

Testkvalitetsprosjektet - del 1: Norske psykologers testholdninger og testbruk



Tanya Marie Ryder

Barne- og ungdomspsykiatrisk avdeling, Sykehuset i Vestfold Postboks 2168, 3103

Tønsberg, Norge

tanya.ryder@siv.no

Denne undersøkelsen viser behov for forbedringstiltak i testfeltet. Testretningslinjer blir ikke fulgt opp tilstrekkelig, og det er vanskelig å finne informasjon om testenets psykometriske egenskaper.

Norsk psykologforeningen (NPF) har lenge vært bekymret for testfeltet og ment at flere aktører må bidra til å heve kvaliteten i et landskap som er preget av manglende oversikt over verktøy, og manglende etterlevelse av internasjonale standarder og retningslinjer for testing. Med bakgrunn i tidligere undersøkelser som viste kvalitetsproblemer i testfeltet, inngikk NPF høsten 2017 et samarbeid med Helsedirektoratet og Folkehelseinstituttet (FHI) om en nyere kartlegging av testfeltet, der 1400 norske psykologer rapporterte om sine holdninger til testkvalitet og om sitt testbruk. Hensikten var å gi en oppdatert oversikt over utfordringer i testfeltet og å identifisere hvilke tester som var mye brukt og samtidig hadde problemer knyttet til opphavsrett, oversettelse, norsk normdata, reliabilitet eller validitet, for å kunne vurdere mulige korrigerende tiltak. Denne artikkelen beskriver prosjektets første del: undersøkelsens utforming, resultatene om psykologers holdninger til testkvalitet, hvilke tester de anvender, samt informasjon avdekket om testenets kvalitetsstatus på Helsebiblioteket, PsykTestBarn og FHI. Resultatene om de hyppigst anvendte testene med kvalitetsspørsmål blir presentert i Testkvalitetsprosjekt del 2 (Ryder, under utgivelse).

Bakgrunn

I Norge utgis et økende antall psykologiske tester, spørreskjemaer, sjekklistes, intervjuer og kartleggingsverktøy som brukes i utredninger innen helse, utdanning og yrkesliv. (Heretter omtales alle disse verktøyene som tester.) Testfeltet er uoversiktlig med mange tester tilgjengelig både via formelle og uformelle kanaler, som testforlag, bøker, forskningsartikler og gjennom kolleger. De fleste testene er utviklet i andre land og er ikke nødvendigvis tilpasset norske forhold.

Funn fra tidligere norske undersøkelser har vist klart problematisk virksomhet i testfeltet. I 2006 gjennomførte NPF (Selvik, 2007) en kvalitetsanalyse av 17 tester brukt innen psykisk helsevern. Konklusjonene avdekket uforsvarlig testbruk som gjaldt «oversettelse, avtale med rettighetshaver, manglende normering, økonomiske/administrative forhold og kopiering» (s. 22). I 2009 deltok 935 medlemmer av NPF i en testundersøkelse der det ble vist til utstrakt bruk av tester uten offisielle norske versjoner, og som ikke hadde gjennomgått undersøkelser av psykometriske egenskaper for norske testtakere (se Vaskinn et al., 2010 og Vaskinn et al., 2012). Også Kunnskapssenteret for helsetjenesten / FHI og RBUP Øst og Sør, i samarbeid med de regionale kunnskapssentrene for barn og unge (PsykTestBarn), har påvist i sine omfattende kartlegginger av kvaliteten på testverktøyene at det er utilstrekkelig informasjon om psykometriske egenskaper og manglende norsk tilpasning av tester hyppig brukt i Norge (Kornør et al., 2015). Disse funnene viste at testpraksis hos både psykologer og testutvikler stred mot internasjonale testretningslinjer.



I 2000 vedtok NPF å gjøre International Test Commissions (ITC) internasjonale retningslinjer for testbruk (ITC, 2001) gjeldende som standarder for norske psykologer. Retningslinjene beskriver testbrukerens ansvar for etisk testbruk, inkludert respekt for opphavsrett. Videre står det at man skal velge «teknisk godt utformede tester som passer til situasjonen» (ITC, 2001, side 13). Dette innebærer at testbrukere skal ta stilling til om det finnes tilstrekkelig informasjon om testens reliabilitet, representativitet av normgruppen i forhold til den målgruppen testen skal benyttes for, samt validitet. Det spesifiseres at man skal «unngå bruk av tester som har inadekvat eller uklar støtte i teknisk dokumentasjon» (side 13).

ITC har også retningslinjer som gjelder oversettelse og tilpasning av tester (ITC, 2017, 2. utgave) som spesifiserer at testutviklere og testforlag skal ha fulgt en standardisert oversettelsesprotokoll og gjort tilgjengelig en manual der kulturell tilpasning av språk og kultur for målgruppen er redegjort for. Testen skal være psykometrisk kvalitetssikret med en undersøkelse av normer, reliabilitet og validitet som er beskrevet i manualen. Videre er European Federation of Psychologists' Associations (EFPA) Test Review Model (2013, v. 4.2.6) en gullstandard for testutvikling. Protokollen beskriver relevante kvalitetsaspekter ved tester og gir en detaljert redegjørelse av formelle krav som stilles til dokumentasjonen av testers psykometriske egenskaper og hvordan standardiseringen (inklusive normdatainnsamlingen) har blitt gjennomført. For å kunne være ansvarlige testbrukere etter ITCs internasjonale retningslinjer for testbruk (ITC, 2001) er norske psykologer avhengig av tilstrekkelig teknisk dokumentasjon for å kunne ta stilling til hvordan tester presterer i en norsk kontekst. For mange tester som er i bruk i Norge, er denne dokumentasjonen utilstrekkelig eller fraværende (Kornør et al., 2015). Dette skaper et dilemma for norske psykologer, som kan oppleve mangel på kvalitetssikrede tester. Situasjonen kan også utgjøre en pasientsikkerhetsrisiko når man tar i bruk verktøy som ikke fungerer som intendert.

