

PERSONLIGHED OG JOBSUCCES

Troels Gottlieb¹ & Jeppe Brændskov Klewe²

Denne artikel behandler spørgsmålet: Kan personlighedstest anvendes til at forudse jobsucces? Til dette formål skitserer vi indledningsvis den historiske person-situation-debat, som efterlod en fundamental skepsis vedrørende personlighedens eksistens. Herefter redegør vi for, hvordan personlighedsbegrebet overlevede denne kritik og bestod som relevant forskningsgenstand. Dette dels på baggrund af en begrebsmæssig konsensus via femfaktormodellen, dels på baggrund af en udvikling af den forskningsmæssige metodik bestående af metaanalytiske og korrektive validitetsgeneraliseringsmetoder. Til undersøgelse af forholdet mellem personlighed og jobsucces resumerer vi efterfølgende de to første metaanalyser over genstanden fra begyndelsen af 1990'erne. Perspektivering foretages med afsæt i Robert Hogans perspektiver i forhold til at indsamle, tolke og anvende psykometriske data i en erhvervsammenhæng. På baggrund af metastudierne konkluderes det, at 1) sammenhængen mellem personlighed og jobsucces er i bedste fald moderat, 2) den "succesfulde medarbejder" varierer fra job til job, 3) ønsker man at opnå stærkere korrelationer, skal erhvervsrettede personlighedstest baseres på en grundigt udarbejdet jobprofil.

Introduktion

En voksende tendens inden for det erhvervspsykologiske felt er kravet om målbarhed. Datadrevet HR har længe været et must på direktiongangene i de større private virksomheder, men etableres nu også i kommunale, regionale og statslige institutioner. Hensigten er at muliggøre monitorering af ikke blot produktivitet og omsætning, men også af komplekse fænomener som trivsel, kvalitet og ledelse via tal og grafer. Sidstnævnte er kerneområ-

-
- 1 Troels Gottlieb er cand.psych., ph.d.-studerende ved instituttet Mennesker og Teknologi, Roskilde Universitet, og har arbejdet som organisationspsykolog både i rollen som ekstern og intern konsulent, herunder med rekruttering.
 - 2 Jeppe Brændskov Klewe er cand.psych. og specialist i arbejds- og organisationspsykologi og har arbejdet med rekruttering i praksis, både som leder og som henholdsvis ekstern og intern konsulent. Arbejder til daglig som chefkonsulent på Bispebjerg og Frederiksberg Hospital.

der for erhvervspsykologien, men qua psykologiens dynamiske natur særlig udfordrende at reducere. Et af de områder, hvor det tilsyneladende er lykkedes erhvervspsykologien metodisk at kvantificere den menneskelige faktor, er ikke desto mindre inden for rekruttering³. Her har man i jagten på de bedste kandidater lige siden 1. Verdenskrig anvendt teorier og metoder med det formål at kunne foretage sammenlignelige målinger og kategoriseringer af intelligens og personlighed.

I denne artikel fokuserer vi på feltet, hvori personlighed kvantificeres og anvendes til at forudsige jobsucces. Sideløbende med at undersøge den empiriske evidens for relationen mellem personlighed og jobsucces beskriver vi metodologien bag denne forskningstradition.

Kritikken af personlighedsbegrebet

Guion og Gottier rejste allerede i 1965 alvorlig tvivl om tre grundlæggende forudsætninger for testning af personlighed, som har vist sig særdeles vanskelige for forskningen at modbevise: 1) Er de teoretiske konstruktioner, som ligger til grund for testene, gyldige? 2) Er kvaliteten og validiteten af de metoder/mål, som anvendes til at kvantificere personligheden, i orden? 3) Er styrken af relationen mellem personlighedstræk og de arbejdsrelaterede kriterier stærk nok til meningsfuldt og etisk forsvarligt at kunne anvendes til at udvælge medarbejdere? Med deres ambitiøst anlagte review af feltet forekom det pludselig særdeles tvivlsomt, at de anvendte personlighedstest reelt havde værdi i rekrutteringssammenhænge, for i blot 12 % af de inkluderede studier var den prædikative værdi marginalt bedre end 0, hvorved hele 88 % af studierne var ude af stand til at tilbagevise nulhypotesen: at der ingen sammenhæng er mellem personlighed og jobsucces. “[...] *Taken as a whole, there is no generalizable evidence that personality measures can be recommended as good practical tools for employee selection*” (s. 159), lød den dengang skelsættende konklusion.

Situation-person-diskussionen

Mischel (1968) udbyggede Guion og Gottiers problematisering af personlighedsvurdering. Dette med en mere generel kritik af en praksis funderet i den antagelse, at mennesker kendetegnes ved individuelle dispositioner med adfærdsmæssig forudsigelighed og kontinuitet igennem livet som sin konsekvens. Præmissen for Mischels kritik var, at individuelle forskelle i person-

3 Cranet-undersøgelsen afdækkede allerede i 2003, at over 50 % af de danske virksomheder anvendte test i deres rekrutteringer. Cranet-undersøgelsen gennemføres årligt af CBS og har som mål at afdække tendenser inden for HRM-feltet (Larsen, 2006).

lighed måtte være konsistente over tid og på tværs af kontekster, skulle det gængse personlighedsbegreb kunne retfærdiggøres. Og dét syntes hans review ikke at kunne bekræfte. For her konstateredes, at validitetskoefficienterne for personlighedsvurderinger sjældent overskred .30 og derved blot forklarede maksimalt op til ti procent af variansen i social adfærd. Herved udfordredes tidens trækpsykologiske og psykodynamiske modeller. Dette til fordel for et mere konstruktivistisk og situationelt paradigme, hvori personlighedstræk ikke var andet end stereotype semantiske kategorier – “[...] *quasi-explanatory ghosts and fictions*” (s. 95). Herved kunne Mischel forklare, hvorfor vi *tror*, der er noget, der hedder personlighed. Argumentet gik på, at menneskets kognitive systemer bygger på kontinuitet i en sådan grad, at vi selv skaber den – især der, hvor den slet ikke er at finde “i virkeligheden”. Dette fx når det gælder personer, både i forhold til oplevelsen af andre og i relation til os selv.

Vi *lader, som om* der er forudsigelighed, lød det ny og tankevækkende perspektiv altså. Og det slog an af flere årsager. Ud over væsentlige empiriske og teoretiske argumenter skal det konstruktivistiske paradigmes gennemslagskraft nok ses i et historisk lys. Det kom på banen samtidig med det kulturpolitiske opgør med den traditionelle psykologis menneskesyn, som oplevedes deterministisk og hæmmende. Uanset årsagerne betød det, at testpsykologien i en årrække var mistænkeliggjort og hensat til en skyggetilværelse (McAdams, 1997).

Et teoretisk gennembrud: Trækpsykologisk konsensus

Fremkomsten af femfaktormodellen (FFM) og metaanalytiske metoder har siden 1980'erne atter legitimeret en differentialpsykologisk personlighedspsykologi, og det er i dag i international sammenhæng et legitimt og veletableret synspunkt, at alle mennesker både besidder og adskiller sig fra hinanden ved forskellige grader af “emotionelle reaktioner⁴, ekstroversion, åbenhed, venlighed og samvittighedsfuldhed” (Nettle, 2011).

Især Costa og McCreas forskning har fra 1980'erne banet vejen for en slags trækpsykologisk renæssance.⁵ Den tyder nemlig på, at træk repræsenteret via psykometriske items er universelle (angiveligt biologiske frem for kulturelle) og følger individer over tid og på tværs af situationer (Digman,

4 Emotionelle reaktioner er den danske NEO PI-R oversættelse af, hvad der i andre sammenhænge kaldes neuroticisme. En høj score på neuroticisme i andre test svarer til en lav score på emotionelle reaktioner.

5 Allerede med Tupes og Cristals (1992) tekniske rapport for det amerikanske flyvevåben fra 1961 var den eksisterende forskning over forskellige personlighedsmodeller faktoranalyseret, og foreslået reduceret fra mange faktorer (her 35), til blot fem (*surgency, agreeableness, dependability, emotional stability og culture*).

1997; Costa & McCrae, 1996). Og blandt de væsentligste argumenter for FFM er, at forskellige observatører, som kender en given testperson godt (fx venner, naboer og kolleger), oplever vedkommende med signifikant ensartethed og med korrelationer mellem .30 og .65 for de fem NEO Personality Inventory-skalaer (McCrae & Costa, 1987). Noget tilsvarende gør sig gældende på facetniveauet, hvorved også kritikken om at være en stærkt reduktionistisk model har fået et modspil (McCrae & Costa, 1992)⁶. Og gode bekendtes vurderinger korrelerer her ikke blot med hinanden, men også med en testpersons egen NEO-PIR-besvarelse med .12-.22, ligesom ægtefællers besvarelser gør det med .23-.70⁷. Endvidere er førnævnte scorer påvist stabile over tid. NEO-PIR's test-retest foretaget efter 30-årsalderen og med minimum seks års udstrækning korrelerer i et *rank-order design* med mellem .78 og .81 (Terracciano, Costa & McCrae, 2006), og *peer ratings* (fx venner, naboer eller kolleger) med seks-syv års interval med medianscorer på omkring .7 (McCrae & Costa, 1992). Endnu stærkere ser det ud, når det gælder ægtefællers testbesvarelser af NEO-PI (McCrae, 2001). "Long term-reliabiliteten" ækvivalerer således mere eller mindre med NEO PI-R's test-retest (short-term/2 ugers) reliabilitet (Costa & McCrae, 1991).

