år deltok. Personene var rekruttert ved spredning av brosjyrer og etter avisoppslag hovedsakelig fra Tønsberg-regionen, men også noen fra Bergen. Yrkesmessig tilhørte de ulike bransjer (industri, offentlig forvaltning, helsevesen, varehandel). Gjennomsnittsalderen for hele utvalget var 33 år (S.A 9 år). Utdanningsnivå varierte fra 9 år til 20, med middelerdi på 14 år (S.A. 2 1/2 år). For å få et grovt estimat på intellektuell evne, ble deltagerne undersøkt på en visuell og en språklig delprøve fra WAIS-r (Wechsler, 1981). Delprøvene billedutfylling og likheter ble valgt, da disse prøvene har en høy g-faktor (Lezak, 1995). Gjennomsnittlig IQ basert på dette estimatet var 105 (S.A. 7,9). Det var ingen forskjeller mellom kvinner og menn i utdanning eller IQ. Kvinnene var i snitt 2 1/2 år eldre enn mennene, men denne forskjellen var ikke statistisk signifikant.

Førtiseks personer var høyrehendte, fire var venstrehendte. Alle deltagere var norsktalende, med normalt syn og hørsel. De undertegnet et informert samtykkeskjema hvor de ble informert om følgende eksklusjonskriterier: a) utviklingsforstyrrelser, tidligere hodeskade eller tidligere/nåværende nevrologisk sykdom, b) nåværende alkohol- eller stoffmisbruk, c) medisinske sykdommer som kunne tenkes å påvirke funksjon i sentralnervesystemet. Deltagerne måtte også bekrefte at de ikke hadde, og ikke tidligere hadde hatt, noen diagnostisert psykisk lidelse.

Nevropsykologiske mål

California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987). CVLT er trolig den hyppigst brukte språklige nylæringsprøven i Norge (se bl.a. Lundervold, Reinvang & Lundervold, 1994). I 2000 ble det publisert en renormering med demografisk korrigerte normer (Norman, Evans, Miller, & Heaton, 2000). Samme år kom en revidert utgave med parallelle varianter av testen (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000). CVLT-II har mindre strenge normer enn den opprinnelige CVLT-versjonen, som kun var normert på 273 personer mellom 17 og 80 år. Som del av normeringen av CVLT-II ble 62 personer testet med både den nye og den gamle versjonen av prøven.

Korrelasjonene mellom råskårene for prestasjonsmålene (målene på hvor godt en person lærer) og på målene av læringsatferd, varierte fra .72 til .80, hvilket tilsier at de to testene egentlig kan betraktes som parallelle former (Delis et al., 2000). I denne artikkelen vil vi derfor sammenligne våre resultater ved bruk av CVLT mot normene for både den opprinnelige og den reviderte versjonen.

Den opprinnelige CVLT-prøven har en rekke ulike prestasjonsmål, hvorav vi har analysert CVLT total innlæring (sum av fem innlæringsforsøk), fri utsatt gjenkalling, utsatt gjenkalling med stikkord og diskriminering (prosent riktige på gjenkjenningsprøven). Prøven gir 12 ulike mål på læringsatferd. Vi har analysert grad av semantisk organisering; seriell organisering; andel av totale svar som kommer fra første fire-del av listen («primacy») og siste del («recency»); konsistens (samsvar mellom svarene på de fem innlæringsforsøkene); læringskurve; intrusjoner (feilsvar); perseverasjoner (gjentakelser); feiltype («bias»: Om man har en tendens til å svare ‘ja’, eventuelt ‘nei’) og antall falske positive responser under gjenkjenning. Formelen for noen av parameterne er endret noe i CVLT-II sammenlignet med den opprinnelige versjonen. Dette gjelder eksempelvis diskrimineringsmålet og beregning av semantisk og seriell læringsstrategi.

Tabell 1. CVLT prestasjon hos kvinner (N = 25) sammenlignet med CVLT- og CVLT II-normer.