Helsedirektoratet ga i 2008 Kunnskapssenteret for helsetjenesten / FHI og RBUP Øst og Sør, i samarbeid med de regionale kunnskapssentrene for barn og unge (PsykTestBarn), oppdraget med å gjennomføre grundige, systematiske vurderinger av norske versjoner av tester. FHI og PsykTestBarn gjør fortløpende vurderinger av tester i henhold til en tilpasset versjon av EFPA's Test Review Model (2013, v. 4.2.6). Det foretas en grundig undersøkelse av om norske eller skandinaviske studier og annen faginformasjon som testmanualer belyser: om opphavsrett til testen er avklart, om godkjente oversettelsesprosedyrer er fulgt, om norsk normering er gjennomført og for hvilke grupper, samt om funn rundt ulike typer validitet og reliabilitet. Arbeidet som FHI og PsykTestBarn gjennomfører, viser også at det er lite kvalitetssikrede data om tester i Norge, og at kulturelle faktorer ved tester gjør konklusjoner i litteraturen fra andre land usikre (FHI, personlig kommunikasjon, 29. september, 2017). Mens FHI's/PsykTestBarns systematiske vurderinger av tester er et nyttig bidrag til testdatabasen, er det et krevende arbeid, og ressursene har ikke vært tilstrekkelig til å vurdere det antallet tester som behøves i et raskt voksende testfelt.

Målet for prosjektet

Med bakgrunn i de tidligere undersøkelsene av norske psykologers testbruk (Selvik, 2007; Vaskinn et al., 2010 og Vaskinn et al., 2012) og kvalitetsundersøkelser utført av FHI og PsykTestBarn (se Kornør et al., 2015) så NPF behov for oppdatert informasjon om testbruk, slik at kvalitetsforbedringstiltak kunne fokuseres der det trengtes mest i samarbeid med aktuelle samarbeidspartnere. NPF tok derfor initiativ til å legge frem en prosjektskisse for Helsedirektoratet og FHI, som begge jobber med pasientsikkerhet, og som har vært pådrivere for testundersøkelsesarbeid. Det ble så avtalt at Helsedirektoratet bidrar

med prosjektmidler til en bred kartlegging, utført av NPF, av hvilke tester norske psykologer bruker, for å kunne identifisere hvilke tester som var mest brukt og samtidig hadde større problemer knyttet til opphavsrett, oversettelse, norsk normdata, reliabilitet eller validitet. I tillegg ble det lagt til spørsmål om psykologenes holdninger til testkvalitet. Hensikten var å gi Helsedirektoratet et grunnlag for å kunne identifisere mulige korrigerende tiltak og igangsette samarbeid for å støtte relevante aktører.



Metode

I undersøkelsen ble 3 hovedkategorier av data samlet inn:

1) Bakgrunnsinformasjon. I tillegg til demografiske data ble det samlet inn opplysninger om psykologenes utdanningsår, spesialiststatus, hovedarbeidssted og fagfelt.

2) Informantenes holdninger og praksis relatert til testkvalitet. Vi inkluderte spørsmål fra en av de tidligere undersøkelsene av norske psykologers testbruk og testholdninger utarbeidet av European Federation of Psychologists' Associations (EFPA) i samarbeid med NPF (Vaskinn et al., 2010 og Vaskinn et al., 2012).

3) Informasjon om de mest anvendte testene. Vi utarbeidet en liste med over 350 tester som er tilgjengelige eller i bruk i Norge, ved å innhente informasjon fra ulike kilder: ni helseforetak/ behandlingsenheter, NAV sentralt, Pedagogisk-Psykologisk Tjeneste Bærum, DNV GL, Selviks rapport fra 2007, EFPA/Psykologforeningens undersøkelse fra 2009 (Vaskinn et al., 2012), ulike lokale og nasjonale retningslinjer for utredning og ulike internettkilder (som blant andre Helsebiblioteket.no, R-BUP.no, Psyktestbarn.no, Folkehelseinstituttet.no, Pearson.no, Hogrefe.no og Statped.no). Resultatet ble en liste som inkluderer utviklingstester, evnetester, kognitive funksjonstester / nevropsykologiske tester, diagnostiske verktøy, kartleggingsskalaer, kliniske intervjuer samt arbeids- og organisasjonstester. Hovedvekten av testene på listen er kliniske tester brukt i helsetjenesten. Noen tester ble vurdert som mindre relevante for prosjektets fokusområde og ble derfor utelatt fra listen for å redusere trettheten ved utfylling av undersøkelsen. Eksempler på dette er forholdsvis enkle symptomscreenere (som GAD-7), subjektive setningsutfyllingsskjemaer (som Rotter Incomplete Sentences Blank), tester som ikke fantes på norsk (som Minnesota Multiphasic Personality Inventory-Adolescent), tester brukt hovedsakelig av pedagoger fremfor psykologer (som Dansk Impresiv Morfologisk Test) og tester med rimelig begrensede målgrupper (som Asperger Syndrome Workplace Interview).

Informanter

NPF sendte ut den elektroniske undersøkelsen til 4915 av sine medlemmer i februar 2018. Vi ekskludert studenter og pensjonister samt psykologer som nylig hadde mottatt en annen undersøkelse fra NPF. Psykologene ble bedt om å indikere hvor ofte de brukte hver test: flere ganger i uken, flere ganger i måneden, flere ganger i kvartalet, sjeldnere enn hvert kvartal og sjeldnere enn hvert år. Vi analyserte 1406 svar, noe som ga en svarprosent på 29.

Analyser

Vi gjorde en beskrivende analyse av testenenes samlede bruksfrekvens ved å summere antall svar i kategorier «flere ganger i uken», «flere ganger i måneden» og «flere ganger i kvartalet».

Resultater



Bakgrunnsinformasjon

Av de rundt 51 % av informantene som oppga informasjon om kjønn, alder og spesialiststatus var 72 % kvinner og 54 % spesialister. I utvalget var 68 % 50 år eller yngre, og alderskategorien med flest psykologer var 31–35 år. Flere som hadde utdannet seg mer nylig, besvarte undersøkelsen.