Herudover er FFM-strukturen, repræsenteret ved den amerikansk udviklede NEO PI-R, transkulturelt genkendelig. Faktoranalysen af de 240 items i fem faktorer kan repliceres på tværs af kulturer (McCrae & Terracciano, 2005), og den genetiske vs. kulturelle effekt (enæggede tvillinger scorer mere ens end tveæggede) synes at være den samme, uanset om der er tale om en opvækst i Canada eller Tyskland (Jang et al., 1998). Sådanne fund har sammen med mere pragmatisk orienteret forskning rettet mod FFM's prædiktative validitet været med til at genoplivet feltet. At modellen også kan prædikere vigtige livsbegivenheder, såsom dødelighed, skilsmisse og arbejdslivsmæssig succes, på linje med både intelligens og SES (socioeconomic status) (Roberts, Kuncel, Shiner, Caspi & Goldberg, 2007), synes i international sammenhæng at have afblæst den ontologiske debat om personligheden.

Et metodisk gennembrud

Udviklingen af validitetsgeneralisering

Samtidig med at FFM blev etableret, pågik en metodisk udvikling i erhvervspsykologiens søgen efter lovmæssigheder. Det situationelle perspektiv – at

6 De fem skalaer består hver af aggregeringen af seks facetter, målt via otte items. Således kan NEO PI-R ikke blot tegne profiler og differentiere mennesker med afsæt i fem overordnede strukturer, men rummer reliable data på hele 30 aspekter.

7 NEO PI-R foreligger i to versioner: en selvrapport (Form S) og en observatørrating (Form R) med samme items fraseret i tredje person.

man ikke meningsfuldt kan finde bestemte træk, der fx på tværs af forskellige jobkategorier, performancekriterier og kontekster forudsiger ledelseseffektivitet – mødtes i 1980'erne af "modprojektet" *validitetsgeneralisering*. Dette i form af en integration af psykometri og differentialpsykologi, fundet i metaanalyser. Oprindeligt var afsættet validitetsstudier over kognitiv formåen (IQ) som prædikator for jobsucces, men siden da er også en række metaanalyser med fokus på personlighedens prædiktive værdi føjet til.

Sandsynlighed frem for signifikans

Pionererne i forhold til at fastholde (erhvervs-)psykologien i en induktiv søgen efter dét iboende, Schmidt og Hunter (1992, s. 1177), siger, at "*the goal in every science is explanation, and explanation is always causal*". Deres tilgang var kontroversiel, blandt andet fordi de gjorde op med den dominerende socialvidenskabelige analyseprocedure i form af nul-hypotese-signifikanstestning (Schmidt & Hunter, 1977). Via aggregeringen af data i metaanalyser påviste de, at signifikans-tankegangen har diskvalificeret en række primærstudier og disses konklusioner, ikke på grund af en faktisk manglende sammenhæng, men på grund af *sampling-error*/for små samples, hvorved forsøgspersonerne er tilfældigt ikke-repræsentative i forhold til populationen (Hunter & Schmidt, 2004). I stedet for som i den klassiske nul-hypotese-signifikanstestning at lade være afgørende, om der med 95 % sandsynlighed kan siges at være en ikke-tilfældig sammenhæng mellem fx intelligens og jobsucces, har Schmidt og Hunter flyttet fokus til *styrken* af en given korrelation. De argumenterer for, at signifikans-tankegangen skaber en u hensigtsmæssig binariet mellem effekt/ikke-effekt (som i øvrigt ikke anvendes i naturvidenskabelige fag som fx kemi og fysik (Schmidt & Hunter, 2001)). For i praksis er det næppe vigtigt, at et givent kriterium har en signifikant sammenhæng med intelligens, hvis korrelationen fx blot er .05. Er sammenhængen derimod .5, signifikans eller ej, er der potentielt penge at spare ved at bringe testen i anvendelse i rekrutteringsarbejdet (Landy, 2003). Hvor det i mange job meget vel kan være vanskeligt at "sætte pris på" jobsucces, eksemplificerer Smith og Smith (2005, s. 173) ikke desto mindre nedenfor, hvordan de statistiske koefficienter i nogle sammenhænge faktisk kan omregnes til kvalificerede estimater af noget så konkret som kroner og øre:

[...] a correlation of, say, .5 between predictor and criterion means that everyone standard deviation increase in test score will result in an improvement of .5 standard deviations in job performance. In high value jobs, the standard deviation of performance can be as much as 1.000.000 dollars, so the difference between two candidates who are separated by 3 IQ point can turn out to be 100.000 dollars per year ($3/15 \times 1.000.000 \times .5$).

Hvor langt fra alle job vil have implikationer af en størrelsesorden som i eksemplet her, kan man, som vi skal eksemplificere det senere via en case om lastbilchauffører, ikke desto mindre nemt forestille sig også lavtlønsjob, hvori der kan være anselige summer at spare ved at undgå fejlrekrutteringer.

Den “sande” vægtede korrelation

Validitetsgeneraliseringsbevægelsen førte til, at der i dag i *fortolkningen* af data ofte refereres til henholdsvis den *observerede* korrelation og den *sande estimerede* korrelation. Dette som konsekvens af Schmidt og Hunters pointe: at den observerede score typisk er så upræcis, at den bør korrigeres – først og fremmest ved at lade undersøgelserne vægte differentieret frem for at rapportere det rene gennemsnit. Et eksempel kunne være et metastudie over fem separate studier, der alle omhandler sammenhængen mellem højde og skostørrelse, hvoraf de fire bygger på et sample a 100 testpersoner, med et fund på .2, men det femte på et sample a 400 testpersoner med et fund på .6. Her bliver den observerede validitet/det rene gennemsnit på .24 $(.2+.2+.2+.6 / 5 = .24)$ umiddelbart til en upræcis repræsentation. Mere troværdigt synes det, at man med en korrigeret/vægtet gengivelse af data kommer tættere på den “sande” sammenhæng. Hvor der for 50 % af testpersonerne har været en effektstørrelse på .2 (de fire studier a hver 100 personer) og for 50 % har været en effektstørrelse af .6 (det ene studie a 400 personer), er den vægtede såkaldte “*true score*” .4 $(.2+.6 / 2 = .4)$ og derved sandsynligvis et bedre estimat end den såkaldte “*observed score*”.

Så langt er korrektions-tankegangen intuitiv. Men i bestræbelserne på at rapportere den “sande estimerede score” har Schmidt og Hunter (1977) tilføjet mere teknisk komplicerede metoder til at bearbejde data. Deres argumentation og metodik er fortsat til debat, men ikke desto mindre anvendes også de nedenfor refererede korrektioner i vidt omfang – bl.a. inden for personlighedsforskningen.

Sampling error og range restriction

Ifølge Hunter og Schmidt (2004) er *sampling-error* en generelt undervurderet faktor, som det er nødvendigt at korrigere for. For metaanalysernes store *sample-sizes* løser kun delvist problematikken med at give et retvisende estimat på en given sammenhængs styrke. Således er et væsentligt argument for at anvende korrektioner i fortolkningen af data i jobeffektivitetsforskningen fænomenet *range restriction*. For hvis hypotesen om, at fx intelligens er prædikator for ledelse, er korrekt, vil grundforskningen over denne sammenhæng uundgåeligt lide under, at de dårligst begavede individer fra den generelle population må antages underrepræsenterede: De vil sjældent hverken kunne eller ville opnå ansættelse som ledere. Men ikke nok med det, så forsvinder i empirien angiveligt også de mindst egnede blandt de faktisk

ansatte ledere. For i denne gruppe går det sandsynligvis mange sådan, at de hurtigt enten siger op på grund af utilstrækkelighedsfølelse eller simpelt hen afskediges på grund af manglende performance. Og her er der jo altså tale om profiler, som er højrelevante for virksomhederne – netop de personer, som man helst vil undgå at ansætte. Når dette potentielt sker parallelt med, at de bedst begavede hurtigt forfremmes eller headhunted af andre firmaer, kommer rådata/den observerede score således til at afspejle en mindre effektstørrelse, end hvis der ikke var range restriction (Smith & Smith, 2005). Og noget lignende gør sig sandsynligvis gældende for personlighedstest – implikationerne heraf vender vi tilbage til i diskussionen.

Målefejl og reliabilitet

Schmidt og Hunter-traditionens korrektioner af den *observerede* korrelation og etablering af den såkaldte *sande estimerede* korrelation foretages således med rødder i klassisk testteori. Ud over range restriction vægtes antagelsen om, at varierende resultater ved gentagne personligheds- eller intelligensstest er en konsekvens af *målefejl* – ikke af, at fx personlighed eller intelligens ændrer sig fra dag til dag. Et eksempel på dette kan være: Forsøgspersoner, der overser et komma i en matematisk prøve eller misforstår et item i en personlighedstest, fordi de er usædvanlig stressede eller trætte netop den dag, de tager testen; Forsøgslederen, som får byttet rundt på to forsøgspersoner og derved tastet rådata forkert i sit statistikprogram; Sekretær-vikaren, der fejlagtigt registrerer omsorgsdage som barnets første sygedag. I klassisk testteori anses sådanne og andre *tilfældigheder* som en præmis: en konstant støj, som psykometrien må forholde sig bedst muligt til (Schmidt & Hunter, 1999).

Når de bedste kognitive test har reliabilitetskoefficienter på omkring .95, betyder det, at henholdsvis 95 % af variansen (mellem forskellige kandidater) i en testscore skyldes variansen i netop det, som der testes for (her kognitive færdigheder), og 5 % af variansen mellem kandidaterne må tilskrives målefejl. I de etablerede FFM-personlighedstest er reliabiliteten for de respektive faktorer gennemsnitligt i spændet mellem .78 (venlighed) og .88 (emotionelle reaktioner) (Judge, Bono, Ilies & Gerhardt, 2002). Som tommelfingerregel kan det derfor siges, at reliabiliteten for et godt personlighedskonstrukt er minimum .7, og at maksimum 30 % af variationen mellem kandidaterne må kunne tilskrives målefejl (Smith & Smith, 2005).