Test	M	Norske kvinner (S.D)	95 % k.i.	CVLT-normer M (S.D.) 35–44 år	Vektet	Avvik fra normene i S.D.	Normer CVLT II Vektet M (S.D.) 30–44	Avvik fra normene i S.D.
Total innlæring	58.4	(9.5)	54.4–62.3	62	7	-0.6	54 (9.0)	+0.4*
Utsatt fri gjenkalling	13.1	(2.4)	12.1–14.1	14.1	ca 1.5	-0.7*	12.6 (2.3)	+0.2
Utsatt stikkord	13.2	(2.2)	12.3–14.1	14.8	ca 1.5	-1.1*	13.1 (2.3)	0
Diskriminering	97.5	(2.9)	96.3–98.6	95.6	Ca 4	+0.5*		
Semantisk organisering	2.6	(0.8)	2.3–2.9	2.3	Ca 0.4	+0.7		
Seriell organisering	1.5	(1.3)	1.0–2.0	1.5	Ca 0.8	0		
% svar fra første fire del	25.0	(2.8)	23.8–26.1	26	Ca 2.5	0	28.5 (5.3)	-0.7*

Test	M	Norske kvinner (S.D)	95 % k.i.	CVLT-normer M (S.D.) 35–44 år	Vektet	Avvik fra normene i S.D.	Normer CVLT II Vektet M (S.D.) 30–44	Avvik fra normene i S.D.
% svar fra siste fire del	28.0	(4.8)	26.4–30.3	28	Ca 2.5	0	26.6 (4.3)	+0.3
Konsistens	88.7	(7.4)	85.7–91.8	85.7	9	+0.2	81.7 (8.6)	+0.8*
Læringskurve	1.3	(0.5)	1.1–1.5	1.3	Ca 0.3	0		
Intrusjoner	1.1	(2.0)	0.3–1.9	1.2	Ca 1.5	-0,1	1.0 (2.5)	+0.1
Perseverasjoner	1.1	(3.5)	2.7–5.7	5.3	Ca 2	-0.5	4.2 (4.0)	0
Feiltype	-0.0	(0.2)	-0.2–0	0.1	Ca 0.2	-0.2		
Falske positive	0.3	(0.5)	0.05–0.50	0.6	Ca 1			

Rey Complex Figure Test and Recognition Trial (RCFT; Meyers & Meyers, 1995;1996). Denne geometriske tegnekopieringsoppgaven med etterfølgende testing av gjenkalling ble laget av Rey (1941) og normert av Osterrieth, (1944). Den ble nylig omfattende renormert av Meyers og Meyers (1996), som også la til en gjenkjenningsoppgave der forsøkspersonene skal gjenkjenne riktige detaljer blant 24 mulige ekte/falske positive. Den nye versjonen har normer fra 6 til 89 år. I undersøkelsen her refereres resultatet av kopieringsprøven, av langtids-gjenkallingsprøven og av gjenkjenningstesten.

Paced Auditory Serial Addition Test (PASAT; Gronwall & Wrightson, 1975; Landrø, Gulowsen & Sletvold, 2004). I denne oppgaven lytter forsøkspersonene til et lydbånd der det leses opp tall mellom en og ni. Oppgaven består i å legge sammen de til enhver tid to sist oppleste tall, og si summen høyt. Etter å ha lagt sammen tall nummer en og to, skal to og tre legges sammen, deretter tre og fire. Det gjelder å ikke blande sammen summeringsoppgaven med det å huske tallene. Således blir dette en test av å gjøre to ting samtidig (Channon, Baker & Robinson 1993), det vil si simultankapasitet eller «the central executive»-elementet i Baddeleys modell for arbeidshukommelse (Baddeley, 1999). I den norske versjonen (Landrø et al., 2004) som her er brukt, leses det først opp 60 tall med en hastighet av ett tall per tredje sekund. Deretter presenteres ytterligere 60 tall med en hastighet av ett tall hvert annet sekund. Siden denne versjonen er identisk med den som har vært normert som del av det nederlandske «Brief Repeatable Battery of Neuropsychological Tests» (Boringa et al., 2000), har vi valgt å sammenligne det norske utvalget med disse normene.

Stroop Color–Word Test. To varianter av Stroop-testen er i klinisk bruk i Norge. Goldens (1978) versjon er designmessig litt annerledes enn den prøven som vi her har brukt, nemlig Kenneth Hugdahls Stroop-test, som er en forkortet utgave av Commali og Kaplans versjon (Lund-Johansen, Hugdahl & Wester, 1996; Mitrushina et al., 1999). For klinikerer som tidvis sammenligner egne testresultater med andres, kan det være viktig å være oppmerksom på at Goldens test måler antall riktige responser innen en tidsperiode, mens Hugdahls test måler antall sekunder på et visst antall responser. Seks fargenavn (hvit, svart, rød, blå, grønn og gul) eller fargeklatter med tilsvarende farge, ble gjengitt til sammen 48 ganger. Først skal deltageren benevne fargeklattene, i neste oppgave lese ordene og i den siste delen si fargen ordet er trykket i. I denne siste oppgaven er det uoverensstemmelse mellom blekkfarge og tekst, slik at deltageren må undertrykke den automatiske impulsen om å lese. Målene vi oppgir her er antall sekunder klienten bruker på å gjennomføre hver prøve.