Tabell 1 viser at flertallet av informantene jobbet i spesialisthelsetjenesten (62 %), med 48 % innenfor psykisk helsevern (29 % med voksne, 19 % med barn og unge). Alle unntatt 11 deltagere oppga sitt fagfelt. Den største enkeltgruppen jobbet med «barn og unge» (25 %). Til sammenlikning jobbet 18 % med voksne og 18 % med psykoterapi.

Tabell 1. **Respondenters hovedarbeidssted**

Hovedarbeidssted	Antall	Prosent
Psykisk helsevern for voksne	413	29,4
Psykisk helsevern for barn og unge	260	18,5
Privat praksis med driftsavtale	102	7,3
Tverrfaglig spesialisert rusbehandling	101	7,2
Primærhelsetjeneste/kommunehelsetjeneste	84	6,0
Somatisk helseavdeling	74	5,3
Privat praksis uten driftsavtale	68	4,8
Pedagogisk-psykologisk tjeneste	65	4,6
Kompetansetjeneste eller annen offentlig tjeneste	26	1,8
Privat næringsliv	25	1,8
NAV (Arbeids- og velferdsforvaltningen)	22	1,6
Privat stiftelse	22	1,6
Familievernkontor	19	1,4
Statlig barnevern	19	1,4
Kommunalt barnevern	10	0,7
Bedriftshelsetjeneste	4	0,3
Annet	90	6,4
Totalt antall svar	1404	99,9
Uteblitte svar	2	0,1

Informantenes holdninger og praksis relatert til testkvalitet

Totalt 715 besvarte spørsmålet «I hvor stor grad legger du vekt på at testene/kartleggingsverktøyene du bruker er kvalitetssikret i forhold til normering, reliabilitet og validitet?». Av disse svarte 68 % «i stor grad», 30 % svarte «i noen grad» og 2 % svarte «i liten grad». Til spørsmålet «I hvor stor grad legger du vekt på at testene/kartleggingsverktøyene du bruker er kvalitetssikret i forhold til opphavsrettigheter

(copyright)?» svarte 37 % «i stor grad» og 41 % «i noen grad», mens 22 % var i liten grad opptatt av opphavsrettigheter.



Informantene ble gitt mulighet til å krysse av for alle de fire svaralternativene på spørsmålet «Hvordan forsikrer du deg om at testene/kartleggingsverktøyene du bruker er kvalitetssikret i forhold til opphavsrettigheter, normering, reliabilitet og validitet?». Svarene fordelte seg nokså jevnt mellom «Stoler på at leverandør har kvalitetssikret produktene sine» (n = 321), «Stoler på at arbeidsgiver har kvalitetssikret de verktøyene som brukes på min arbeidsplass (n = 289) og «Sjekker personlig at verktøyene jeg bruker er kvalitetssikret» (n = 318). Svaralternativet «Har ikke noe bevisst forhold til dette» ble avkrysset av 35 av informantene. I fritekstfeltet ble det registrert 59 kommentarer som omhandlet alt fra at man ikke var opptatt av opphavsrett, til at man brukte ulike kilder som PsykTestBarn, Helsebiblioteket, testmanualer og vitenskapelige artikler for å finne informasjon om testkvalitet. Noen kommentarer refererte til arbeidsgivers og psykologers manglende oversikt over tester av god kvalitet.

Majoriteten av informantene oppga at de bruker tester i sitt arbeid (90 %, n = 1271). Av disse oppga 49 % at de bruker tester flere ganger i uken, 36 % flere ganger i måneden og 10 % flere ganger i kvartalet.

De 40 mest brukte testene (se Tabell 2) er beskrevet under i ulike testkategorier:

Tabell 2. De 40 mest brukte testene

Test	Antall psykologer som bruker testen minst flere ganger i kvartalet (N = 1406)
1. Beck Depression Inventory-II (BDI-II)	299
2. Beck Anxiety Inventory (BAI)	223
3. Mini Internasjonalt Nevropsykiatrisk Intervju plus (M.I.N.I plus)	207
4. Wechsler Adult Intelligence Scale-Fourth edition (WAIS-IV)	198
5. Symptom Checklist 90-Revised (SCL-90-R)	197
6. Adult ADHD Self-Report Scale-v1.1 (ASRS-v1.1)	195
7. Mini Internasjonalt Nevropsykiatrisk Intervju (M.I.N.I.).	192
8. Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)	159
9. Behavior Rating Inventory of Executive Function (BRIEF; parent and teacher version; 5-18 år)	147
10. Achenbach System of Empirically Based Assessment – Child Behavior Checklist (ASEBA – CBCL)	144
11. Montgomery Åsberg Depression Rating Scale (MADRS)	144
12. Alcohol Use Disorder Identification Test (AUDIT)	142
13. Achenbach System of Empirically Based Assessment – Teacher Report Form (ASEBA – TRF)	138
14. Conners Continuous Performance Test 3 (Conners CPT 3)	134
15. Kartlegging av traumeerfaringer-Barn/Foreldre (KATE-B/F)	130
16. Wechsler Intelligence Scale for Children-Fourth edition (WISC-IV)	119
17. Achenbach System of Empirically Based Assessment – Youth Self-Report (ASEBA – YSR)	118
18. Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V)	117
19. Diagnostisk intervju for utredning av ADHD hos voksne (DIVA 2.0)	117
20. Behavior Rating Inventory of Executive Function-Self-Report Version (BRIEF-SR; 11-18 år)	111
21. Drug Use Disorders Identification Test (DUDIT)	110
22. Achenbach System of Empirically Based Assessment – Caregiver-Teacher Report Form (ASEBA – C-TRF)	105
23. Delis-Kaplan Executive Function System (D-KEFS)	101
24. Behavior Rating Inventory of Executive Function-Adult Version (BRIEF-A; 18–90 år)	99
25. Child and Adolescent Trauma Screen (CATS)	97
26. ADHD Rating Scale-IV: Hjemmeversjon (ADHD-RS-IV: Hjemme)	95
27. ADHD Rating Scale-IV: Skoleversjon (ADHD-RS-IV: Skole)	91
28. Rey Complex Figure Test and Recognition Trial (RCFT)	88
29. California Verbal Learning Test-II (CVLT-II)	83
30. Schedule for Affective Disorders and Schizophrenia for School Aged Children (6–18 years): Present and Lifetime Version (Kiddie-SADS-PL)	78
31. Wisconsin Card Sorting Test (WCST)	70
32. Behavior Rating Inventory of Executive Function-Preschool Version (BRIEF-P; 2–5:11 år)	69
33. Vineland Adaptive Behavior Scales-Second edition (Vineland-II)	68
34. Wechsler Preschool and Primary Scale of Intelligence-Fourth edition (WPPSI-IV)	64
35. Norsk revidert Mini Mental Status Evaluering (MMSE-NR3)	61
36. Autism Diagnostic Interview-Revised (ADI-R)	61
37. Wechsler Memory Scale-III (WMS-III)	58
38. Impact of Event Scale-Revised (IES-R)	58
39. FAS test (COWAT; verbal flyt)	57
40. Autism Spectrum Screening Questionnaire (ASSQ)	56

