

Att välja rätt person till rätt arbetsuppgift kan ha en stor positiv effekt på både arbetsmiljö, produktivitet och effektivitet. Det visar en rad forskningsresultat. Ändå tillämpas inte den kunskapen inom arbetslivet. Motståndet mot standardiserade bedömningar vinner till förmån för intuitiva helhetsbedömningar. Det talade den amerikanska psykologen och forskaren **Scott Highhouse** om vid ett seminarium på Stockholms universitet som hölls i december 2013. Här är ett referat av seminariet.

Därför använder vi inte evidensbaserade urvalsmetoder i arbetslivet

Seminariet handlade om evidensbaserat beslutsfattande vid urval. Professor Scott Highhouse kombinerar psykologisk forskning om omdöme och beslutsfattande (judgement and decision making theory) med forskning om personbedömning och urval. Han försöker besvara frågan om varför vi inte tillämpar befintliga kun-

skaper för att göra bättre urval i arbetslivet och menar att om vi tillämpade evidensbaserade urvalsmetoder skulle vi uppnå ökad produktivitet och andra positiva effekter. Arbets- och organisationspsykologins största framgång enligt Scott Highhouse, är att den lyckats minska felmarginalen vid prediktion av arbetsprestation. Urval är en insats

som – om den görs rätt – kan ha en stor positiv effekt på arbetsmiljö, produktivitet och effektivitet i organisationer. För att vara ännu tydligare, bra urval har ofta större betydelse för organisationer och organisationers intressenter än olika former av utvecklingsinsatser. Vad är då rätt sätt att göra urval? Och varför tillämpas inte kunskapen?





” Bra urval har ofta större betydelse för organisationer och organisationers intressenter än olika former av utvecklingsinsatser.

Två skolor – kliniska eller mekaniska bedömningar

Inom arbets- och organisationspsykologin har under ett antal decennier funnits två skolor inom bedömning och urval. Den ena traditionen har kallats klinisk, holistisk eller individuell bedömning och den andra mekanisk, statistisk eller psykometrisk bedömning.

- **Kliniska bedömningar** kan definieras som bedömningar av olika personliga attribut som en expert anser vara av betydelse i ett visst sammanhang. Bedömningen är en intuitiv eller åtminstone inte en standardiserad sammanvägning av dessa olika aspekter.

- **Mekaniska bedömningar** kan definieras som standardiserade sätt att

mäta och sedan statistiskt sammanväga de aspekter som forskningen visat är av betydelse för urvalsbeslut. Dessa två skolor har funnits parallellt under lång tid.

Scott Highhouse beskriver de två bedömningstraditionerna i en artikel 2002: "Henry Murray ses som förgrundsgestalten inom den kliniska el-

” När man undersökt hur bedömningsexperter inom olika områden kommer fram till beslut har man funnit att de förlitar sig på begränsad information, är oklara över och inte kan redovisa hur de har kommit fram till sina bedömningar, att olika bedömare med samma information kommer fram till olika slutsatser samt att de blir mer säkra på sina ursprungliga slutsatser när irrelevant information tillförs.

ler holistiska bedömningstraditionen. Murray arbetade på den amerikanska byrån för strategisk service (the Office of Strategic Services, OSS) under andra världskriget. På OSS utvecklades ett program för urval av spioner. Här medverkade psykologer och dessa psykologer var influerade av personlighetsteori och psykoanalys. Inga psykologer från det arbetspsykologiska kunskapsfältet medverkade. Murrays holistiska approach byggde på idén om psykologens förmåga att observera mönster i en persons beteende och att integrera dessa observationer till en bedömning av prestation och effektivitet.”

I en OSS-rapport från 1948 ger Henry Murray uttryck för detta förhållningssätt på följande sätt: "And so it could be said that this method calls for the improvement of the psychologist as observer and interpreter rather than the improvement of mechanical instruments and test materials."

Det program man tog fram för urval av spioner innehöll en mängd olika övningar. Slutbedömningen gjordes med hjälp av diskussioner som skulle leda till konsensusbeslut. Men vid senare utvärdering av denna metod lyckades inte forskarna visa att dessa metoder hade tillfredsställande tillförlitlighet. Man fann också att en mer standardiserad del bestående av en dags testningar hade betydligt högre tillförlitlighet än den andra delen om tre dagar

av mer kliniska bedömningar. De kliniska bedömningarna tycks alltså ha sänkt tillförlitligheten (validiteten) i urvalsbeslutet. Trots dessa tillkortakommanden fortsatte kliniska helhetsbedömningar att vinna mark. Man ansåg att det var viktigt att se människan som en organisk helhet vars beteende är resultatet av både personlighet och omgivning. Förespråkarna för denna helhetsbedömning satte själva bedömaren som den centrala gestalten att göra denna helhetsbedömning och den psykologiska utbildningen som undervisade denna kliniska kunskap blev sättet att skaffa sig denna unika kompetens.