CalCAP. Fra California Computerized Assessment Package (Miller, 1993) valgte vi ut fire delprøver som målte enkel reaksjonstid og gradvis mer krevende selektiv reaksjonstid. Disse reaksjonstidsprøvene er i bruk hos en rekke klinikere. Testen har kun normer for menn. I den enkle reaksjonstidstesten bes deltageren om å trykke på tastaturet hver gang et tall dukket opp på data-skjermen. I neste oppgave (selektiv reaksjonstid) skal deltageren kun respondere på syv-tallet. Tilslutt følger to sekvensielle oppgaver der deltageren henholdsvis skal respondere overfor to like tall etter hverandre, og når neste tall er i stigende rekkefølge. Oppgavene representerer et kontinuum fra automatisk til mer komplisert informasjonsprosessering.

Tabell 2. CVLT-prestasjon hos menn (N = 25) sammenlignet med CVLT- og CVLT II-normer.

Test	M	Norske menn (S.D)	95 % k.i.	CVLT-normer M (S.D.) 35–44 år	Vektet	Avvik fra normene i S.D.	Normer CVLT II Vektet M (S.D.) 30–44	Avvik fra normene i S.D.	
Total innlæring	56.1	(9.6)	52.1–60.1	57	7.0	-0,1	48.8	10,5	+0,7*
Utsatt fri gjenkalling	13.0	(2.2)	12.1–13.9	12.8	2.0	+0,1	11.1	2,5	+0,8*
Utsatt stikkord	13.1	(2.4)	12.1–14.1	13.3	1.5	-0,1	12.1	3	+0,3
Diskriminering	96.8	(3.7)	95.3–98.3	96.8	4.0	0			
Semantisk organisering	2.0	(1.0)	1.6–2.4	2.3	.4	-0,7			

Test	M	Norske menn (S.D.)	95 % k.i.	CVLT-normer Vektet M (S.D.) 35–44 år	Avvik fra normene i S.D.	Normer CVLT II Vektet M (S.D.) 30–44	Avvik fra normene i S.D.	
Seriell organisering	2.5	(1.9)	1.6–3.5	1.5	0.8		+1,2	
% svar fra første firedel	2.8	(5.3)	25.8–30.2	29.0	4.0	30.0	5	-0,4
% svar fra siste firedel	28.0	(3.5)	26.3–29.1	29.0	2.6	27.7	5,5	+0,1
Konsistens	83.8	(7.3)	80.8–86.8	87.0	9.0	81.3	8	+0,3
Læringskurve	1.3	(0.4)	1.1–1.5	1.3	0.3			0
Intrusjoner	1.1	(1.6)	0.4–1.8	2.1	2.0	1	2,5	+0,1
Perseverasjoner	1.1	(2.4)	2.6–4.6	5.4	2.0	4.3	3	-0,2
Feiltype	0.0	(0.2)	-0.1–0.1	0.1	0.2			-0,4
Falske positive	0.7	(1.2)	0.2–1.2	1.0	2.0			

Recognition Memory Test (RMT; Warrington, 1984). Testen består av en visuell (ansikter) og en språklig (ord) delprøve. Klienten presenteres 50 kort med et ord eller et bilde av et ansikt. Kortene presenteres ett for ett, i ca. fem sekunder. For å sikre seg at klienten er oppmerksom og bearbeider stimuli på et «dypt» semantisk nivå, bes klienten vurdere hvert kort med hensyn til om ordet eller personen er sympatisk/behagelig. Umiddelbart etter presentasjonen av kortene, får klienten presentert enten 50 ordpar eller ansiktspar, og bes om å identifisere ordet eller ansiktet som hun/han tidligere har sett i hvert par. I denne artikkelen presenteres resultatene fra hver delprøve separat.

Tabell 3. RCFT, PASAT, Stroop og RMT: Sammenligning av norske forsøkspersoner (N = 50) og normer.

Test	Norske forsøkspersoner M	Normer (S.D.)	95 % k.i.	Vektet M	(S.D.)	Avvik fra normene
RCFT						
Kopiering	35.7	(1.2)	35.4–36.1			
Langtidsgjenkalling	4.5	(5.8)	22.8–26.1	23.1	4.81	+0.3
Gjenkjenning	21.1	(1.7)	20.6–21.5	21.3	Ca 1.51	-0.1
PASAT						