De mest anvendte testene



Kartleggingsskjemaer. Mest brukt av alle tester i listen var Beck Depression Inventory (BDI-II; Beck & Steer, 2005) og Beck Anxiety Inventory (BAI; Beck, Steer & Brown, 2005), som brukes til kartlegging av symptomer en ser relativt ofte hos ungdom og voksne i psykisk helsevern. Blant kartleggingsskjemaer som brukes til å screene for et bredt spekter av problematikk, var Symptom Checklist 90-Revised (SCL-90-R; Derogatis, 2010) for voksne samt fire versjoner i Achenbach System of Empirically Based Assessment (ASEBA) for barn og unge (Achenbach et al., 2000, 2001).

Kartleggingsskjemaer som ofte anvendes i utredning av mulig ADHD-problematikk inkluderte Adult ADHD Self-Report Scale-v1.1 (ASRS-v1.1; Kessler et al., 2005), fire versjoner av Behavior Rating Inventory of Executive Function (BRIEF; Parent and Teacher Version, Gioia et al., 2000; Self-Report Version, Guy et al., 2004; Adult Version, Roth et al., 2005 og Preschool Version, Gioia et al., 2003) og hjemme- og skoleversjoner av ADHD Rating Scale-IV (ADHD-RS-IV; DuPaul et al., 1998). To kartleggingsskjemaer brukt til screening av rusproblematikk ble rapportert: Alcohol Use Disorder Identification Test (AUDIT; Babor et al., 2001) og Drug Use Disorders Identification Test (DUDIT; Berman et al., 2003). Videre ble det oppgitt tre skjemaer brukt i utredning av traumer: Kartlegging av traumeerfaringer-Barn/Foreldre (NKVTS, 2015a/2015b), Child and Adolescent Trauma Screen (Sachser et al., 2017) og Impact of Event Scale-Revised (Weiss et al., 1997). De andre kartleggingsskjemaene blant de 40 mest brukte var Montgomery Åsberg Depression Rating Scale (MADRS; Montgomery et al., 1979), utviklet for å følge symptomendringer ved bruk av antidepressiv medisin, og Autism Spectrum Screening Questionnaire (Ehlers et al., 1999).

Kliniske intervjuer. Mini Internasjonalt Nevropsykiatrisk Intervju, plus- og standardversjoner (M.I.N.I.; Sheehan et al. 1998), var de mest brukte kliniske intervjuene. Andre kliniske intervjuer for voksne inkluderte Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II; First & Gibbon et al., 1997) og Diagnostisk intervju for utredning av ADHD hos voksne 2.0 (Kooij et al., 2010). Schedule for Affective Disorders and Schizophrenia for School Aged Children (6 - 18 years): Present and Lifetime Version (Kiddie-SADS-PL; Kaufman, 1997) var det mest brukte kliniske intervju for barn/unge. De to andre hyppig brukte intervjuene var Vineland Adaptive Behavior Scales-Second edition (Sparrow et al., 2011) og Autism Diagnostic Interview-Revised (Rutter et al., 2003), som brukes i utredning av henholdsvis adaptive ferdigheter og autismespektervansker.

Intelligenstester. Blant de 40 mest brukte testene som fremkom i undersøkelsen, var det kun Wechsler-testene som ble anvendt til utredning av intelligens. Wechsler Adult Intelligence Scale-Fourth Edition (Wechsler, 2011) kom på fjerdeplass blant alle testene. For barn og unge var Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV; Wechsler, 2009) og den nyere WISC-V (Wechsler, 2017), som ble publisert på norsk året før undersøkelsen ble sendt ut, brukt i nesten like stor grad. Wechsler Preschool and Primary Scale of Intelligence-Fourth edition (Wechsler, 2015) til bruk hos barn opptil 7 år og 7 måneder var noe mindre brukt enn de andre, antagelig grunnet mindre målgruppe.

Nevropsykologiske tester. De mest utbredte nevropsykologiske testene brukes til å vurdere kognisjon, eksekutiv fungering, orientering, hukommelse, visuospatiale evner, oppmerksomhet, innlæring og verbal flyt. Flere av disse testene kan brukes både for barn/unge og voksne. Blant tester som kan anvendes i utredning av barn fra 6 års alder fant vi Rey Complex Figure Test and Recognition Trial (RCFT; Meyers et al., 2004) og Wisconsin Card Sorting Test (WCST; Heaton et al., 1993).

Innrapporterte tester som kan brukes fra 8 års alder var Conners Continuous Performance Test 3 (Conners CPT-3; Conners, 2014), Delis-Kaplan Executive Function System (D-KEFS; Delis et al., 2005) og FAS Test. FAS Test, opprinnelig utviklet av Benton et al. som Controlled Oral Word

Association test (1989), kan brukes som en selvstendig test, men den finnes også som en del av D-KEFS nevropsykologisk batteri (D-KEFS Verbal Flyt – Ordflyt av Delis et al., 2005).