Arbetspsykologisk och psykometrisk tradition

Parallellt med denna tradition utvecklade den arbetspsykologiska, psykometriska traditionen metoder, samlade data och forskade. Förgrundsgestalten inom denna tradition anses vara *Gordon Allport*. Två stora program som han var inblandad i genomfördes i USA efter kriget, det ena på Standard Oil i New Jersey, det andra på Sears, Roebuck and Company.

Målet med dessa projekt var att göra valida prediktioner av arbetsprestation. Man använde ett stort antal mått såsom verbal förmåga, begåvning, chefsattityder, personlighet och liknande. De test som användes visade sig

ha hög validitet, det vill säga de lyckades mäta faktorer som sedan visade sig vara avgörande för arbetsprestation.

Sears, Roebuck and Company anlätade 1942 psykologen och psykometrikern *L L Thurstone* för att sätta ihop ett testbatteri för chefsurval: begåvnings-test, personlighetstest, värderingstest och preferenstest. Sambandet mellan testbatteriet och chefskap uppgick till .75 (där 1.00 är ett perfekt samband). Redan på 1940-talet kunde man alltså identifiera en generell faktor av ledaregenskaper.

På 1950-talet gjordes ett ambitiöst program där man kombinerade klinisk och psykometrisk metodik i individuell bedömning vid Psychological Research Services of Western Reserve University. Man besökte det företag där man skulle göra urval för att ta reda på företagets struktur och "personlighet" och satte sedan ihop ett batteri utifrån detta. Det kunde innehålla begåvnings-test och självskattningstest. En utvärdering konstaterade att den subjektiva kliniska slutbedömningen hade lägre validitet än de enskilda testen var för sig.

I England tillsatte man the War Office Selection Board (WOSB). De flesta bedömare hörde till den psykiatriska traditionen snarare än den arbetspsykologiska och psykometriska traditionen. Man utvecklade metoder för bedömning och urval utan att fokusera på standardisering och reliabilitet.

I slutet av 1980-talet och början av 1990-talet "återupptäcktes" klinisk/holistisk personbedömning av forskare och arbetspsykologer. Psykologerna *A M Ryan* och *P R Sackett* (1987, 1989, 1992, 1998) skrev en rad artiklar i vilka de gav uttryck för sin förvåning över att individuell bedömning inte uppmärksammades av forskare och inte ens beskrevs av sina utövare trots att många psykologer tjänade sitt uppehälle på det.

Scott Highhouse hänvisar i slutet av sin artikel från 2002 till *A W Astins* (1961) diskussion om hur psykoterapi trots svagt stöd för dess effektivitet kunde blomstra under 1950-talet. *A W Astin* kommer i denna analys fram till att området uppnått vad han kallar "funktionell autonomi". Detsamma menar Scott Highhouse har inträffat för området individuell eller klinisk bedömning. Utövare utan kunskaper om

bedömning, psykometri och statistik strömmar till detta verksamhetsområde och – märkligast av allt – de som verkar inom området *avkrävs ingen redovisning av det empiriska stödet för sina insatser*.

Fallgropar i omdöme vid urvalsbedömning

Ett av psykologins stora bidrag till samhället är forskningen och kunskapen om människans omdöme och beslutsprocesser. Genom detta bidrag har vi lärt oss mer om hur vårt omdöme påverkas av ovidkommande faktorer, hur svårt vi har att uppskatta sannolikhet att förutsäga saker som kommer att hända och hur vi föredrar och låter oss påverkas mer av berättelser än av enstaka fakta även när det är fakta som är relevanta för bedömningen. Denna psykologiska kunskap förklarar varför tillförlitligheten sjunker när vi försöker "tolka" testresultat och annan data som insamlats.