Test	Norske forsøkspersoner			Vektet M	(S.D.)	Avvik fra normene
	M	Normer (S.D.)	95 % k.i.			
3 sekunder	52.0	(8.7)	49.5–54.4	48.7	(10.7)	+0,3
2 sekunder	43.4	(9.5)	40.7–46.1	38.2	(11.0)	+0.5
Stroop Farge	25.5	(5.1)	24.1–27.0	25.4	4.2	0
Stroop Ordlesing	16.5	(3.1)	15.6–17.4	16.4	3.0	0
Stroop Farge–Ord	43.7	(9.6)	41.0–46.4	45.5	11.3	+2
RMT						
Ord	49.0	(1.2)	48.6–49.3	46.6	33	+0,8*
Ansikter	42.4	(4.3)	41.1–43.6	43.4	3.53	-0,3

Dataanalyser

Testresultatene for hver enkeltperson ble skåret ut fra normene angitt i testmanualene. Det ble beregnet gjennomsnitt, standardavvik og 95. prosentil konfidensintervall for forsøkspersonene. Normverdier utenfor dette konfidensintervallet betraktes som avvikende fra den norske gruppen. Testresultatene ble korrelert med bakgrunnsvariablene kjønn, alder og utdanning.

For å lette sammenligningen med gjeldende normer, ble normene vektet for å ta hensyn til forsøkspersonenes spredning over flere aldersgrupper med ulike normer. Tilsvarende vektning er ikke gjort med hensyn til standardavvik, siden det ikke er noen systematisk økning i standardavvik under 50 år. I tabellene har vi derfor for hver test valgt ut standardavviket for den høyest frekvente aldersgruppen. Der det ikke er oppgitt desimaler i håndbøkene, er standardavviket oppgitt som en cirkaverdi, beregnet ved å halvere differansen mellom normverdiene for T 40 og T 60-nivået. De norske forsøkspersonenes avvik fra normene er beregnet i standardskårer, (z-skårer) basert på de estimerte standardavvikene. Vi har også beregnet andelen av de norske forsøkspersonene som har en differanse som er større enn 10 poeng mellom Ansikts- og Ord-delprøvene på RMT. Klinisk brukes denne differansen som indikasjon på lateraliserde utfall i venstre eller høyre hjernehemisfære. Ti poeng svakere skåre på ansiktsgjenkjenningssprøven regnes som en signifikant forskjell (5. prosentilnivå).

Resultater

CVLT-resultatene er splittet opp på kvinner og menn, slik det også er i testmanualen. Resultatene for kvinner fremkommer av Tabell 1. Tre av skårene for kvinnene var avvikende fra de opprinnelige CVLT-normene ved at de amerikanske normene var utenfor 95 prosentil-intervallet for de norske kvinnene. Sammenlignet med normene presterte de norske kvinnene svakere med hensyn til fri utsatt gjenkalling og gjenkalling ved stikkord. Kvinnene i dette utvalget presterte derimot bedre med hensyn til gjenkjenning.

I forhold til CVLT-II avvek de norske kvinnenes resultater fra normene på et prestasjonsmål og to strategimål: Total innlæring var bedre enn normene. Kvinnene hadde en lavere prosent svar fra første firedel av listen, og bedre konsistens enn normene skulle tilsi (normene er utenfor 95. prosentil-intervallet til de norske kvinnene).

CVLT-resultatene for mennene presenteres i Tabell 2. Ingen av disse resultatene avvek fra de opprinnelige CVLT-normene. Sammenlignet med CVLT-II-normene, presterte derimot de norske mennene bedre på to av prestasjonsmålene, nemlig total innlæring og utsatt fri gjenkalling. Ingen av strategimålene avvek fra CVLT-II-normene.

Det var ingen statistisk signifikant forskjell mellom kvinner og menn på noen av resultatmålene. Med hensyn til målene på læringsatferd hadde kvinnene en statistisk signifikant høyere grad av semantisk organisering av lærestoffet på CVLT, mens det var en tendens i retning av høyere seriell organisering hos menn ($p = .056$). Kvinner hadde en statistisk signifikant høyere grad av konsistens over de fem innlæringsforsøkene i CVLT. Menn hadde en høyere prosentats svar fra den siste firedelen av listen.

Ingen av CVLT-målene korrelerte signifikant med alder i dette utvalget. Total innlæring og utsatt fri gjenkalling korrelerte med utdanning (henholdsvis $r = .30, p = .03$ og $r = .31, p = .03$).

Tabell 3 viser resultatene for Rey Complex Figure Test, Pasat, Stroop test og Recognition Memory Test. Rey Utsatt Minne korrelerte med utdanning ($r = -.20, p = .01$), men ikke med alder og kjønn. Også PASAT-resultatene korrelerte kun med utdanning (3 sek.: $r = .35, p < .01$, 2 sek.: $r = .23, p < .001$). Stroop Fargebenevning korrelerte med kjønn ($r = -.25, p < .01$). Ordlesing korrelerte med utdanning ($r = -.28, p = .02$). Farge-ord korrelerte med utdanning og alder (utdanning: ($r = -.18, p = .03$, alder: ($r = .17, p = .04$).