Med hensyn til kriteriet – de forskellige KPI'er – er det sædvanligvis endda vanskeligere at opnå en høj reliabilitet, såvel at få et præcist billede af, hvad koefficienten for denne må forventes at være. Diskussionen om, hvordan personlighed måles, har taget fokus, hvorfor det er en relativt ny erkendelse, at også spørgsmålet om, hvad jobperformance egentlig er, er særdeles kompliceret at forholde sig til. Fx synes det med Hough (1992) at styrke validiteten mellem prædiktor og jobsucceskriterier, når også sidstnævnte ordnes som væsensforskellige kategorier. Parallelt med trækmodellens metodiske udgangspunkt, den leksikale pointe, at sprogets mange adjektiver kan

reduceres til relativt få træk, har Viswesvaran (2001) således operationaliseret 486 kriterier, som har været anvendt i litteraturen, til et domæne af blot ti væsensforskellige dimensioner. Det være sig: generel performance, produktivitet, jobkendskab, interpersonel kompetence, administrativ kompetence, kvalitet, kommunikation, ledelse og regelretted.

Uanset kriterium er de mest typiske data såkaldte peer ratings, som fx kan bestå af den evalueredes leders, kollegers, medarbejders eller kunders subjektive vurderinger af vedkommendes performance. Altså anvendes oftest subjektive vurderinger, og heller ikke her data med optimal reliabilitet. Og forskellige kriterier har forskellige reliabiliteter. Fx er reliabiliteten højere, når der spørges til kvalitet eller produktivitet, end når fokus er mere komplekse størrelser som kommunikation og interpersonel kompetence (Viswesvaran et al., 1996). Men som generelle estimater er et godt bud, at test-retest-korrelationer for supervisor-ratings er omkring de .80 (Salgado & Moscoso, 2003), og inter-rater-korrelationen i området af .5 (Viswesvaran et al., 1996) og .6 (Salgado & Moscoso, 1996).

Validitetsgeneralisering har således primært handlet om at imødegå *sampling-problemet*. Dette for det første ved via metaanalyse at sikre tusindvis af testpersoner og derved statistisk power. For det andet har man villet imødegå *reliabilitetsspørgsmålet*, og ovenfor nævnte hypotese, at selv når de bedste test anvendes, så underrapporterer rådata den faktiske sammenhæng mellem testscore og et givent kriterium. Det er på den baggrund, at Hunter og Schmidt (2004) har udviklet en formel til korrektion af data, med hvilken den observerede effektstørrelse divideres med kvadratroden af reliabiliteten for performancekriteriet. Murphy (2003) illustrerer metoden med, at hvis fx 100 studier af det strukturerede interview har en vægtet gennemsnitlig effektstørrelse på .45, og reliabiliteten for performancekriteriet (fx sygefravær) er .7, så er den estimerede sande score (det bedste estimat af validiteten af de strukturerede interview, som undersøges) sandsynligvis tættere på .54 (.45 divideret med kvadratroden af .7).

Hvor metoden har vundet stor opbakning, herunder at det i APA's Society for Industrial and Organizational Psychology (2003) "Principles for the Validation and Use of Personnel Selection Procedures" lyder, at "*adjustment of the validity coefficient for criterion unreliability should be made if an appropriate estimate of criterion reliability can be obtained*", så medfølger dog det paradoks, at måleusikkerheden med denne tilgang altid kommer effektstørrelsen til gode. Havde man således i eksemplet ovenfor haft endda bedre data, fx en reliabilitet på .9 (i stedet for .7), var den "sande effektstørrelse" blevet .54 (i stedet for .45). Herudover mener fx Murphy og DeShon (2000), at metodens baggrund i klassisk reliabilitetsteori forsimples problemstillingen og dette med u hensigtsmæssige estimater til følge. De påpeger, at der kan være mange andre årsager end tilfældige målefejl til, at reliabiliteten af jobperformance ratings er lav. *Job-appraisals* er underlagt mange andre motiver end at beskrive den evalueredes performance, såsom: Hvad vil moti-

vere, såre eller måske endda bidrage til at få afskediget personen, som bliver evalueret? Desuden er appraisals i nogen grad influeret af både den evaluerede og dem, som evalueres, med hensyn til disses køn, alder, etnicitet, værdier, relation til hinanden m.v. Murphy (2008) argumenterer derfor for, at det snarere handler om at udvikle dataindsamlingsprocessen på en måde, som medierer de komplekse forhold, der omgiver virksomheders anvendelse af ratings, end det er et simpelt spørgsmål om målefejl og en generel psykometrisk korrektion heraf. Fokus bør i stedet være på at fjerne organisatoriske og subjektive barrierer for *biased ratings* ved eksempelvis at motivere og træne dem, der rater, til at give ærlige og retfærdige vurderinger.

FFM og de metaanalytiske metoder har tilsammen udgjort nye muligheder for at gå på tværs af forskning, som førhen var præget af mangfoldige personlighedsmodeller. Førnævnte Guion og Gottiers (1965) review stod mere eller mindre som det eneste af sin art i næsten 30 år, hvad der er et godt billede på behovet for de nye metoder. Tidligere var det en nærmest uoverstigelig opgave at samle og analysere data fra et så omfattende og komplekst felt, som fx ledelsesforskningen er det. På bagkant af Guion og Gottiers pessimistiske konklusioner skulle ingen nyde noget af at ofre yderligere på en "tabt sag". Men fra 1991 så feltet altså anderledes tilgængeligt ud. For med FFM kunne den historiske forskning nu kategoriseres i fem træk, og kombineret med principperne fra validitetsgeneralisering var der nu tale om et langt bedre afgrænset felt, teoretisk som metodisk. Via metaanalytiske synteser af feltets primære studier og den heraf opnåede optimerede repræsentativitet trådte hidtil oversete signifikante sammenhænge samt stærkere og mere konsistente korrelationer end i fx Guion og Gottiers og Mischels klassiske reviews nu frem (Rothstein & Jolley, 2003).

Personlighed og jobsucces i et metaanalytisk perspektiv

De første studier, hvor Hunter og Schmidts metoder bragtes sammen med personlighedsbegrebet, var Barrick og Mounts (1991) metaanalyse *The Big Five Personality Dimensions and Job Performance* og Tett et al.s (1991) *Personality Measures as Predictors of Job Performance – a Metaanalytic Review*.

Især Barrick og Mount (1991) er fortsat flittigt citeret. Deres indledende litteratursøgning førte til, at hele 117 studier fra 1950'erne til 1980'erne inkluderedes, der herved dækkede imponerende 23.994 respondenter. Disse studier kategoriserede via tre typer performancekriterier (jobsucces, træningssucces og personaledata) og fem jobgrupper (professionelle, politi, managere, salg og skilled/semi-skilled), og de anvendte personlighedstræk blev klassificeret i FFM-dimensionerne (emotionelle reaktioner, ekstroversion, åbenhed, venlighed og samvittighedsfuldhed samt kategorien "diverse" i de tilfælde, hvor man i det primære studie ikke anvendte træk, der klart lod sig indordne i FFM). Herudover anvendte de Hunter og Schmidts (2004) fulde

validitetsgeneraliserings-procedurer. Dette begrundet med, at de korrigerede scorer tjente undersøgelsens teoretiske sigte og det overordnede forsknings-spørgsmål bedst, nemlig: *Hvad er den faktiske korrelation mellem personlighed og jobperformance?* Arbejdshypoteserne og fundene i forlængelse af disse var:

(1) *Samvittighedsfuldhed* (tendensen til at være planlæggende, vedholdende, ansvarlig, hårdarbejdende m.v.) og *emotionelle reaktioner* (tendensen til at være selvmedlidende, bekymret, temperamentsfuld, nervøs m.v.) forventedes valide prædiktorer for *alle* jobkategorier og succeskriterier (det vil *altid* være en arbejdslivsmæssig fordel at være planlæggende m.v., ligeså en ulempe at være selvmedlidende m.v.). Som forventet viste samvittighedsfuldhed sig valid på tværs af både jobgrupper og performancekriterier (.13 råscore/.22 korrigeret), men det gjorde emotionelle reaktioner til gengæld ikke (.05/.08).

(2) De øvrige personlighedsdimensioner (ekstroversion, venlighed og åbenhed) forventedes relaterede til jobperformance i *visse* jobtyper og i forbindelse med *visse* succeskriterier. Fx var hypotesen, at *ekstroversion* og *venlighed* ville vise sig valide prædiktorer for jobsucces i stillinger som salg og ledelse, hvori der er hyppig kontakt med andre mennesker, men ikke nødvendigvis i stillinger som ingeniør eller fabriksarbejder. Åbenhed – tendensen til at være nysgerrig, hurtigt opfattende m.v. – forventedes at prædikere jobtrænings-succes på tværs af jobgrupper, da sådanne personer forudsås motiverede for at deltage i træningsprogrammer og for at lære nyt. Deres hypoteser blev bekræftet delvist. Ekstroversion viste sig valid i forhold til salg (.09/.15) og ledelse (.11/.18), hvorimod den forventede sammenhæng med venlighed udeblev både for salg (.00/.00) og ledelse (.05/.10). Overraskende viste ekstroversion sig yderligere at have en sammenhæng med jobtræning (.15/.26), der herudover, som forventet, også korrelerede med åbenhed (.14/.25).