Forskare har tittat specifikt på urvalsexperters beslutsprocesser. När man undersökt hur bedömningsexperten inom olika områden kommer fram till beslut har man funnit att de förlitar sig på begränsad information, är oklara över och inte kan redovisa hur de har kommit fram till sina bedömningar, att olika bedömare med samma information kommer fram till olika slutsatser samt att de blir mer säkra på sina ursprungliga slutsatser när irrelevant information tillförs (R Hastie & R M Dawes, 2001). Det finns alltså inget stöd i forskningen för att den professionella bedömaren har en unik förmåga att väga samman intryck, ställa relevanta frågor och till sist med hjälp av sitt omdöme fatta bättre urvalsbeslut. Snarare visar forskningen att professionella bedömare överskattar sin förmåga.

En annan fallgrop vid urvalsbedömning är att vi lätt tror att det går att göra helt säkra bedömningar och prediktioner (förutsägelser). Vi inser inte att urval är en fråga om sannolikhet och att vi kan förbättra urvalet på gruppnivå men att varje enskilt urval inte med hundraprocentig säkerhet blir rätt. Att använda standardiserade datainsamlingsmetoder och standardiserade metoder för att väga samman data är ett sätt att öka sannolikheten för rätt urvalsbedömningar, men vi får leva

med att ett visst antal urvalsbeslut per hundra urval blir fel.

Ett annat felaktigt antagande är att erfarenhet gör oss bättre på att predicera prestationer på lång sikt. Även detta antagande saknar stöd i forskningen (Camerer & Johnson, 1991).

I boken *Thinking, fast and slow* (2011) skriver *Daniel Kahneman* om bedömningsexperten i kapitlet "The illusion of validity". Kahneman menar att vi kan göra "intuitiva" prediktioner endast i miljöer där vi i princip får omedelbar återkoppling, till exempel inom vård eller vid olyckor. Detta är inte fallet i urvalssammanhang. Vi får inte omedelbar feedback på hur en person presterar över tid. Vi får oftast inte feedback alls på om och hur mycket vår bedömning förbättrar urvalet jämfört med om vi hade singlat slant. Sådan återkoppling får vi bara om vi gör systematiska utvärderingar.

Daniel Kahneman skriver också om "an illusion of skill", det vill säga en idé om expertkunnande som saknar empirisk grund. Han beskriver sitt eget arbete som urvalspsykolog inom den israeliska armén. Djupt inom sig anade han att han inte kunde bedöma vilka soldater som skulle komma att bli bättre

Evidensbaserat beslutsfattande vid urval

Under de 70 år som följt efter de tidigare nämnda urvalsprojekten i USA har man varje gång man jämfört kliniska/holistiska och mekaniska/psykometriska bedömningar kommit fram till samma resultat. Forskningen ger tydligt besked, mekanisk sammanvägning vinner överlägset när samma information bedöms (P E Meehl, 1954, W M Grove, D H Zald, B S Lebow, B E Snitz & C Nelson, 2000, N R Kuncel, D M Klieger, B S Connelly & D S Ones, 2013).

Holistiska bedömningar sänker validiteten jämfört med om man bara använt enskilda testresultat eller – ännu bättre – vägt ihop dem på ett standardiserat sätt. Enligt N R Kuncel med flera (2013) sänker klinisk sammanvägning av bedömningsdata träffsäkerheten med 25 procent. I dessa 25 procent förlorad säkerhet i bedömningen ligger stora ekonomiska och mänskliga värden.

Forskningen inom området är alltså solklar och ger oss besked om hur urvalsprocessen ska utföras för största möjliga träffsäkerhet. När det gäller datainsamlingen bör den ske på ett strukturerat och standardiserat sätt, det vill säga samma slags information ska samlas in för varje enskild kandidat som söker jobbet. Forskningen visar också att även kombinationen av de olika delarna information ska ske på ett standardiserat sätt. Varje informationskälla som används ska ha ett empiriskt bevisat samband med det bedömaren vill

” Det finns alltså inget stöd i forskningen för att den professionella bedömaren har en unik förmåga att väga samman intryck, ställa relevanta frågor och till sist med hjälp av sitt omdöme fatta bättre urvalsbeslut. Snarare visar forskningen att professionella bedömare överskattar sin förmåga.

re än andra och att det varje år visade sig att han heller inte hade kunnat det, men att han trots detta började om med samma urvalsprocedurer år efter år. Kahneman visar att graden av tilltro till den egna bedömningen inte har något samband med riktigheten i den egna bedömningen (vare sig man är säker eller som i hans fall tvivlar).

turet och standardiserat sätt, det vill säga samma slags information ska samlas in för varje enskild kandidat som söker jobbet. Forskningen visar också att även kombinationen av de olika delarna information ska ske på ett standardiserat sätt. Varje informationskälla som används ska ha ett empiriskt bevisat samband med det bedömaren vill



förutsäga. Om det inte finns ett konstaterat samband mellan det man bedömer (prediktorn) och det man vill förutsäga (till exempel arbetsprestation) bör det inte ingå i datainsamlingen.