Tabell 4. CalCAP: Sammenligning av norske menn (n = 25) og norske kvinner (n = 25) med amerikanske normer.

Test	Norske kvinner			Norske menn			Normer menn M (S.D.) 35–44 år	Vektet	Avvik fra normene i S.D.
	M	(S.D.)	95 % k.i.	M	(S.D.)	95 % k.i.			
Enkel reaksjonstid	335.8	(81.3)	301.4– 370.1	287.1	(40.2)	270.1– 304.1	367.4	97.22	-,83*
Valg reaksjonstid	431.6	(59.5)	406.5– 456.7	396.3	(44.8)	377.4– 415.2	408.1	43.72	-,27
Sekvensiell reaksjonstid I	543.5	(88.2)	506.3– 580.7	471.8	(97.7)	430.5– 513.0	544.5	93.12	-,78*
Sekvensiell reaksjonstid II	614.5	(90.5)	576.3– 652.8	531.3	(116.0)	482.4– 580.3	604.3	114.12	-,64*

Ord-delprøven fra Recognition Memory Test var bedre i det norske utvalget enn i normene. På RMT-ansikter presterte forsøkspersonene numerisk sett svakere enn normene, men denne forskjellen var ikke statistisk signifikant. At delprøvene avvek hver sin vei, påvirker frekvensen av personer med avvikende differanse mellom delprøvene. Utfra normenes kriterier for avvik, hadde 20 % av forsøkspersonene svakere skåre på ansiktsprøven sammenlignet med RMT-ord. Begge delprøvene i RMT korrelerte signifikant med utdanning, men ikke med alder eller kjønn (Ord: $r = .19$, $p = .02$, ansikter: $r = .17$, $p = .03$).

Tabell 4 viser resultatene for CalCAP. De norske mennene presterte bedre enn normene på tre av fire delprøver. Mennene var statistisk signifikant raskere enn kvinnene på alle delprøvene. Kvinnenes prestasjoner var på nivå med normene for amerikanske menn. Ingen av delprøvene korrelerte med alder eller utdanning.

Diskusjon

Kjønnforskjeller i CVLT: De norske kvinnene avviker fra de opprinnelige CVLT-normene på tre av 14 mål og fra CVLT-II-normene på tre av åtte analyserte parametre. Mennene avvek ikke fra de opprinnelige normene på noe mål, men avvek fra CVLT-II-målene på to av åtte parametre.

Hovedsakelig samsvarer altså de norske kontrollene med de amerikanske normene. Avvikene indikerer imidlertid at den opprinnelige CVLT-versjonens normer for gjenkalling var for strenge for kvinner, og reiser spørsmål om den nye versjonen er for snill når det gjelder innlæring både for kvinner og menn.

I den opprinnelige versjonen (Delis et al., 1987) var forskjellen mellom menn og kvinner ca. ett standardavvik. I normene for eldre (Paolo, Tröster & Ryan, 1997) varierer kjønnsforskjellen fra ca. et halvt standardavvik til henimot et helt. I de demografisk korrigerede normene (Norman et al., 2000) og i CVLT-II er forskjellen redusert til i underkant av et halvt standardavvik. Kvinnene i vår undersøkelse skårer numerisk sett bedre enn mennene på alle prestasjonsmål, men ingen av kjønnsforskjellene ble statistisk signifikante. Noe av dette kan tilskrives lav statistisk styrke. Det er likevel klart at forskjellen i absolutt prestasjon er mindre i vårt norske utvalg enn man ville forventet ut fra de amerikanske normene. At kvinnene er noe eldre enn mennene (2 1/2 år) i vårt utvalg, kan spille en mindre rolle, men selv ved kovariasjon av alder, ble ikke kjønnsforskjellene signifikante. Aldersendringer er forøvrig et meget robust funn når det gjelder hukommelsestester. Ut fra de demografisk korrigerede normene, svekkes fri utsatt minne med et halvt standardavvik mellom 20 og 49 års alder (Norman et al., 2000). Liten statistisk styrke kombinert med manglende spredning på alder er trolig grunnen til at aldersforskjeller ikke ble statistisk signifikante i dette utvalget.

