

Kan vi leve med manglende norsk normering på WISC, WAIS og WPPSI?

Tor Herman Andreassen

Øyvind Fallmyr
oeyfal@piv.no

«Kvalitetskontroll av personlighetstester» het en reportasje i Tidsskriftet (11/05). Den handlet om hvordan Det Norske Veritas og Sertifiseringsrådet for testbruk i Norge kan tilby sertifisering av tester i arbeidslivet. Kravene til validitet, reliabilitet og normering skal være strenge, og det skal bli vanskeligere for de useriøse aktørene å selge sine tester.

I Norge mangler vi i dag normerte evnetester for barn og voksne, og de som brukes, vil neppe klare sertifiseringen på grunn av manglende lokal normering. Dette innebærer at de mest sårbare barna og ungdommene innen pedagogisk psykologisk tjeneste, psykisk helsevern for barn og unge, habiliteringstjenesten og de statlige kompetansesentrene ikke tilbys god nok utredning, fordi testverktøyet ikke fungerer. Situasjonen for voksne tør jeg ikke engang skrive om.

For noen uker siden kjøpe jeg min første badevekt. I butikken hadde de seks ulike modeller med priser fra 199 til rundt tusen kroner. Som alltid når jeg gjør viktige innkjøp, prøver jeg varene. Jeg oppdaget at ingen viste det samme, og at det skilte nesten syv kilo mellom snilleste og strengeste vekt. Hvordan velger man i en slik situasjon? Og hvordan tolker jeg tallet som dukker opp på vekta?

I jobben som skolepsykolog bruker jeg evnetestene WISC-III og WPPSI-II, og er usikker på hvordan målingene skal tolkes. Jeg har ingen garanti for at de ulike IQ-verdiene og de oppgitte konfidensintervallene stemmer med virkeligheten.

Hvorfor trenger vi evnetester?

Evnetester er avgjørende for å kunne identifisere og bekrefte en rekke tilstander og vansker som barn og unge strever med.

- 1. Dette gjelder selvsagt psykisk utviklingshemning og generelle lærevansker, hvor evnenivået er et hovedkriterium som vurderes.
- 2. For en rekke lærevansker er det nødvendig med en evnevurdering for å stadfeste om det dreier seg om spesifikke vansker og ikke generelle vansker. Dette gjelder blant annet dysleksi, spesifikk artikulasjonsforstyrrelse, ekspressiv språkforstyrrelse, impressiv språkforstyrrelse, spesifikk staveforstyrrelse, dyskalkuli og ikke-verbale lærevansker.
- 3. Evnetester er også nødvendige for å kunne stille diagnoser som ADHD og Asperger syndrom.
- 4. Evnetester brukes nesten alltid i nevropsykologiske undersøkelser, blant annet som et referansemål for å tolke testresultater på læring, oppmerksomhet, sansepersepsjon og motorikk.
- 5. Evnetestene hører dessuten vanligvis med i utredning av tilstander som autisme, cerebral parese, føtalt alkoholsyndrom og hodeskader.
- 6. I PP-tjenesten hjelper testene oss å finne riktig ambisjonsnivå i den tilpassede undervisningen, testene kan si noe om barnets ressurser og hva som er deres beste læringsstil.

Hva er galt med WISC-III?

Bortsett fra at WISC-IV sannsynligvis er bedre, er WISC-III helt utmerket. Det alvorlige er at den ikke er normert i Norge. Forskning på WISC-III gjort i Vestfold (Sundberg, Egeland, Andreassen & Stensli, 2006), indikerer at den svenske normeringen sannsynligvis er for streng, spesielt for ungdomsskolegruppen. Videre synes det som indeksene som skal avdekke modalitetsspesifikke vansker (visuospatiale eller språklige vansker), ikke har god nok validitet. Vi kan med andre ord neppe tolke WISC-III normativt, dvs. sammenligne et barns resultater med andre barn på samme alder, og heller ikke tolke profilen ipsativt ved å tolke forskjeller innad i enkeltprofiler som indikasjoner på spesifikke vansker. Hva står vi da igjen med? Skal vi gå tilbake til WISC-R, som har foreldede normer? Skal vi bruke den som et observasjonskartleggingsverktøy som kan si noe om relasjon til tester, oppgaveløsningsatferd og arbeidsstil?

Hvilke konsekvenser har evnetester som ikke fungerer?

Standardiserte tester ble oppfunnet fordi utviklingsbedømming og vurderinger av psykologiske funksjoner har mange feilkilder, blant annet ekspertenes skjønn. Standardiseringen skal ha en kalibrerende effekt på det nødvendige skjønn. Selv om evnetesting kun er en av mange informasjonskilder i utredning, diagnostikk og tiltaksplanlegging, vil konsekvensene av dette uføret lett bli alvorlige for klientene våre. I et intervju i Tidsskriftet (Strand, 2005) bagatelliserer både Assessios Dag Øyvind Engen Nilsen og psykolog Astri J. Lundervoll problemene ved å vise til at alle tester har feilmarginer. Andreassens (2006) historier fra sin praksis, i dette nummeret av Tidsskriftet, er skremmende eksempler på hvor alvorlig situasjonen er. Når feilmarginene blir så store og bruken av skjønn blir tilsvarende dominerende, må man kanskje spørre om godt skjønn må koste over 13 000 kroner, som WISC-III koster. Følgende konsekvenser er sannsynlige:

- 1. Uriktige funksjonsbeskrivelser og verdiløse tiltaksforslag.
- 2. Screeningfunksjon og evne til å generere hypoteser blir svekket.
- 3. Testens evne til å gi grunnlag for diagnostikk av en rekke vansker blir uforsvarlig usikre.
- 4. Testens posisjon som selve standarden andre tester vurderes mot, er ødelagt. Med en godt normert evnetest kan man lettere supplere med amerikansknormerte tester (K-ABC, Leiter, TONI), fordi vi kan sammenligne med standarden.
- 5. Mennesker som har juridiske rettigheter til pedagogisk, psykososial, medisinsk og økonomisk hjelp, fanges ikke opp.
- 6. Mennesker som ikke har en psykisk utviklingshemning, risikerer å få diagnosen.