(3) Det sidste forskningsspørgsmål vedrørte ikke en konkret hypotese, men spørgsmålet om, hvorvidt validiteten af FFM-dimensionerne var ens for subjektive (fx performance ratings) som objektive (fx turnover) kriterier. Konklusionen lød, at personlighed prædikerer subjektive kriterier bedre end objektive.⁸

Samme år kom Tett et al.s (1991) studie *Personality Measures as Predictors of Jobperformance – a Metaanalytic Review* (n = 13.521). I tråd med Barrick og Mounts (1991) undersøgelse var sigtet både at bestemme personlighedstrækkenes prædiktive validitet af jobperformance, defineret ved hen-

8 Diskussionen om objektive vs. subjektive kriterier er klassisk og omhandler, hvorvidt performance bedst måles og forstås ud fra *organizational records* (fx direkte mål af produktivitet eller personaledata i form af sygefravær m.v.), som typisk betragtes som mere objektive data end organisationers forskellige spørgeskemabaserede evalueringer (fx appraisal- eller lederevalueringer, trivselsmålinger m.v.) (Viswesvaran, 2001).

holdsvis subjektive og objektive kriterier, og at undersøge, hvordan forskellige typer studier af sammenhængen mellem personlighed og jobsucces modererer den gennemsnitlige validitet. Dette ligeledes med grundlæggende afsæt i Schmidt og Hunters metoder – og så alligevel med væsentlige forskelle. Særlig iøjefaldende var eksklusionen af eksplorativt studier, som forskerne her fandt stærkt begrænset validitet sammenlignet med teori-båren forskningsstrategi. Om deres konfirmatoriske tilgang forklarede de:

Each study was classified as being either theoretically or empirically driven. Studies in which the researchers indicated either an explicit or implicit rationale for examining specific traits in relation to performance in a particular job were considered to be theoretically driven (i.e., confirmatory). Studies in which the researchers administered a broad array of personality measures (typically all scales of a multi-scale questionnaire) without any clear rationale for expecting significant correlations for some but not other traits were considered to be empirically driven (i.e., exploratory (s. 713).

Særlig kontroversielt ved Tett et al.s konfirmatoriske ideal var, at de som følge heraf kun anvendte absolutte scorere. Denne tilgang gjorde, at forskerholdet havde en forventning om at finde korrelationer for *alle* FFM's dimensioner samt finde endnu stærkere korrelationer, end Barrick og Mount havde gjort. Statistisk fifleri mente nogle, men de teoretiske argumenter var på plads: Fx synes det intuitivt, at ekstroversion kan hjælpe en sælger i kontakt med mange kunder, men omvendt stå i vejen for en data-programmørs koncentration gennem lange arbejdsdage alene foran skærmen. Derfor var det for Tett et al. ingen trussel mod trækteoriens kohærens, når et givent træk viste sig at korrelere positivt med jobsucces i én funktion, men negativt i en anden.

Begge hypoteser blev bekræftet i studiet. Konklusionen lød derfor, at personlighed prædikerer jobsucces, samt at der via en konfirmatorisk tilgang og anvendelse blot af absolutte værdier var grund til en vis optimisme på feltets vegne. Endda en større af slagsen end den, der tidligere på året var opstået med Barrick og Mount. Hvor Barrick og Mounts vægtede og korrigerede gennemsnitlige korrelation var på .11 (med et spænd fra .04 for træk relateret til åbenhed til .22 for træk relateret til samvittighedsfuldhed), var Tett et al.s fund .29⁹ (med et spænd på .16 for træk relateret til ekstroversion til .33 for venlighed). Her forekom venlighed – tendensen til at være medfølelse, godmodig og samarbejdende – altså at være den væsentligste faktor.

9 Efter metodekritik (Ones, Mount, Barrick & Hunter, 1994) foretog Tett, Jackson, Rothstein & Reddon (1994) en reanalyse af 1991-studiet, hvori den gennemsnitlige korrelation for hhv. konfirmatoriske og eksplorativt studier korrigeredes fra .29 og .12 til .24 og .04.

Kritikken, der ikke vil forsvinde

Hvor metaanalyserne således på ny sandsynliggjorde relevansen af en personlighedspsykologisk tilgang til forudsigtelse af organisatorisk adfærd, er kritikken dog aldrig forstummet (se fx Morgeson et al., 2007; Morgeson, Campion, Hollenbeck, Murphy & Schmitt, 2007). Der er trods alt tale om lave til moderate¹⁰ gennemsnitlige korrelationer, som sjældent overskrider .3. Og dette vel at mærke som korrigerede effektstørrelser, hvorved metoden, og dermed også effektstørrelserne, fortsat kan diskuteres. På den ene side kan der argumenteres for, at personlighed er en valid prædikator i forhold til jobsucces, og at gentagne metaanalytiske påvisninger af sådanne korrelationer må kaldes konsistente fund (Tett & Christiansen, 2007; Ones, Dilchert, Viswesvaran & Judge, 2007). Modsat kunne fokus også lægges på, at fundene, som de refererede undersøgelser eksemplificerer, er inkonsistente: Hverken effektstørrelser (.12 eller .24) eller konklusioner (samvittighedsfuldhed eller venlighed som det væsentligste træk) er identiske. Og dette, til trods for at der i kraft af den metaanalytiske metode er tale om studier, som dækker meget af den samme litteratur.

Også de direkte involverede forskerhold kritiserede i sin tid hinandens metoder og deraf følgende resultater (Ones, Mount, Barrick & Hunter, 1994; Tett, Jackson, Rothstein & Reddon, 1994). Siden har Barrick og Mount (2003) dog slået fast, at de to 1991-metaundersøgelser bør anses som komplementære: De umiddelbart forskelligartede fund var i højere grad en konsekvens af forskellige typer undersøgelser end udtryk for metodiske og teoretiske problemer ved den personlighedspsykologisk baserede organisationspsykologiske forskning. I et tilbageblik på det historiske validitets-generaliserings-ideals indtræden i AO-psykologien siger Barrick og Mount (s. 202-203):

Although there are numerous judgment calls that must be made in any metaanalysis, this is especially true in personality research given the countless number of personality traits that exists and the ambiguity regarding their labels and definitions. Similarly, judgment calls also are made on the criterion side, where little consensus exists regarding a taxonomy of performance dimensions. Because these decision rules obviously have a bearing on the outcome of the meta-analysis, subsequent researchers may believe that different conclusions would be reached if different decision rules were used.

10 Den statistiske konvention for "prosa-oversættelse" af korrelationer: < 0.3 = svag 0.3-0.5 = moderat > 0.5 = stærk (Cohen, 1970).

Perspektivering

Når psykologiske test i en erhvervssammenhængs umiddelbare loft på .3 har været kritiseret for blot at kunne forudsige op til 9 % af variansen mellem individer, skal det hertil nævnes, at det er en særlig psykologisk tradition at tage kvadratroden af korrelationen for at estimere en tests forklaringspotentiale. Således har det også været argumenteret, at en direkte oversættelse af træk-korrelationer er mere retvisende: At fx en .3-træk-korrelation forklarer helt op til 30 % af variansen (Ozer, 1985). Et argument, der dog samtidig diskvalificerer argumenterne for at korrigerer data.

Statistiske modeller er ikke facts; de er og bliver modeller, der repræsenterer en videnskabelig og relativ "realitet". Uanset uenighederne hvad dette angår skal det siges, at den dominerende socialvidenskabelige konvention lyder, at .3 trods alt er at betragte som en moderat korrelation (Cohen, 1970). Og set i lyset af, at psykologiske tests validitet er sammenlignelige med medicinske (Meyer et al., 2001), opstår spørgsmålet, hvorfor det tilsyneladende .3-loft egentlig skulle være diskvalificerende for anvendeligheden af personlighedstest. Fx nævner Meyer eksempler på publicerede studier, der blot har kunnet påvise (ikke-korrigerede) sammenhænge mellem: kemoterapi og overlevelse ifm. brystkræft på .03; rygning og lungecancer på .08; alkoholbrug i graviditeten og følgende præmatur fødsel på .09 og antihistaminer og nedsat nys og løbende næse på 0.11. Korrelationerne her er svage og så alligevel stærke nok til, at man inden for lægevidenskaben finder, at de kan danne grundlaget for den medicinske praksis.

Metaanalyser synes isoleret set at være velegnede til at afklare eventuelle signifikante sammenhænge. Men, må der holdes fast i, så mister man i aggregeringen af studier samtidig specifitet, og således kan metastudier som metodik heller ikke tilbyde definatorisk sande svar a la: Hvis A så B; hvis du modtager kemo, så overlever du; hvis du ansætter den mest dominante leder i ansøgerfeltet, så får vedkommende succes i jobbet osv. Med andre ord vil der i den metaanalytiske logik aldrig være korrelationer i nærheden af 1.0. Langtfra endda, når feltet er psykologi og genstanden dynamisk og situationelt påvirkelig i en grad, som fx adfærd er det. På denne baggrund synes stærke korrelationer at være et problematisk succeskriterie for personlighedspsykologisk tests anvendelighed.