Det er uklart hva som er grunnlaget for eventuelle kjønnsforskjeller i læring, og de fremkommer heller ikke i alle undersøkelser av språklig læring. I Children's Auditory Verbal Learning Test-2 (Talley, 1992), som er en CVLT-lignende språklig læringstest for barn mellom 6 og 18 år, forklarte kjønn mindre enn en prosent av variansen i testresultater. I Våle-testen (Andreassen & Øksenholt, 2002), som er en ny norsk læringsprøve normert på 300 skolebarn fra 2. til 8. klasse, er det en liten kjønnsforskjell i favør av jentene i 2. klasse, men ikke på senere alderstrinn.

Når det gjelder målene på hukommelsesatferd, fremkommer flere statistisk signifikante kjønnsforskjeller. For semantisk og seriell organisering er kjønnsforskjellen enda større i det norske utvalget enn i normene. Kramer, Delis og Daniel (1988) fant også at kvinner i større grad enn menn organiserte innlæringen semantisk. Samtidig som kvinnene og mennene avviker fra normene hver sin vei, er standardavviket dobbelt så stort for de norske forsøkspersonene som for normene. På individnivå skal en derfor være forsiktig med kjønns-stereotyper av typen «kvinner bearbeider mening» mens «menn pugger». Det store standardavviket kan tyde på at mange nordmenn har vansker med inndelingen i fire meningsfulle kategorier. Klinikere kan ofte oppleve at klienter ikke spontant

opererer med «krydder og urter» som en kategori, og at mange tenker på «paprika» som en grønnsak som ikke naturlig hører sammen med krydder.

Er CVLT-normene for strenge?

At vi finner at de amerikanske normene er gyldige for mennene i utvalget, er overraskende tatt i betraktning at flere forskere har kritisert de opprinnelige normene for å være for strenge også for menn (Elwood, 1995). Wiens, Tindall og Crossen (1994) undersøkte et større utvalg av yngre, jobbsøkende personer (hovedsakelig menn). De fant at selv om personene hadde gjennomsnittlig IQ og 14 års utdanning, presterte de noe svakere enn normene skulle tilsi. Paolo et al.'s (1997) normering av CVLT for eldre personer ga gjennomgående noe snillere normer enn de opprinnelige. I en sammenligning av CVLT og Wechsler Memory Scale-revised, fant Randolph et al. (1994) at en gruppe eldre personer presterte mer enn et standardavvik over gjennomsnittet på WMS-r og gjennomsnittlig på CVLT. Bosnes og Ellertsen (2003) fant den samme forskjellen mellom CVLT og WMS-r i et norsk klinisk utvalg. De konkluderte at for strenge normer øker risikoen for falske positive. På den andre siden hadde de 20 kontrollpersonene i Lundervold et al.s (1994) undersøkelse 76 % riktige svar på innlæringsdelen av CVLT. Tilsvarende tall for vårt utvalg er 72 %. Lundervolds utvalg var gjennomsnittlig 39 år gamle og hadde 12 års utdanning. De 30 unge (gjennomsnittlig 15,7 år) kontrollene i Øie, Sundet og Runds (1999) studie av pasienter med schizofreni og ADHD, presterte også antydningvis bedre enn vårt utvalg med hensyn til innlæring. De presterte nøyaktig likt vårt utvalg med hensyn til fri gjenkalling. I Norman et al.'s (2000) kritikk av CVLT-normene, trekker de særlig frem at etniske minoriteter skal ha vært underrepresentert i det opprinnelige normeringsutvalget. Å undervurdere betydningen av etnisitet kan føre til at personer feilklassifiseres som å ha et hukommelsesproblem. I kontrollgrupper til forskningsprosjekter er man vanligvis svært nøye på å kontrollere for fremmedspråklighet. Med dette forbeholdet, samt at vårt utvalg er noe høyere utdannet (14 år) enn snittet i befolkningen (men gjennomsnittlig i IQ), gir ikke dette samlede norske normalutvalget støtte for antagelsen om at de opprinnelige CVLT-normene generelt har vært for strenge. Derimot føyer vår studie seg inn i rekken studier som har vist at kvinnenormene har vært for strenge.

Hvis den opprinnelige CVLT-ver-sjonen ikke hadde for strenge normer, betyr det at den ny versjonen kan være for snill. Ut fra Norman et al.s (2000) formel for demografisk korrigerede skårer, svarer ett års utdanning til nær en t-skåre. At vårt norske utvalg er noe høyere utdannet enn det amerikanske, kan likevel ikke fullt ut forklare gapet mellom våre resultater og CVLT-II-normene.

Det synes derfor påkrevd å få ytterligere normaldata, fortrinnsvis med CVLT-II-ordlistene, for å ytterligere avklare mistanken om at CVLT-II-normene kan bidra til falske negative.