Mer information är alltså inte automatiskt bättre information, tvärtom pekar forskningen på att "extrainformation" ges en tyngd i beslutsprocessen som den inte alls motsvarar. När man samlat in all information som bevisats ha ett samband med det man vill förutsäga ska den framtagna informationen vägas samman enligt en på förhand framtagna statistisk formel.

förvanskar något som är mycket komplext. Mekaniska bedömningar baserade på statistik förenklar för bedömare i själva urvalsögonblicket, men de formler som ligger till grund för prediktion rymmer enorma mängder data och information om relationen mellan olika data och allt detta motsvarar med råge den komplexitet som en så kallad helhetsbedömning omfattar. Faktum kvarstår, samhället skulle vinna mycket på ett mer evidensbaserat beslutsfattande vid urval. Då skulle diskrimineringen minska och effektiviteten öka i hela arbetslivet.

” Vi inser inte att urval är en fråga om sannolikhet och att vi kan förbättra urvalet på gruppnivå men att varje enskilt urval inte med hundra procentig säkerhet blir rätt.

Enligt vår mening bör leverantören av urvalsprocessen göras ansvarsskyldig, med andra ord kunna redogöra för hur och med vilket stöd han eller hon har kommit fram till sin slutsats. Det bör gå att spåra processen fram till beslut och om flera bedömare använder samma metod på samma data ska de komma till samma slutsats det vill säga vara reliabla.

Motståndet styr mer än förnuftet

Varför har den intuitiva helhetsbedömningen så stark ställning när all forskning visar att vi alla skulle tjäna på en mer standardiserad bedömning? När det gäller strukturerade datainsamlingsmetoder kan det bero på att människor motsätter sig begränsningar av vilken information de ska samla in. När det gäller kombinationen av data, alltså att strukturera beslutsprocessen, finns det om möjligt ett ännu större motstånd baserat på uppfattningen att detta förenklar och därmed

REFERENSER

Astin A W (1961). The functional autonomy of psychotherapy. *American Psychologist*, 16, 75–78.

Camerer C F, Johnson E J (1991). The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In Ericsson K A, Smith J (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 195–217). Cambridge, Cambridge University Press.

Grove, W M, Zald, D H, Lebow, B S, Snitz, B E, & Nelson, C (2000). Clinical versus mechanical prediction. *Psychological Assessment*, 12, 19–30.

Hastie, R, & Dawes, R M (2001). *Rational choice in an uncertain world*. Thousand Oaks, CA: Sage.

Highhouse, S. (2002). Assessing the candidate as a whole: A historical and critical analysis of individual psychological assessment for personnel decision making. *Personnel Psychology*, 55, 363–396.

Kahneman, D (2011). *Thinking fast and slow*. New York: Farrar, Strauss & Giroux.

Kuncel, N R, Klieger, D M, Connelly, B S, & Ones, D S (2013, September 16). Mechanical Versus Clinical Data Combination in Selection and Admissions Decisions: A Meta-Analysis. *Journal of Applied Psychology*. Advance online publication, doi: 10.1037/a0034156.

Meehl P E (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.

Murray, H (1948) *Assessment of men: Selection of personnel for the Office of Strategic Services*. New York, Rinehart.

Ryan A M, Sackett P R (1987). A survey of individual assessment practices by I-O psychologists. *Personnel Psychology*, 40, 455–488.

Ryan A M, Sackett P R (1989). Exploratory study of individual assessment practices: Interrater reliability and judgments of assessor effectiveness. *Journal of Applied Psychology*, 74, 568–579.

Ryan A M, Sackett P R (1992). Relationships between graduate training, professional affiliation, and individual psychological assessment practices for personnel decisions. *Personnel Psychology*, 45, 363–385.

Ryan A M, Sackett P R (1998). Individual assessment: The research base. In Jeanneret R, Silzer R (Eds.), *Individual psychological assessment: Predicting behavior in organizational settings* (pp. 54–87). San Francisco: Jossey-Bass.



MARIA ÅKERLUND
leg psykolog



SARA HENRYSSON EIDVALL
leg psykolog, specialist i
arbetslivets psykologi



ANDERS SJÖBERG
docent i psykologi