Hvem har ansvaret?

Assessio Norge AS har hatt rettighetene og solgt testene, og har ikke tatt et ansvar for å rydde opp. Assessio har lagt to premisser til grunn for å si at WISC-III er gangbart i Norge: den svenske normeringen er god nok, og norske barn skiller seg ikke fra de svenske. Når så mange som halvparten av de spurte svenske foreldrene takket nei til å delta, kan det vel være en risiko for at de ressurssterke er overrepresentert, og at det bidrar til at testen antagelig er for strengt normert? At norske og svenske barn presterer likt, er også diskutabelt, sett i lys av de store skoleundersøkelsene som er gjennomført. TIMSS (Trends in International Mathematics and Science Study) -undersøkelsen fra 2003 viste for eksempel at blant norske 4.-klassinger var det 30 % som klarte regnestykket 15 x 9, blant svenske, danske og finske barn var det mellom 60 og 70 % som klarte det. PISA-undersøkelsen (Programme for International Student Assessment) fra 2003 konkluderer med at norske elever presterer betydelig

svakere enn de andre nordiske landene. TIMSS-undersøkelsen konkluderer med at trening på elementære ferdigheter vektlegges lite i norsk skole, og at norske elever skårer påfallende svakt i emnet problemløsning som særlig måler elevenes evne til analytisk resonnering.

Den enkelte psykolog står ansvarlig for kvalitet og konsekvenser av sitt arbeid. Vi som bruker testene, har selv måttet finne ut at resultatene er feilaktige, og selv måttet rydde opp og nyansere gale slutninger. Noen har erstattet WISC-III med gamle WISC-R, som er for snill. Andre har begynt å innkalle på ny for å reteste og endre diagnoser.

Assessio burde gått ut med advarsel om sannsynlige feil på produktet. Det kan synes som det er Assessios politikk å ikke informere om feil, selv der disse er åpenbare. Det er godt over ett år siden jeg gjorde dem oppmerksomme på at i deltesten «Ordforståelse» er det feilaktig oppgitt stopp etter fire gale svar på registreringsskjemaet (det riktige skal være fem). De har verken endret feilen eller informert om den, og selger fremdeles skjemaene.

Psykologforeningen henvendte seg for et par år siden til Helse- og omsorgsdepartementet, Helsedirektoratet og Justisdepartementet og meldte sin bekymring for kvaliteten på evnetestene som brukes i Norge. Jeg ønsker meg en tilbakemelding fra Testpolitisk utvalg, og eventuelt Fagetisk råd om hvilke retningslinjer som bør gjelde i denne unntakstilstanden. Kan vi fortsette å bruke testen? Under hvilke betingelser? Er det kun totalskåren som kan benyttes for barna på barnetrinnet, forutsatt at deres evnenivå ligger over 75–80? Kanskje vi bør slutte å bruke WISC-III, for å ivareta våre fagetiske retningslinjer.

Hva blir veien videre?

I stedet for å bruke mer tid og penger på WISC-3 og WPPSI-R bør vi forvente å få oversatt og normert WISC-4 og WPPSI-3, som lenge har vært i bruk andre steder i verden.

Nasjonen som har råd til ny hoppbakke i Oslo til over 300 millioner kroner, må ha råd til testverktøy som sikrer at de mest sårbare blant oss får gode nok utredninger. Det bør være et statlig ansvar å sørge for at standardiserte evnetester fungerer.

Til slutt et tankeeksperiment: Hva ville skjedd om legenes blodtrykkmålere, røntgenapparater, CT- og MR-maskiner hadde vært feilkalibrerte og gitt gal informasjon, kanskje var de for strenge (falskt positive) eller for snille (falskt negative), men ingen visste. Hva ville skjedd om landets snekkere eller sveisere hadde fått tommestokker av strikk, vinkler som hadde ti grader slakk eller vatre uten gassbobler? Hva ville skjedd om vekta i kjøttdisken viste 40 % for mye (tilsvarende Andreassens eksempel hvor WISC 3 viser IQ 53 og WAIS 3 viser IQ 81 på samme elev)?

Konsekvensene ville kanskje blitt operasjoner der det ikke var grunnlag, mangel på behandling der det var påkrevd, hus ville kanskje klart seg, men kundene ville klaget. Er det noen som klager på oss? Vekta i kjøttdisken ville blitt tatt ut av bruk. En annen konsekvens ville selvsagt vært at myndighetene hadde bevilget penger, engasjert forskere, satt himmel og jord i bevegelse for å finne feilene, og rettet dem opp.

Og hvilken badevekt jeg kjøpte? Jeg gjorde som enkelte PP-kontorer som har gått over til WISC-R, jeg kjøpte selvsagt den snilleste vekta, den var også den billigste.

Øyvind Fallmyr
PPT, Tønsberg kommune
Pb 2410, 3104 Tønsberg

E-post oyvind.fallmyr@tonsberg.kommune.no