I tillegg til forskjellen mellom gjennomsnittsskårene i den opprinnelige CVLT og CVLT-II, vil leseren kunne se av Tabell 1 og Tabell 2, at standardavvikene er større i CVLT-II. Det betyr at forskjellen mellom normsettene blir større jo større avviket er fra middelverdien. Et klinisk eksempel kan illustrere hvor ulike konsekvenser det har å legge det ene eller andre normsett til grunn: En femti år gammel mann som tidligere drev sitt eget firma, er nå sykemeldt grunnet kognitive vansker etter en trafikkulykke der han pådro seg en nakkeskade og muligens en lett hode-skade. Utsatt minne på CVLT er syv ord. Ut fra de opprinnelige CVLT-normene er dette to standardavvik under middelverdien for alder og kjønn, det vil si at det utgjør et sikkert utfall. Ut fra CVLT-II-normene er prestasjonene imidlertid kun et halvt standardavvik under gjennomsnittet. Undersøkelsen vil dermed ikke gi støtte til at den målte hukommelsesfunksjon er vesentlig dårligere enn sammenligningsgruppens.

RCFT, Stroop og RMT

Det er et godt samsvar mellom RCFT-normene og prestasjonene til de norske forsøkspersonene. Når det gjelder Hugdahls Stroop-test, presterer det norske utvalget nøyaktig som normene ville forutsi. Dette gir økt troverdighet til denne prøven, en test som er viktig i undersøkelse av konsentrasjonsevne og psykomotorisk tempo.

Vi har sammenlignet våre PASAT-resultater med den nederlandske versjonen, bygd på undersøkelse av 134 voksne personer, med noe høyere alder enn vårt utvalg (gjennomsnittlig 45,8 år, med et spenn fra 22–73 år). Aldersforskjellen til vårt utvalg på 12 år kan trolig forklare at det nederlandske utvalget presterer numerisk sett (men ikke statistisk signifikant) svakere enn vår gruppe.

Når det gjelder RMT, gir studien grunnlag for å advare om at normene kan hende ikke er gyldige. Det norske utvalget skårer statistisk signifikant forskjellig fra normene på RMT-ord. Det er en tak-effekt som gir testen dårlige psykometriske egenskaper. Når standardavviket blir så lite, blir resultatene lite robuste på enkeltindividnivå. For klinikerer er det vanskelig å argumentere for at klienter som kun gjør fire feil av femti mulige, skulle ha et hukommelsesproblem, selv om de ut fra de normalgruppens resultater avviker med mer enn to standardavvik. At RMT-ord og -ansikt-er avviker hver sin vei i det norske utvalget sammenlignet med de opprinnelige normene, svekker troverdigheten av testen når det gjelder å diagnostisere lateraliserte eller modalitetsspesifikke

hukommelsesvansker. At 20 % av utvalg-et står i fare for å få diagnostisert en høyresidig hjerneskade, tilsier stor tilbakeholdenhet når det gjelder å anvende diskrepans-grenseverdiene.

CalCAP

Normgrunnet for CalCAP består utelukkende av menn. Likevel er spesifisering av kjønn en av bakgrunnsvariablene som skal legges inn forut for kjøring av testprogrammet. Vår undersøkelse viser at kvinner responderer noe senere enn menn. De kliniske implikasjonene av funnene er at positive funn på CalCAP har stor troverdighet når det gjelder menn, og at prøven også kan brukes for kvinner.

Jens Egeland

Forsknings- og utviklingsenheten

Psykiatrien i Vestfold HF

Pb. 2267, 3103 Tønsberg

Tlf 33 01 80 00

E-post Jens.Egeland@piv.no

Referanser

- Andreassen, T. H., & Øksenholt, S. I. (2002). *Våle testen. En norsk test på hukommelse og læringsevne*. Tolvsrød, Tønsberg: Psykologtjenester.
- Baddeley, A. (1999). *Essentials of human memory*. West Sussex, UK: Psychology Press.
- Boringa, J. B., Lazeron, R. H. C., Reuling, I. E. W., Adler, H. J., Pfenning, L. E. M. A., Lindeboom, J., de Sonneville, L. M. J., Kalkers, N. F., & Polman, C. H. (2001). The brief repeatable battery of neuropsychological tests: Normative values allow application in multiple sclerosis clinical practice. *Multiple Sclerosis*, 7, 263–267.
- Bosnes, O., & Ellertsen, B. (2003). California Verbal Learning Test og Wechsler Memory Scale-revised: Prestasjoner i et klinisk utvalg i Norge. *Tidsskrift for Norsk Psykologforening*, 40, 590–597.
- Brager-Larsen, L. M., Sundet, K., Engvik, H., Ørbeck, B., & Bang, R. N. (2001). Psychometric properties of a Norwegian research version of the Wechsler abbreviated scale of intelligence. *7. Nordic Meeting in Neuropsychology*, Oslo. Poster-presentasjon.