California Verbal Learning Test-II (CVLT-II; Delis et al., 2004) og Wechsler Memory Scale-III (WMS-III; Wechsler, 2008) kan anvendes i utredning av personer i alderen 16 år og over. Norsk revidert Mini Mental Status Evaluering (MMSE-NR3; Strobel et al., 2008) brukes i vurdering av ulike typer kognitiv svikt, oftest hos voksne.

Ytterligere analyser. Flere av testene som var i utstrakt bruk i 2009 (Vaskinn et al., 2012), var det også i 2018: Wechsler tester, BDI-II, BAI, begge versjonene av M.I.N.I., SCL-90-R, ASRS-v1.1, SCID-II, BRIEF, ASEBA, MADRS, AUDIT, DUDIT, Conners CPT (nå i sin tredje utgave), begge versjoner av ADHD-RS-IV, RCFT, CVLT-II, Kiddie-SADS-PL, WCST, MMSE (nå i sin tredje versjon) og WMS (også i sin tredje versjon). Imidlertid ser det ut til at noen tester brukes langt mindre nå enn det som ble rapportert om i 2009. Blant disse er Ravens Matriser, hvor 38 % av informantene oppga å bruke dette i 2009 mot 7 % i 2018 (Raven, 1998). Minnesota Multiphasic Personality Inventory-2 ble rapportert i bruk av 27 % i 2009 mot 9 % i 2018 (Butcher et al., 2004) og NEO Personality Inventory-R ble brukt av 12 % i 2009, mens NEO-PI-3/NEO-FFI-3 ble brukt av 3 % i 2018 (McCrae et al., 2017). Siden testundersøkelsen i 2009 har det kommet noen nye tester på norsk som i dag er i hyppig bruk, deriblant nyere versjoner av Wechsler intelligens tester, Kartlegging av traumeerfaringer-Barn/Foreldre (NKVTS, 2015a/2015b), Child and Adolescent Trauma Screen (Sachser et al., 2017) og Diagnostisk intervju for utredning av ADHD hos voksne (DIVA 2.0; Kooij et al., 2010).

FHI og PsykTestBarn har gjort tilgjengelig, frem til begynnelsen av juli 2020, 15 systematiske vurderinger blant de 40 mest brukte testene fra denne undersøkelsen. Samtidig ble det gjennomført 15 slike vurderinger av tester som ikke var i bruk i det hele tatt hos respondentene, og 14 vurderinger av tester brukt av kun 8 eller færre psykologer (minst flere ganger i kvartalet, tester på plass 170 eller lavere).

Diskusjon

Mens denne undersøkelsen har en høyere svarprosent (29 %; N = 1406) enn Vaskinn og kollegas tilsvarende testundersøkelse fra 2009 (2010 og 2012; 15 %, N = 943), er den likevel å betrakte som relativt lav. Dette kan antyde et visst seleksjonsproblem hvor primært psykologer med særlig interesse for testing deltok i undersøkelsen. Lav svarprosent gjør det vanskelig å anslå hvorvidt funnene fra undersøkelsen er representative. Psykologer dekker en stor bredde når det gjelder spesifikke fagområder, aktuelle pasient-/klientpopulasjoner og personlige erfaringsbakgrunner. Denne spennvidden i virkefelt vil nødvendigvis medføre en viss heterogenitet når det gjelder testbruk, og det kan således argumenteres at en full representativitet hverken er et mulig eller hensiktsmessig mål for en undersøkelse som denne. Materialet er imidlertid av en slik størrelse og samsvarer med data fra andre kilder i en slik grad at det gir et illustrerende bilde av testbruk blant psykologer.

Selv om de alle fleste av informantene oppga at de bruker tester i sitt arbeid, besvarte kun halvparten etterfølgende spørsmål om testholdninger. Av informantene som svarte på spørsmål om testholdninger, oppga 98 % at de i stor eller noen grad er opptatt av at testene de bruker er kvalitetssikret med hensyn til normering, reliabilitet og validitet. Færre i dette utvalget (78 %) ga tilsvarende rangering på spørsmålet om hvorvidt man er opptatt av om testene man bruker er avklart med hensyn til opphavsrettigheter. At færre av psykologene som deltok i undersøkelsen rangerer spørsmålet relatert til opphavsrett som viktig, fremstår som et paradoks, da opphavsrett er relatert til psykometriske kvaliteter ved tester. Bruk av verktøy der manualen mangler, for eksempel ved

anvendelse av fotokopierte spørreskjemaer og intervjuer av ukjent opphav, medfører uunngåelig at psykologen ikke har noen mulighet for å kontrollere testens psykometriske kvaliteter. Manualer inneholder rettledning i testadministrasjon, normtabeller og beskrivelse av normdatagrunnlaget og som regel resultat fra validitets- og/eller reliabilitetsstudier. Uten manual har således ikke testbruker noe grunnlag for å vurdere om en test er tilpasset eller egnet for den populasjonen den aktuelle testtaker tilhører.

Problemer knyttet til opphavsrett har tidligere blitt omtalt i Tidsskriftet for Norsk Psykologforening (Vaskinn et al., 2012; Egeland et al., 2015), og våre funn viser at situasjonen ikke har bedret seg de siste årene. SCID-II (First & Gibbon et al., 1997) har i mange år vært, og er fortsatt, blant de hyppigst anvendte testene i Norge, til tross for at norske utgaver av dette intervjuet aldri har vært autorisert og ikke skal brukes eller distribueres (American Psychiatric Association Publishing, personlig kommunikasjon, 07. desember, 2020; Selvik, 2007; Ryder, under utgivelse). Å bruke SCID-II er i strid med ITCs internasjonale retningslinjer for testbruk som NPF har vedtatt som rettledende. At tester som SCID-II brukes av mange psykologer indikerer at det fortsatt eksisterer en problematisk avstand mellom testbrukeres praksis og etikk, lovverk og rettledende prinsipper for testsikkerhet.