Set i det lys bør rationalet for at tage test i anvendelse måske snarere baseres på, hvor meget en given test løfter sig fra 0.0, end at have 1.0 som sit vejledende ideal. Hermed ville udgangspunktet også blive, at enhver test, som bidrager til at distancere en vigtig beslutning fra chanceniveau, principielt må overvejes bragt i anvendelse. Med et sådant udgangspunkt bliver kravene til korrelationsniveauet (og altså til validitet), ikke et spørgsmål om svag, moderat eller stærk korrelation, men noget dynamisk, som besvares af den givne situation, vi som praktikere står i. Fx vil der formentlig være forskel på kravene til en test, der har til formål at selektere indkaldelsen af 100

ud af 200 ansøgere til jobsamtale forud ansættelsen som arbejdsmand på et større byggeri, som skal skydes i gang nu og her, end til en, der har til formål at udvælge én astronaut til en mission til Mars. I begge situationer vil undgåelse af procedurefejl og deraf følgende ulykkesrisiko være af central betydning for ansættelsen af den rette kandidat. Jf. FFM vil vi i begge situationer være særlig interesserede i at få udvalgt samvittighedsfulde personer. På byggepladsen vil vi dog bedre kunne leve med, at der slipper nogle "falsk positive" om bord – her vil en eventuel arbejdsulykke ikke nødvendigvis blive ekstremt dyr, invaliderende eller ligefrem dødbringende. Måske er værdien af at bringe en billig og simpelt administrerbar test her tilstrækkelig. Alternativet er at ringe til alle 200 ansøgere for at få et billede af, hvem de personlighedsmæssigt er, eller blot tage den karakteristisk, de i ansøgningen giver af sig selv, for gode varer, hvilket heller ikke er tilgange med hverken en prangende reliabilitet eller validitet. På rumfartøjet derimod lever vi dog hellere med "falsk negative" og forsinkede rekrutteringsprocesser. Her bliver en fejlrekruttering irreversibel: Vi risikerer, at de mange års træning vil være spildt, og at rumskibet slet ikke kommer hjem igen, skulle vi have bemandedet det med en skødesløs person. Her er idealet selvsagt mindst en moderat validitet, og vi vil i øvrigt give os rigelig tid til at følge grundigt op med samtlige (relativt få) kandidater, som på papiret både har den optimale uddannelse, fysik, personlighedsprofil m.m.

Feltets fortsatte udfordringer

Den amerikanske professor i psykolog Robert Hogan er én af få, som har viet sin karriere til at udvikle personlighedsteori specifikt møntet på anvendt organisationspsykologi. Det gør han på et teoretisk, empirisk og til tider også debatterende plan, hvori trækmodellen både ses som anvendelig og problematisk. Hans skepsis vedrører både muligheden for og anvendeligheden af en rendyrket *intra*-psykologisk model i form af at se træk som neuropsykologiske entiteter. Ikke desto mindre er han fortaler for anvendelse af personlighedsteori i en organisationspsykologisk sammenhæng og har som den første til at udvikle en erhvervsrettet FFM-baseret personlighedstest, the Hogan Personality Inventory, massiv kommerciel succes med sit forehavende. I en ledelsesteoretisk sammenhæng synes trækkene nemlig oplagte at forstå i et *inter*-psykologisk perspektiv og i en funktion af evaluerende termer: "[...] they reflect social consensus regarding an actor's behavior as it bears on the welfare of the group", siger han (Hogan, 1982). Således beskriver trækkene en persons renommé, ikke nødvendigvis vedkommendes identitet.

Hvor Hogan er med på, at der i mainstream-tilgangen til feltet potentielt overkorrigeres (Hogan, Hogan & Roberts, 1996), anser han ikke desto mindre korrektionerne som et nødvendigt onde; anbefalelsesværdige, alternativt

til at leve med underrapporteringer i form af de rå observerede scorer. Baggrunden herfor er, ud over de ovenfor refererede reliabilitets- og validitetsproblematikker, som karakteriserer feltet, en række ikke-statistiske problemer. Alle handler de populært sagt om metaanalysernes tendens til at få "blandet æbler og appelsiner sammen". Han peger på, at personlighedsskalaerne, der anvendes i primærstudierne, typisk er udviklet på baggrund af forskellige formål, bagvedliggende teorier og psykometriske konstruktioner, hvorfor der i metaanalyserne er en risiko for, at disse mis-kategoriseres. Dette, fx når trækket *skarpsindighed* fra én test (CPI) kategoriseres som FFM's *samvittighedsfuldhed* og ikke som *åbenhed*, der ifølge Hogan ville være det, personlighedsteoretisk set, korrekte at gøre.

En anden udfordring er at få klassificeret jobkategorierne hensigtsmæssigt. Hvor metaanalyserne kan siges at have sine statistiske fordele i forhold til at kunne imødekomme sample-size-problematikken, mister de til gengæld domænespecifik information, når de operationaliserer de enkelte job i mere overordnede jobkategorier, samtidig med at de opererer med generelle jobsucceskriterier frem for specifikke objektive som subjektive mål (Hogan, Davies & Hogan, 2007). Metaanalyser kan ligefrem tegne et misvisende billede, når idealet om metodisk set at forholde sig deskriptiv kammer over og bliver ateoretisk og derved ligefrem mekanisk. Hogan nævner således eksempler på, at fx militær-mekanikere i metaanalyser er endt i kategori med skolelærere og socialarbejdere. Rent objektivt kategoriseret er alle tre grupper "*professionals*". Men de tre job er ikke desto mindre også forskellige og mekanikerjobbet decideret væsensforskellig fra de to øvrige (Hogan, Hogan & Roberts, 1996). På denne måde må den metaanalytiske aggregering, og bestræbelserne på at nå både mere generaliserede og statistisk signifikante fund, medføre, at effektstørrelserne risikeres mindsket sammenlignet med primærstudiernes mere præcise afgrænsning af kontekst.

Sidst men ikke mindst står kriterieproblemet: Hvordan håndteres det, at man i nogle primærstudier anvender subjektive variable, som fx peer-ratings, mens andre bruger objektive kriterier, som fx salgstal eller sygefravær? Et eksempel kunne være, når der i aggregeringen af studier fx sammenblandes ét, som måler lærernes effektivitet via elevernes karaktergennemsnit, et andet via trivslen i klassen og et tredje via rektors rangordning af, hvem af lærerne hun personligt mener præsterer bedst, næstbedst, tredjebedst osv. Ifølge Hogan er det med sådanne problematikker ikke den psykometriske jobeffektivitets-forskning, der i sig selv er angribelig, men snarere selve den specifikke bagvedliggende forskningsmæssige praksis, som bør kritiseres (Hogan, DeSoto, Solano, 1977). Hogan (2005, s. 333) siger:

[...] researchers often fail to align predictors with criteria; this results in using measures of conscientiousness to predict service orientation, or measures of extraversion to predict training performance. The resulting correlations are low and critics then use them to indict personality re-

search rather than the personality researchers. And then many researchers ignore the problem of bidirectionality – sometimes measures of conscientiousness are positively correlated with outcomes and sometimes negatively, but for sound theoretical reasons in both cases. Imagine that conscientiousness is negatively correlated with one kind of performance (rated creativity), but positively correlated with another kind of performance (rated compliance with rules). This suggests that conscientiousness is a robust predictor of performance. However, if you simply add the two sets of correlations together, they cancel each other, and lead to the conclusion that personality is a weak predictor of performance.

Med de korrekte forbehold er Hogans konklusion vedrørende erhvervsrettet psykometri således, at dens berettigelse er veldokumenteret og langt stærkere, end hvad den rent akademiske tilgang synes at varsle, såvel at personlighedstest anvendt af kompetente personlighedspsykologer kan udgøre et væsentligt bidrag, både i rekrutterings- og udviklingssammenhænge.

Hvilke forventninger kan vi have til personlighedstest?

Hvor personlighedsbegrebets comeback i organisationspsykologien i høj grad skete på baggrund af Schmidt og Hunters (Hunter, 1986; Hunter & Schmidt, 1996; Schmidt & Hunter, 1998) arbejde med, via intelligensbegrebet, at påvise én generelt prædiktiv faktor for jobsucces, er det problematisk, at man tilsyneladende har arvet en lignende ambition for personlighed. Det virker, som om praksis styres af, at kunderne/virksomhederne betragter personlighed, på samme måde som man ofte betragter intelligens, hvilket fx de såkaldte integritetstest kunne være udtryk for. Disse er i bund og grund et empirisk funderet sammenkog af facetter fra FFM-trækkene emotionelle reaktioner, venlighed og pålidelighed (Salgado, Moscoso & Alonso, 2013). Men spørgsmålet er, om det faktisk giver mening at lede efter én “P-faktor” – ét universelt træk, hvorom man på tværs af arbejdsroller og situationer altid kan sige: Jo højere score, des bedre.¹¹ Fx er der både teoretiske og em-

11 Digman (1997) peger dog på, at FFM’s faktorer korrelerer i en sådan grad, at disse empirisk set er underordnet en “big two” bestående af, hvad han kalder alpha (venlighed, samvittighedsfuldhed og følelsesmæssige reaktioner), og beta (ekstroversion og åbenhed), som har at gøre med hhv. vores evner til at socialisere og til at udvikle os. Rushton og Irwing (2008) used structural equation models to test the hypothesis that a General Factor of Personality (GFP) argumenterer sågar for en “*general factor of personality*”, der med en høj score, analogt til antagelsen om en intelligensmæssig G-faktor, gør, at “de bedst egnede” er selvtillidsfulde og på én og samme tid “[...] *altruistic, emotionally stable, agreeable, conscientious, extraverted, intellectually open, mentally tough, and emotionally intelligent*” (Rushton et al., 2009, s. 356).

piriske tegn på, at høje “integritetsscorer” virker kontraproduktive, når jobperformance har med kreativitet at gøre (Hogan & Brinkmeyer, 1997). Tilsvarende skal den innovative, visionære og strategisk stærke leder næppe være samvittighedsfuld i en grad, hvor vedkommende vanskeligt kan abstrahere fra detaljer, procedurer og autoriteter (Robertson et al., 2000). I et praksisperspektiv taler perspektivet på den måde ind i en diskussion om specifikke vs. generelle kategorier og i det komplicerede forhold, at hvor der på den ene side er situationer, hvor et givent træk har den bedste validitet, så er der samtidig andre, hvor de underliggende og mere snævre facetter står endda stærkere, både når det gælder forudsigelse og forståelse (Tett, Steele & Beauregard, 2003). Således minder Hogan og Judge (2012, s. 4) om, at *“there is no such thing as an ideal personality score – scores must be interpreted in the context of total profiles”*.