- Channon, S., Baker, J. E., & Robinson, M. M. (1993). Working memory in clinical depression: An experimental study. *Psychological Medicine*, *23*, 87–91.
- Delis, D. C., Kramer, J. H., Freeland, J., & Kaplan, E. (1998). Integrating clinical assessment with cognitive neuroscience: Construct validation of the California Verbal Learning Test. *Journal of Consulting and Clinical Psychology*, *56*, 123–130.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test (CVLT) manual*. New York: The Psychological Corporation.
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test, Second Edition (CVLT-II) Adult Version, manual*. New York: The Psychological Corporation.
- Egeland, J., Rund, B. R., Sundet, K., Landrø, N. I., Asbjørnsen, A., Hugdahl, K., Lund, A., Roness, A., & Stordal, K. I. (2004). Prosesseringstempo og oppmerksomhet ved schizofreni og depresjon. *Tidsskrift for Norsk Psykologforening*, *41*, 552–560.
- 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., Sundet, K., Rund, B. R., Asbjørnsen, A., Hugdahl, K., Landrø, N. I., Lund, A., Roness, A., & Stordal, K. I. (2003). Sensitivity and specificity for memory dysfunction in schizophrenia: A comparison with major depression. *Journal of Clinical and Experimental Neuropsychology*, *25*, 79–93.
- Golden, C. J. (1978). *Stroop color and word test*. Manual, Odessa, FL: Psychological Assessment Resources.
- Gronwall, D., & Wrightson, P. (1975). Cumulative effects of concussion. *Lancet*, sat 22. November, 995–997.
- Kramer, J. H., Delis, D. C., & Daniel, M. (1988). Sex differences in verbal learning. *Journal of Clinical Psychology*, *44*, 907–915.
- Landrø, N. I., Gulowsen, C., & Sletvold, H. (2004). Depressive symptoms account for deficient information processing speed but not for impaired working memory in early phase multiple sclerosis. *Journal of the Neurological Sciences*, *217*, 211–216.
- Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). Oxford, UK: Oxford University Press.

- Lund-Johansen, M., Hugdahl, K., & Wester, K. (1996). Cognitive function in patients with Parkinson's disease undergoing stereotaxic thalamotomy. *Journal of Neurology, Neurosurgery & Psychiatry, 60*, 564–571.
- Lundervold, A. L., Reinvang, I., & Lundervold, A. (1994). Characteristic patterns of verbal memory functions in patients with Huntington's disease. *Scandinavian Journal of Psychology, 35*, 38–47.
- Meyers, J. E., & Meyers, K. R. (1995). *Rey Complex Figure Test and Recognition Trial – Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Meyers, J. E., & Meyers, K. R. (1996). *Supplemental normative data for Children*. Odessa, FL: Psychological Assessment Resources.
- Miller, E. N. (1993). *Calcap: California Computerized Assessment Package Manual (rev. 4.0)*. Los Angeles, CA: Eric N. Miller, PhD. and Norland Software.
- Mitrushina, M. N., Boone, K. B., & D'Elia, L. F. (1999). *Handbook of normative data for neuropsychological assessment*. New York: Oxford University Press.
- Osterrieth, P. A. (1944). Le test de copie d'une figure complexe. *Archives de Psychologie, 30*, 206–356.
- Paolo, A. M., Tröster, A. I., & Ryan, J. L. (1997). California Verbal Learning Test: Normative data for the elderly. *Journal of Clinical & Experimental Neuropsychology, 19*, 220–234.
- Rey, A. (1941) L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives de Psychologie, 28*, 286–340.
- Talley, J. L. (1992). *Children's Auditory Verbal Learning Test-2*. Odessa, FL: Psychological Assessment Resources.
- Tjemsland, E., Walbækken, K., & Egeland, J. (2001). Norwegian norms for phonetic (FAS) and semantic (animals & clothes) word production stratified by education level. 7. *Nordic Meeting in Neuropsychology*, Oslo. Poster-presentasjon.
- Warrington, E. K. (1984). *Recognition Memory Test*. Windsor, U.K.: NFER-Nelson.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. New York: The Psychological Corporation.
- Øie, M., Sundet, K., & Rund, B. R. (1999). Contrasts in memory functions between adolescents with schizophrenia or ADHD. *Neuropsychologia, 37*, 1352–1358.