Ved spørsmålet om «Hvordan forsikrer du deg om at testene/kartleggingsverktøyene du bruker er kvalitetssikret i forhold til opphavsrettigheter, normering, reliabilitet og validitet?» ser man en endring fra resultatene i Vaskinn og kollegas testbruksundersøkelse fra 2009, hvor 42 % av informantene oppga at de selv tok ansvar for å sjekke testkvalitet, og 20 % oppga at de stolte på at arbeidsgiver hadde kvalitetssikret testene (2010). I vår undersøkelse falt henholdsvis 33 % og 30 % av svarene på disse alternativene, noe som indikerer at psykologer tar mindre selvstendig ansvar enn for ti år siden og lener seg mer på arbeidsgivers vurderinger eller bestemmelser. Funnet kan imidlertid være uttrykk for ulike forhold, som at testfeltet oppfattes mer uoversiktlig, å foreta selvstendige kvalitetsvurderinger oppleves mer utfordrende, at det er mindre personlig engasjement eller bevissthet rundt valg av tester blant psykologer, eller økt tillit til arbeidsgivers faglige vurderinger. I tillegg kan tendensen knyttes til at det ved flere klinikker og institusjoner legges føringer fra arbeidsgiver om hvilke tester som kan eller skal brukes.

Hvem som egentlig har ansvar for kvalitetssikring av tester fremstår som uklart og omstridt. Helsedirektoratet har i brev til Helse- og omsorgsdepartementet plassert ansvar for kvalitetssikring av tester hos helseforetak og den enkelte testbruker (Helsedirektoratet, 2016). Her presiseres det at «Helsedirektoratet / offentlig helseforvaltning skal ikke ha en godkjenningsrolle av verktøy i bruk i Norge» og videre at «Helseforetakene og den enkelte helsearbeider som anvender et verktøy, må selv sørge for at bruken av verktøyet er innen rammen for «god klinisk praksis» og innenfor de rettighetsrammene som finnes». Flere problemer er knyttet til disse formuleringene. NPF har ment at det ikke er holdbart å legge alt ansvar for testing over på den enkelte testbruker når informasjon om testkvalitet er så manglende, og når det som finnes er så uoversiktlig. I det åpne svarfeltet i vår undersøkelse bekrefter informantene at både psykologer og arbeidsgivere mangler oversikt over hvilke tester som har hvilke psykometriske kvaliteter. Det var også uttrykt et ønske om en kvalitetssikret testbase. Helsebiblioteket.no, et offentlig finansiert nettsted drevet av FHI som tilbyr gratis tilgang til fagkunnskap, har påtatt seg oppgaven med å publisere ulike tester og lenker til tester. Helsebibliotekets tilknytning til FHI kommuniserer en kvalitetsgodkjenning fra offentlig helseforvaltning, ikke minst når det står at Helsebiblioteket «skal bidra til å heve kvaliteten på helsetjenestene og til å bedre pasientsikkerheten gjennom å gi enkel tilgang til oppdaterte og kvalitetsvurderte kilder på nett for alt helsepersonell i Norge». I vår undersøkelse ble Helsebiblioteket.no nevnt som en faglig kilde for testopplysninger. I den grad testbrukere benytter seg av denne nettressursen for å skaffe seg slik

informasjon, er det avgjørende at den faktisk evner å tilby oppdatert og kvalitetssikret informasjon som ikke bryter med åndsverkloven. Dette prosjektet viser dessverre at dette ikke alltid er tilfellet, som ved Helsebibliotekets tidligere lenke til den aldri-godkjente SCID-II og nåværende lenker til M.I.N.I. versjoner som er både utdaterte og som forfatteren ikke har gitt lov til å offentliggjøre (David Sheehan, personlig kommunikasjon, 11. desember, 2020; Ryder, under utgivelse). Helsedirektoratet har selv anbefalt bruk av SCID-II i nasjonale veiledere tross manglende tillatelse til å distribuere eller bruke norske utgaver av dette intervjuet (Helsedirektoratet, 2012 og 2013). Disse eksemplene viser at kvalitetssikring er en formidabel utfordring som krever enorme ressurser som den enkelte testbruker ikke har til rådighet. FHI og PsykTestBarn har i stor grad bidratt til tilgjengeliggjøring og økt forståelse for testers psykometriske egenskaper, slik at kvalifiserte testbrukere kan ta informerte valg om tester. Dessverre viser det seg at det brukes forskningsmidler på mange tester som ut fra informantenes svar viser seg å være, i beste fall, sjelden i bruk. Vår undersøkelse bidrar til informasjon om hvilke tester som bør prioriteres for slike systematiske kartlegginger.

Gjennom det siste tiåret kan det se ut som om utstrakt bruk av tester blant psykologgruppen har vært stabil, da 90 % av informantene oppga å bruke tester i sitt arbeid både i testbruksundersøkelsen fra 2009 (Vaskinn et al., 2010) og i 2018. Blant de 40 hyppigst anvendte testene i vår undersøkelse fant vi både spørreskjemaer, intervjuer og testleder-administrerte verktøy. Flertallet av instrumentene var kliniske tester, som er naturlig siden majoriteten av informantene jobbet i spesialisthelsetjenesten. Noen av testene brukes imidlertid også i skolevesenet. Kun 4 % av informantene i utvalget jobbet innenfor arbeids- og organisasjonspsykologi, og følgelig ble bruk av arbeidspsykologiske tester relativt sjelden rapportert om.

For nesten alle tester oppga en mindre andel av informantene i 2018 enn i 2009 at de brukte de ulike testene (Vaskinn et al., 2012). Noe av forklaringen på dette kan relateres til at det er flere tester tilgjengelig nå enn for 10 år siden, og en konsekvens av dette er at hver enkelt test blir brukt av færre psykologer. Vi fant også at enkelte tester i 2009 med kjente kvalitetsproblemer fortsatt ble profilert av Helsedirektoratet og Helsebiblioteket, og at de ble hyppig brukt av denne studiens informanter i 2018.