En interaktionsmodel under udvikling

Hogans (2005, s. 406) overbevisning er, at “[...] *when personality research is conducted correctly, the validity coefficients match those obtained in good research using measures of cognitive ability*”. Med Borman og Motowidlos (1997) distinktion mellem *“task”* og *“contextual”* performance har der sågar været argumenteret for, at der er sider af performancebegrebet, som kun personlighed må forventes at kunne forklare og prædikere. For hvor task-performance direkte vedrører organisationens kerneopgave og altså typisk er kendetegnet ved noget teknisk eller fagligt, så fokuseres der med contextual performance på den værdi, der fx ligger i at have medarbejdere, som frivilligt påtager sig ekstraopgaver og sågar hjælper deres kolleger med at performe. (Joyce) Hogan og Rybicki (1998) har med afsæt i sådan operationalisering illustreret, hvordan person-situation-diskussionen, såvel som præmissen om én p-faktor, er misforstået, og at der er brug for at kunne forklare en *interaktion*. Den kontekstuelle performance handler således ikke nødvendigvis om at være samvittighedsfuld. Det kan komme an på konteksten, hvad der fordrer fx engagement og samarbejde. Er den organisatoriske kontekst den, at forfremmelser blot følger anciennitet, er samvittighedsfuldhed oplagt en afgørende motivationskilde for værdigenererende adfærd, som ligger ud over den enkeltes medarbejders formelle ansvarsområder. Følges sådan adfærd derimod med mulighederne for at opnå status, kan ambition virke som en endda stærkere “driver”.

I tråd med vores argument ovenfor bygger Hogans optimisme og anbefalinger for genstandsfeltet paradoksalt på en erkendelse af de begrænsninger, som følger med tidens metaanalytiske idealer, hvor jobanalysen ikke har nogen særlig plads (Hogan, Davies & Hogan, 2007). Med tidens evidensbegreb synes studier, hvis metodik er struktureret af den anvendte teoris forud-

sigelse af effektiv ledelse, mistænkelige. De ekskluderes måske endda, og relevante informationer fra lokale validitetsstudier, eller mere case- og praksisbaserede erfaringer for den sags skyld, forsvinder. Konsekvensen er, at det situationelle aspekt underprioriteres. I dette perspektiv undervurderes den viden, som fx en god business-case kan rumme. Følgende case illustrerer således, hvordan en personlighedsteoretisk tilgang kan hjælpe til at træffe velargumenterede strategiske beslutninger, såsom i rekrutteringssammenhænge:

Tre år efter at have bragt HPI-testen i anvendelse til screening af deres lastbilchauffør-ansøgere havde et amerikansk logistikfirma nedbragt antallet af ulykker med 2/3 og forsikringsomkostningerne med 75 %. Ansøgere, som scorede under det 40. percentil for følelsesmæssig balance (emotionelle reaktioner i FFM), *omgængelighed* (venlighed i FFM) og *pålidelighed* (samvittighedsfuldhed i FFM), kom ikke i betragtning. Dette for at undgå en lav *stresstolerance* (og derved forebygge fravær og kompensationskrav mod virksomheden), højne en *serviceminded adfærd* (herunder tolerance, høflighed og interpersonel varme) og forebygge *kriminell adfærd* (som tyveri og narkotikabrug) (Sinclair & Hogan, 1996; Hogan 1998).

Et af få konfirmatoriske studier, der dog er nået igennem til publikation (Hogan, 2005), er (Joyce) Hogan og Hollands (2003) metaanalyse "Using theory to evaluate personality and job-performance relations: A socioanalytic perspective" (N = 5242). Heri forsøgte de af Hogan nævnte metaanalytiske problemstillinger imødekommet. Dette dels ved at inddrage blot én anerkendt personlighedstest (HPI¹²). Herved er alle deltagere testet med den samme test, hvorved man er ude over problemet med at aggregere data fra væsens-forskelligt "måleudstyr". Men samtidig gjorde Hogan og Holland en dyd ud af at anlægge en teoretisk forskningsstrategi, den socioanalytiske (Hogan, 1982; 1996; Hogan, Jones & Cheek, 1985). Sigtet med dette perspektiv er at forklare individuelle forskelle i jobsucces. Teorien lyder, at *to get along* (behovet for social accept) og/eller *to get ahead* (behovet for status) motiverer menneskers organisatoriske adfærd. Således er man ikke blot på udkig efter to forskellige typer performance, når man fx hhv. ønsker sig, at der lægges mange timer på jobbet, eller at kollegial støtte til at hjælpe andre med at nå deres deadlines vægtes. Man er også på udkig efter to forskellige personlighedsprofiler. For antagelsen er, at førstnævnte adfærd drives af metabehovet to get ahead og sidstnævnte af to get along, og at disse tendenser meget vel kan stå i et modsætningsforhold til hinanden: Status kan vække misundelse, hvorfor individuel succes risikerer at komme på bekostning af social accept. Det teoretiske/konfirmatoriske design bestod derfor i, at primærstudiernes kriterier kategoriseredes som enten værende overvejende *get-*

12 HPI-testen er baseret på FFM og Hogans socioanalytiske teori. Den er en 7-faktor 206 items S/F personlighedstest, som er designet til at måle den normale personlighed i en erhvervskontekst (Hogan & Hogan, 2007).

ting ahead- eller *getting along-*adfærd, samt ved at knytte disse an til enten FFM's ekstroversion og åbenhed (synes at kunne have med opnåelsen af status at gøre), følelsesmæssig balance, venlighed og samvittighedsfuldhed (synes at kunne have med social accept at gøre).

Korrelationerne mellem personlighed og jobsucces oversteg med denne tilgang de klassiske (ateoretiske) fund, hvor *getting ahead-* og *getting along-*tendenser sammenblandes, selv om de egentlig er forskellige typer jobperformance og har at gøre med forskellige personlighedstræk. Hogan og Holland (2003) fandt således effektstørrelser på: .43/.25 for følelsesmæssig balance (emotionelle reaktioner i FFM), .35/.20 for ambition (ekstroversion i FFM), .34/.18 for omgængelighed (venlighed i FFM), .36/.22 for pålidelighed (samvittighedsfuldhed i FFM) og .34/.20 intellektuelt åben (åbenhed i FFM). Et bemærkelsesværdigt fund var i øvrigt, at ud over at pålidelighed/samvittighedsfuldhed, som også tidligere påvist, gik på tværs af jobkategorier, så fremstod i dette studie også følelsesmæssig balance/emotionelle reaktioner som en generel prædikator.

Som nyere eksempler på, hvorledes teoretiske bidrag bedre synes at kunne udfolde interaktions-perspektivet, kan nævnes forskning i den såkaldte *adaptive performance*, der med den senmoderne organisations kendetegnende omskiftelighed synes at være blevet et nødvendigt supplement til de ovenfor omtalte kategoriseringer task og contextual performance. Adaptiv performance vedrører jobadfærd knyttet til reaktivt at kunne håndtere uventede forhold (som korrelerer stærkest med følelsesmæssig balance) såvel som proaktivt at kunne initiere forandringer med sigte på at tilpasse sig fremtidens forhold (som korrelerer stærkest med ambition). Med dette fokus bliver den klassiske Barrick og Mount-lærdom, at pålidelighed prædikerer jobeffektivitet (i de fleste job), potentielt anakronistisk (Huang et al., 2013). Et andet eksempel på, hvorledes feltet i dag kombinerer Barrick og Mounts reduktive metoder (kombinationen af FFM med Schmidt og Hunters validitetsgeneralisering), atter med en interaktionsmodel for øje, kunne være Shaffer og Postlethwaites (2013) metaanalyse over hypotesen, at de kognitive krav i jobbet har betydning for, hvilken personlighedsprofil der egner sig bedst. Konklusionen fra dette studie blev, at FFM-dimensionen samvittighedsfuldhed er en stærk prædikator i relation til jobsucces, når det gælder rutinebaserede job, men til gengæld er en svag prædikator, når det gælder job, der kræver en høj grad af kognitiv formåen.

Diskussion

Befolkningen er naturligvis ikke, som beskrevet i Huxleys (1932/1968) science fiction *Fagre nye verden*, alfaer og betaer: Nogle, som kan blive ledere, modsat flertallet, som er arbejdere. Det er et skræmmebillede og har ingenting at gøre med, at vi som organisationspsykologer kan bidrage med

personvurdering og især til udvælgelsen af de rette ledere. Dette, uanset at personlighed ikke kan forklare enhver organisatorisk succes eller elendighed. De omstændigheder, som en given leder agerer i, er centrale for de resultater, som det er muligt at skabe, og for, hvilken type ledelse der vil blive accepteret i organisationen – og derved effektiv. Men det ændrer ikke ved, at lederens personlighed *også* er af central betydning. Til tider *er* særligt egnede lederes personlige karakteristika en væsentlig del af forklaringen på en given virksomheds succes. Og omvendt er der i rigt mål eksempler på, at lederens outrerede personlighed vitterlig er *den* bedste forklaring på en virksomheds deroute. Fx var Stein Bagger naturligvis ikke den eneste ansvarlige for det korrumperede IT Factory. Men mon ikke han var, om ikke *dén* eneste, så en af få, der kunne udnytte at være til stede i en organisation, der åbenbart rummede mulighederne for, at en mand med netop hans personlighed fik gode muligheder for at orkestrere et korrumpert bedrag.

Når man i en organisation skal overveje rekrutteringsstrategier, synes det at være væsentligt at tage hensyn til både aspekter vedrørende de kommende medarbejderes personlighed og aspekter vedrørende de situationer, medarbejderne skal agere i. Disse aspekter skal kunne interagere konstruktivt. Som det er i dag, er personlighedsbegrebet bedre teoretisk defineret og afgrænset end situationsbegrebet, hvorfor en forbedret anvendelse af personlighedsteorien måske i virkeligheden kræver fokus på at udvikle forbedrede modeller for det sidstnævnte. Som det er nu, kan vi derfor bedre foretage en vurdering af, om en leder fx kan siges at være stærkt dominant, omgængelig, empatisk, resultatorienteret m.v., og med større sikkerhed vurdere, hvilke implikationer det har for vedkommendes ledelse, end vi kan reducere og beskrive de omstændigheder eller situationer, vedkommende skal agere i. Vi kan således nemmere intervenere evidensbaseret målrettet på individplanet (ansætte enten en dominant eller en omgængelig leder), end vi kan, når vi forsøger at analysere og intervenere med en ambition om at håndtere en given social kontekst. Så det skal vi gøre – samtidig med at forbedrede forskningsresultater og gode interaktionistiske modeller udvikles. Fx er der brug for: at *vide* meget mere om personlighedstræks interaktioner med hinanden frem for trækkenes generelle betydning hver for sig; et teams personlighedsprofil, der i dag ofte adresseres som et gennemsnit af summen af en given gruppes individer; om sammenhængen mellem træk og destruktiv adfærd (Fruyt & Salgado, 2003).

Virksomheder og organisationer har ofte urealistiske forventninger til anvendelsen af personlighedstest. En personlighedstest skal på én og samme tid helst være hurtig at gennemføre, nem at anvende, effektiv og billig. Når praksis domineres af dette ønske, skyldes det håbet om, at personlighedstestene kan reducere kompleksiteten og derved gøre rekrutteringsarbejdet lettere. Sagen er bare den, at en hensigtsmæssig anvendelse af personlighedstest, som fx beskrevet i casen med lastbilchaufførerne ovenfor, på det praktiske plan snarere gør rekrutteringsarbejdet mere besværligt. Dels kræ-

ver udviklingen af en dækkende jobprofil med efterfølgende kobling til de personlighedsdimensioner, som testen opererer med, et væsentligt forarbejde. Og dels kræver den professionelle anvendelse af personlighedstest en høj grad af teoretisk og praktisk ekspertise, som de færreste virksomheder besidder. Endelig kan en personlighedstest ikke stå alene eller kompensere for de øvrige typiske procedurer i forsøget på at udvælge den bedst egnede kandidat. Der skal fortsat foretages strukturerede interview, trækkes referencer, læses CV og eventuelt færdighedstestes. Set ud fra dette perspektiv har personlighedstest vanskeligt ved at leve op til virksomhedernes forventninger. Man kunne måske endda tale om, at anvendelsen af personlighedstest i et vist omfang bidrager med det modsatte af det, der efterspørges: De fører ikke hverken til hurtige eller nemme rekrutteringspraksisser.

At forkaste personlighedstest på den baggrund vil ikke desto mindre være uhensigtsmæssigt. Kløgtigt anvendt og med den rette investering i solid fagpsykologisk kompetence, gode test og ikke mindst tid kan test bidrage til ønsket om effektivitet. Dette, både når det drejer sig om den rent økonomiske side af sagen i form af højere produktivitet, og når fokus er på de rent menneskelige faktorer, fx ved fejlrekruttering af medarbejdere og ledere og mistrivsel og opsigelser, når den forkerte kandidat hyres.

Den relevante målsætning for personlighedstest handler om, i hvilket omfang de kan bidrage til at finde bedst mulige *matches* mellem person, rolle og organisation. Foruden en specifik jobanalyse bliver personlighedstest til forudsigtelse af performance meningsløs. Leder vi fx mest af alt efter en lærer, som kan sikre a) højt karaktergennemsnit, b) gode relationer til forældrene, c) trivsel, d) faglige udfordringer til de fagligt stærke, e) rummelighed i mødet med klassens udsatte – eller leder vi efter f) en, som mestrer administration? En god begavelse vil som regel være en fordel uanset omstændighederne. Men personer, der ligefrem excellerer i alle dele samtidigt, er ikke desto mindre sjældne, og virksomheder står fra tid til anden i situationer, hvor det synes mest formålstjenligt at prioritere det ene over det andet. Og i forhold til hvem der har den relevante motivation for at lykkes med dette eller hint, mangler intelligensbegrebet forklaringspotentiale.

Konklusion

Artiklen har haft til formål at 1) undersøge, om der er empirisk evidens for, at personligheden kan kvantificeres, og 2) at undersøge forholdet mellem personlighed og jobsucces.

Empirien tyder på, at FFM afspejler stabile egenskaber: træk, som er relativt stabile over tid, og som bliver beskrevet med nogen ensartethed, af de mennesker, vi omgiver os med. Samtidig er trækkene fra FFM at finde på tværs af kulturer og med tilsyneladende arvelighed, fx når enæggede og tve-

æggede tvillinger sammenlignes. Den NEO PI-R-baserede forskning,¹³ som siden 1990'erne har set dagens lys, tyder på en vis konsensus om, at FFM-modellen, de forskellige NEO-inventories og psykometrien bag udgør en meningsfuld personlighedspsykologisk differentialpsykologi.

Ser man på den kvantificerede evidens for, om der er sammenhæng mellem personlighed og jobsucces, bliver svaret mere tvetydigt. Det synes hverken rimeligt at svare højt og klart "ja!" eller "nej!" I stedet må konklusionen være: Der *er* sammenhæng mellem scorer i personlighedstest og jobsucces – men den er i bedste fald moderat.

Således har de kvantitative metoder heller ikke her kunnet afklare psykologiens filosofiske problem med at forstå, hvad det vil sige at være menneske. Den kvantitative metodik, som der her er redegjort for, den metaanalytiske bearbejdning af erhvervspsykologisk psykometri, bringer væsentlig statistisk styrke, men er samtidig en voldsom reduktion. Så selv hvis metoderne forbedres yderligere, er personlighed en så kompleks og dynamisk genstand, at den kvantitative model langt fra altid er meningsfuld og anvendelig. Den kan i hvert fald ikke stå alene. Paradoksalt nok bliver det derfor relevant at overveje, om den empiriske traditions idealer her må tages med et gran salt. I vores optik må det vedvarende overvejes, hvornår metaanalysen er det rette analyseniveau, og om den autoritative status i toppen af tidens videnshierarki nu også sikrer den bedst mulige evidens. Således vurderer vi eksempelvis, at korrelationer mellem personlighed og jobsucces (når nu det metaanalytisk er bekræftet, at de er der) atter bør studeres ved at kigge på personlighedens betydning i specifikke kontekster. Stopper vi ved de metaanalytiske fund, at samvittighedsfuldhed og venlighed generelt er gode egenskaber for en medarbejder at besidde, er den akademiske psykologi endnu ikke nået meget længere end til, hvad erfarne rekrutteringsmedarbejdere altid har vidst: at en vis tendens til at være planlæggende, vedholdende, ansvarlig og hårdarbejdende, samt medfølelse, godmodig og samarbejdende, i de fleste sammenhænge er gode egenskaber at have med sig i arbejdslivet.

REFERENCER

- Barrick, M.R., & Mount, M.K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1), 1-26.
- Barrick, M.R., & Mount, M.K. (2003). Impact of metaanalysis methods. In K.R. Murphy (Ed.), *Validity Generalization: A Critical Review*. New Jersey: Psychology Press, 197-221.
- Borman, W.C., & Motowidlo, S.J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, 10(2), 99-109.
- Cohen, J. (1970). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

13 Godt 3000 videnskabelige artikler er at finde i bibliografien for NEO-inventories: <http://www.sigmaassessmentsystems.com/bibliographies/neo3bibliography.pdf>