Konklusjon

Denne studien viser behovet for økt bevissthet om internasjonale testretningslinjer, både hos testutviklere og hos testbrukere. Mange av de testene som brukes av norske psykologer, er ikke nødvendigvis kvalitetssikret med hensyn til formell oversettelsesprotokoll, normeringsdata, reliabilitet og validitet (Ryder, under utgivelse). I tillegg gir det grunn til bekymring at det eksisterer en testbruksnorm som tillater omfattende misbruk av åndsverk, noe som bryter med internasjonale testretningslinjer. Videre kan resultatene tyde på at testfeltet er preget av manglende oversikt over kvalitetsinformasjon og utydelighet rundt hvem som har ansvar for kvalitetssikring av tester. Denne studien kan således tyde på at det er behov for tilgjengelig og oversiktlig teknisk testdokumentasjon fra testutviklere, slik at testbrukere kan ta informerte valg og ha tillit til testresultater. Et annet spørsmål knytter seg til hvor stort kvalitetsproblem som kan være heftet ved de mest anvendte testene. I del 2 av dette prosjektet er de 40 mest brukte testene gjennomgått for en belysning av de 10 testene med størst kvalitetsproblemer. Målet er å gi aktuelle aktører oversikt over hvilke tester som bør prioriteres for forbedringstiltak

Takk

Takk til Helsedirektoratet som bidro med midler som gjorde prosjektet mulig. Det takkes for skrivestipend fra Barne- og ungdomspsykiatrisk avdeling ved Sykehuset i Vestfold. Takk til Norsk psykologforening: TPU-medlemmer (2017–2019) for innspill til undersøkelsen og analysemetoder, Andreas Høstmælingen og Ole Tunold for bistand med forberedelse og utsendelse av undersøkelsen samt datainnsamling og databearbeidelse, sentralstyret for skrivestipend og Mikael J. Sømhovd for kommentarer til manuskriptet.

MERKNADER: Forfatteren har avgitt signert interessekonflikterklæring og oppgir ingen interessekonflikter. Innhold i rapporten er utarbeidet av forfatteren uten påvirkning fra de økonomiske bidragsyterne. Denne rapporten er ikke fagfellevurdert.

Referanser

- Achenbach, T. M., McConaughy, S., Ivanova, M. & Rescorla, L. (2011). *Manual for the ASEBA Brief Problem Monitor™ (BPM)*. Research Center for Children, Youth, and Families, University of Vermont.
- Achenbach, T. M. & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. Research Center for Children, Youth, and Families, University of Vermont.
- Achenbach, T. M. & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Research Center for Children, Youth, and Families, University of Vermont.
- Babor, T. F., Higgins-Biddle, J. C., Saunders, A. B. & Monteiro, M. G. (2001). *The Alcohol Use Disorders Identification Test: Guidelines for use in primary care* (2. utg.). The World Health Organization (WHO).
- Beck, A. T. & Steer, R. A. (2005). *Beck Anxiety Inventory*. Pearson Assessment.
- Beck, A. T., Steer, R. A. & Brown, G. K. (2005). *Beck Depression Inventory-Second Edition*. Pearson Assessment.
- Benton, A. L. & Hamsher, K. (1989). *Multilingual Aphasia Examination (2nd ed.)*. AJA Associates.
- Berman, A. H., Bergman, H., Palmstierna, T. & Schlyter, F. (2003). *The Drug Use Disorders Identification Test: Manual*. Karolinska Institutet. <https://doi.org/10.1037/t02890-000>
- Butcher J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G. & Kaemmer, B. (2004). *Minnesota Multiphasic Personality Inventory-2, norsk manual*. Hogrefe Psykologiförlaget AB.
- Conners, C.K. (2014). *Conners' Continuous Performance Test 3rd Edition*. Hogrefe Psykologiförlaget AB.
- Delis, D. C., Kaplan, E. & Kramer, J. H. (2005). *Delis-Kaplan Executive Function System, norsk versjon*. Pearson Assessment.
- Delis, D. C., Kramer, J. H., Kaplan, E. & Ober, B. A. (2004). *California Verbal Learning Test, Second Edition (CVLT-II), norsk versjon*. Pearson Assessment.
- Derogatis, L. R. (2010). *Symptom Checklist-90-R (SCL-90-R). Norsk versjon. Manual for administrering og skåring*. NCS Pearson, Inc. <https://doi.org/10.1002/9780470479216.corpsy0970>
- DuPaul, G. J., Power, T. J., Anastopoulos, A. D. & Reid, R. (1998). *ADHD Rating Scale-IV: Checklists, Norms, and Clinical Interpretation*. Guilford. <https://doi.org/10.1037/t00680-000>
- Egeland, J., van Delft, C., Føllestad, H., Holen, A., Helland, S. & Engen Nilsen, D. Ø. (2015). *Det må være et myndighetsansvar å sette nødvendige prosesser i gang for å sikre tilgang på kvalitetssikrede verktøy*. <https://www.psykologforeningen.no/medlem/testbruk/uautoriserte-tester-og-kartleggingsverktoy>