- Costa, P.T., & McCrae, R.R. (1991). Trait psychology comes of age. *Nebraska Symposium on Motivation. Nebraska Symposium on Motivation, 39*, 169-204.
- Costa, P.T., & McCrae, R.R. (1996). Toward a new generation of personality theories: Theoretical contexts for the five factor model. In J.S. Wiggins (Ed.), *The Five Factor Model of Personality. Theoretical Perspectives* (pp. 51-87). New York: Guilford Press.
- Digman, J.M. (1997). Higher-order factors of the big five. *Journal of Personality and Social Psychology, 73*, 1246-1256.
- Fruyt, F.D., & Salgado, S.F. (2003). Applied personality psychology: Lessons learned from the IWO field. *European Journal of Personality, 17*, 123-131.
- Guion, R.M., & Gottier, R.F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18*, 135-164.
- Hogan, R. (1982). A socioanalytic theory of personality. *Nebraska Symposium on Motivation, 55-89*.
- Hogan, R. (1996). A socioanalytic perspective on the Five-Factor Model. In J.S. Wiggins: *The Five Factor Model of Personality. Theoretical Perspectives* (pp. 51-87). New York: Guilford Press.
- Hogan, R. (1998). Reinventing personality. *Journal of Social and Clinical Psychology, 17*(1), 1-10.
- Hogan, R. (2005). In defense of personality measurement: New Wine for Old Whiners. *Human Performance, 18*(4), 331-341.
- Hogan R., Hogan, J. (2007). *Hogan Personality Inventory*. Tulsa: Hogan Assessment Systems.
- Hogan, R., Hogan, J., & Roberts, B.W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51*(5), 469-477.
- Hogan, R., Jones, W.H., & Cheek, J.M. (1985) Socioanalytic Theory: An alternative to armadillo psychology. In B.R. Schlenker (Ed.), *The Self and Social Life*. New York: McGraw-Hill Book Company, 175-198.
- Hogan, R., & Judge, T. (2012). Personality and leadership. In M.G. Rumsey (Ed.), *The Oxford Handbook of Leadership* (pp. 1-12). Oxford University Press.
- Hogan, R. DeSoto, C.B., & Solano, C. (1977). Traits, tests, and personality research. *American Psychologist, (April)*, 255-264.
- Hogan, J., & Brinkmeyer, K. (1997). Bridging the gap between overt and personality-based integrity tests. *Personnel Psychology, 50*, 587-599.
- Hogan, J., Davies, S., & Hogan, R. (2007). Generalizing personality-based validity evidence. In M. McPhail (Ed.), *Alternative Validation Strategies. Developing New and Leveraging Existing Validity Evidence* (1st Ed., pp. 181-229). London: John Wiley & Sons, Inc.
- Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology, 88*(1), 100-112.
- Hogan J., & Rybicki, S.L. (1998). Relations between contextual performance, personality, and occupational advancement. *Human Performance, 11*(2/3), 189-207.
- Hough, L.M. (1992). The Big Five personality variables – construct confusion: Description versus prediction. *Human Performance, 5*(1/2), 139-155.
- Huang, J.L., Ryan, A.M., Zabel, K.L., & Palmer, A. (2013). Personality and adaptive performance at work: A meta-analytic investigation. *Journal of Applied Psychology, 99*(1), 162-179.
- Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340-362.
- Hunter, J.E., & Schmidt, F.L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

- Hunter, J.E., & Schmidt, F.L. (1996). Intelligence and job performance: Economic and social implications. *Psychology Public Policy, and Law*, 2(3/4), 447-472.
- Huxley, A. (1932/1968). *Fagre nye verden*. 5. oplag. København: Aschehoug.
- Jang, K.L., McCrae, R.R., Angleitner, A., Riemann, R., & Livesley, W.J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: support for a hierarchical model of personality. *Journal of Personality and Social Psychology*, 74(6), 1556-65.
- Judge, T.A., Bono, J.E., Ilies, R., & Gerhardt, M.W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87(4), 765-780.
- Landy, F.J. (2003). Validity generalization then and now. In K.R. Murphy, *Validity Generalization: A Critical Review*. New Jersey: Psychology Press, 155-195.
- McAdams, D. (1997). A conceptual history of personality psychology. In R. Hogan, J. Johnson & S. Briggs (Ed.), *Handbook of Personality Psychology*. New York: Academic Press, 3-31.
- McCrae, R.R. (2001). Traits Through Time. *Psychological Inquiry*, 12(2), 85-87.
- McCrae, R.R., & Costa, P.T. (1992). Discriminant validity of NEO-PIR facet scales. *Educational and Psychological Measurement*, 52(1), 229-237.
- McCrae, R.R., & Costa, P.T. (1987). Validation of the Five-Factor Model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81-90.
- McCrae, R.R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: data from 50 cultures. *Journal of Personality and Social Psychology*, 88(3), 547-61.
- Meyer, G.J., Finn, S.E., Eyde, L.D., Kay, G.G., Moreland, K.L., Dies, R.R., Eisman, E.J., Kubiszyn, T.W., & Reed, G.M. (2001). Psychological testing and psychological assessment. *American Psychologist*, 56(2), 128-165.
- Mischel, W. (1968). *Personality and Assessment*. New York: Riley.
- Morgeson, F.P., Campion, M.A., Dipboye, R.L., Hollenbeck, J.R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60(3), 683-729.
- Morgeson, F.P., Campion, M.A., Hollenbeck, J.R., Murphy, K., & Schmitt, N. (2007). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology*, 60, 1029-1049.
- Murphy, K., & DeShon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology*, 873-900.
- Murphy, K.R. (2003). The logic of validity generalization. In K.R. Murphy (Ed.), *Validity Generalization: A Critical Review*. New Jersey: Psychology Press, 1-31.
- Murphy, K.R. (2008). Explaining the weak relationship between job performance and ratings of job performance. *Industrial and Organizational Psychology*, 1(2), 148-160.
- Nettle, D. (2011). *Personlighed*. Virum: Dansk Psykologisk Forlag.
- Ones, D.S., Dilchert, S., Viswesvaran, C., & Judge, T. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, 60(4), 995-1027.
- Ones, D.S., Mount, M.K., Barrick, M.R., & Hunter, J.E. (1994). Personality and job performance: A critique of the Tett, Jackson, and Rothstein (1991) meta-analysis. *Personnel Psychology*, 47(1), 147-156.
- Ozer, D.J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin*, 97(2), 307-315.
- Roberts, B.W., Kuncel, N.R., Shiner, R., Caspi, A., & Goldberg, L.R. (2007). The power of personality. The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2(4), 313-345.

- Robertson, I.T., Baron, H., Gibbons, P., MacIver, R., & Nyfield, G. (2000). Conscientiousness and managerial performance. *Journal of Occupational and Organizational Psychology*, 73, 171-180.
- Rothstein, M.G., & Jellay, B.R. (2003). The challenges of aggregating studies of personality. In K.R. Murphy (Ed.), *Validity Generalization: A Critical Review* (pp. 223-262). New Jersey: Psychology Press.
- Rushton, J.P., Bons, T.A., Ando, J., Hur, Y.M., Irwing, P., Vernon, P.A., & Barbaraneli, C. (2009). A general factor of personality from multitrait-multimethod data and cross-national twins. *Twin Research and Human Genetics: The Official Journal of the International Society for Twin Studies*, 12(4), 356-365.
- Rushton, J.P., & Irwing, P. (2008). A general factor of personality (GFP) from two meta-analyses of the Big Five: Digman (1997), and Mount, Barrick, Scullen, and Rounds (2005). *Personality and Individual Differences*, 45(7), 679-683.
- Salgado, J., & Moscoso, S. (1996). Meta-analysis of interrater reliability of job performance ratings in validity studies of personnel selection. *Perceptual and Motor Skills*, 83, 1195-1201.
- Salgado, J., & Moscoso, S. (2003). Information exchange article test-retest reliability of ratings of job performance dimensions in managers. *International Journal of Selection and Assessment*, 11, 98-101.
- Salgado, J., Moscoso, S., & Alonso, P. (2013). Subdimensional structure of the Hogan, personality inventory. *International Journal of Selection and Assessment*, 21(3).
- Schmidt, F.L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47(10), 1173-1181.
- Schmidt, F.L., & Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62(5), 529-540.
- Schmidt, F.L., & Hunter, J.E. (1998) The validity and utility of selection methods in personnel psychology: practical and theoretical findings. *The Psychology Bulletin*, 124, 262-274. American Psychological Association.
- Schmidt, F.L., & Hunter, J.E. (1999). Theory testing and measurement error. *Intelligence*, 27(3), 183-198.
- Schmidt, F.L., & Hunter, J.E. (2001). Meta-analysis. In N. Anderson, D.S. Ones, H. K. Sinangil & C. Viswesavaran (Ed.), *Handbook of Industrial, Work and Organizational Psychology*, 1. Personnel Psychology. London: SAGE Publications Inc., 53-70.
- Shaffer, J.A., & Postlethwaite, B.E. (2013). The validity of conscientiousness for predicting job performance: A meta-analytic test of two hypotheses. *International Journal of Selection and Assessment*, 21(2), 183-199.
- Sinclair, R., & Hogan, R. (1996). The answer is still yes: But what was the question? *The Journal of Applied Behavioral Science*, 32(4), 434-440.
- Smith, M., & Smith, P. (2005). *Testing people at work. Competencies in psychometric testing*. Oxford: Blackwell Publishing.
- Society for Industrial and Organizational Psychology (2003). *Principles for the Validation and Use of Personnel Selection*. Bowling Green: Society for Industrial and Organizational Psychology.
- Terracciano, A., Costa, P.T., & McCrae, R.R. (2006). Personality plasticity after age 30. *Personality & Social Psychology Bulletin*, 32(8), 999-1009.
- Tett, R.P., & Christiansen, N.D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology*, 60(4), 967-993.
- Tett, R.P., Jackson, D.N., & Rohstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44, 703-742.

- Tett, R.R., Jackson, D.N., Rothstein, M., & Reddon, J.R. (1994). Meta-analysis of personality-job performance relations: A reply to Ones, Mount, Barrick, and Hunter (1994). *Personnel Psychology*, *47*(1), 157-172.
- Tett, R.P., Steele, J.R., & Beauregard, R.S. (2003). Broad and narrow measures on both sides of the personality-job performance relationship. *Journal of Organizational Behavior*, *24*, 335-356.
- Tupes, E.C., & Christal, R.E. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality*, *60*(2), June 1992, pp. 225-251.
- Viswesvaran, C. (2001). Assessment of individual job performance: A review of the past century and a look ahead. In N. Anderson, D.S. Ones, H.K. Sinangil, & C. Viswesvaran (Ed.) *Handbook of Industrial, Work and Organizational Psychology*, 1. Personnel Psychology, London: SAGE Publications Inc., 110-126.
- Viswesvaran, C., Schmidt, F.L., & Ones, D.S. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, *81*(5), 557-574.