- Ehlers, S., Gillberg, C. & Wing, L. (1999). A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. *Journal of Autism and Developmental Disorders*, 29(2), 129–141. <https://doi.org/10.1023/A:1023040610384>
- European Federation of Psychologists' Associations (EFPA). (2013). *EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests: Test Review Form and Notes for Reviewers* (v. 4.2.6).
- First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W. & Benjamin, L. S. (1997). *Structured clinical interview for DSM-IV axis I personality disorders, (SCID-II)*. American Psychiatric Association.
- Gioia, G. A., Espy, K. A. & Isquith, P. K. (2003). *BRIEF-P: Behavior Rating Inventory of Executive Function-Preschool Version*. Psychological Assessment Resources.
- Gioia, G. A., Isquith, P. K., Guy, S. C. & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function*. Psychological Assessment Resources. <https://doi.org/10.1076/chin.6.3.235.3152>
- Guy, S., Isquith, P. & Gioia, G. (2004). *BRIEF-SR: Behavior Rating Inventory of Executive Function-Self-Report Version*. Psychological Assessment Resources.
- Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G. & Curtiss, G. (1993). *Wisconsin card sorting test manual: Revised and expanded*. Psychological Assessment Resources.
- Helsedirektoratet. (2012). *Nasjonal faglig retningslinje for utredning, behandling og oppfølging av personer med samtidig ruslidelse og psykisk lidelse – ROP-lidelser*. <https://www.helsedirektoratet.no/retningslinjer/samtidig-ruslidelse-og-psykisk-lidelse-rop-lidelser>
- Helsedirektoratet. (2013). *Nasjonal faglig retningslinje for utredning, behandling og oppfølging av personer med psykoselidelser*. <https://www.helsedirektoratet.no/retningslinjer/psykoselidelser>
- Helsedirektoratet. (2016, 10. juni). *Helsedirektoratets rolle vedrørende ulike tester og verktøy innen psykisk helse og rus*. Brev med referanse 15/1915 - 5.
- International Test Commission. (2001). International Guidelines for Test Use [ITC retningslinjer for testbruk]. *International Journal of Testing*, 1(2), 93–114. https://doi.org/10.1207/S15327574IJT0102_1
- <https://www.intestcom.org/page/17>
- International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests (Second edition)*. <https://www.intestcom.org/page/16>
- Kaufman, J., Birmaher, B., Brent, D., Rao, U. & Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial Reliability and Validity Data. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36(7), 980–988. <https://doi.org/10.1097/00004583-199707000-00021>
- Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B. & Walters, E. E. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine*, 35, 245–256. <https://doi.org/10.1017/S0033291704002892>
- Kooij, J. J. S. & Francken, M. H. (2010). *Diagnostisk intervju for utredning av ADHD hos voksne*. DIVA Stiftelsen.
- Kornør, H., Jozefiak, T., Hanssen-Bauer, K. & Hysing, M. (2015). Forsvarlig testbruk. *Tidsskrift for Norsk psykologforening*, 52(7), 600–601.
- McCrae, R. R. & Costa, P. T. (2017). *NEO Personality Inventory-3, norsk manual*. Hogrefe Psykologiförlaget AB.
- Meyers, J. E. & Meyers, K. R. (2004). *Rey Complex Figure Test and Recognition Trial, norsk versjon*. Hogrefe Psykologiförlaget AB.

- Montgomery, S. A. & Åsberg, M. (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry*, 134, 382–389. <https://doi.org/10.1192/bjp.134.4.382>
- Nasjonalt kunnskapssenter om vold og traumatisk stress (NKVTS). (2015a). *KATE-Barn: Kartlegging av traumatiske erfaringer. Barneverjon.*
- Nasjonalt kunnskapssenter om vold og traumatisk stress (NKVTS). (2015b). *KATE-Foreldre: Kartlegging av traumatiske erfaringer. Foreldreversjonen.*
- Raven, J. C. (1998). *Raven's Matriser*. Pearson Assessment.
- Roth, R. M., Isquith, P. K. & Gioia, G. A. (2005). *BRIEF-A: Behavior Rating Inventory of Executive Function-Adult version*. Psychological Assessment Resources.
- Rutter, M., Le Couteur, A. & Lord, C. (2003). *Autism diagnostic interview-revised*. Western Psychological Services.
- Ryder, T. M., under utgivelse. Testkvalitetsprosjekt – del 2: Testene i behov av kvalitetstiltak. *Tidsskrift for Norsk psykologforening*
- Sachser, C., Berliner, L., Holt, T., Jensen, T. K., Jungbluth, N., Risch, E., Rosner, R. & Goldbeck, L. (2017). International development and psychometric properties of the Child and Adolescent Trauma Screen (CATS). *Journal of affective disorders*, 210, 189–195. <https://doi.org/10.1016/j.jad.2016.12.040>
- Selvik, A. (2007). Forsvarlig bruk av tester, strukturerte intervjuer og spørreskjemaer – kvalitet og bruksrett i helsevesenet. Internrapport Psykologforeningen. https://ressursside.no/handbok-filer/forsvarlig_bruk_av_tester_npf.pdf
- Sheehan, D. V., Lecrubier, Y., Harnett-Sheehan, K., Amorim, P., Janavs, J., Weiller, E., Hergueta T., Baker, R., Dunbar, G. (1998). The Mini International Neuropsychiatric Interview (M.I.N.I.): The Development and Validation of a Structured Diagnostic Psychiatric Interview. *Journal of Clinical Psychiatry*, 59(20), 22–33. <https://doi.org/10.1037/t18597-000>
- Sparrow, S. S., Cicchetti, D. V. & Balla, D. A. (2011). *Vineland Adaptive Behavior Scales- Second edition, norsk versjon*. Pearson Assessment.
- Strobel, C. & Engedal, K. (2008). *MMSE-NR. Norsk revidert Mini Mental Status Evaluering. Revidert og utvidet manual.*
- Vaskinn, A. & Egeland, J. (2012). Testbrukerundersøkelsen: En oversikt over tester brukt av norske psykologer. *Tidsskrift for Norsk psykologforening*, 49(7), 658–665.
- Vaskinn, A., Egeland, J., Nielsen, G. H. & Høstmælingen, A. (2010). Norwegian psychologists' use of and opinions on tests and testing. *Tidsskrift for Norsk psykologforening* 47(11), 1010–1016.
- Wechsler, D. (2008). *Wechsler Memory Scale-Third edition, norsk versjon*. Pearson Assessment.
- Wechsler, D. (2009). *Wechsler Intelligence Scale for Children-Fourth edition, norsk versjon*. Pearson Assessment. <https://doi.org/10.1037/t15169-000>
- Wechsler, D. (2011). *Wechsler Adult Intelligence Scale-Fourth edition, norsk versjon*. Pearson Assessment.
- Wechsler, D. (2015). *Wechsler Preschool and Primary Scale of Intelligence-Fourth edition, norsk versjon*. Pearson Assessment.
- Wechsler, D. (2017). *Wechsler Intelligence Scale for Children-Fifth edition, norsk versjon*. Pearson Assessment.
- Weiss, D. S. & Marmar, C. R. (1997). The Impact of Event Scale-Revised. I J. P. Wilson & T. <https://doi.org/10.1037/t12199-000>
- M. Keane (red.), *Assessing psychological trauma and PTSD* (s. 99–411). Guilford Press.

