

VU Research Portal

Beoordelen loont

Jansen, Paul

2021

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Jansen, P. (2021). *Beoordelen loont*.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

prof.dr. Paul Jansen

BEOORDELEN LOONT

Rede uitgesproken bij zijn afscheid als hoogleraar Bedrijfspsychologie bij de School of Business and Economics van de Vrije Universiteit Amsterdam op dinsdag 14 september 2021.

BEOORDELEN LOONT

1. Gedrag tegenover prestatie/resultaat.

Beoordelen is het doen van uitspraken doen over personen. Die uitspraken vinden expliciet plaats in een sociale werkcontext. Dat maakt dat je als manager vastzit aan je beoordelingsuitspraak. Je committeert je eraan. Maar de uitspraak heeft ook consequenties voor de beoordeelde, in de zin dat deze een promotie wel of niet krijgt of een bonus in het vooruitzicht wordt gesteld. Searle (1998) noemt zulke taalhandelingen performatieve taaldaden. Het (uit)spreken heeft tot doel iets te veranderen in de wereld. Van belang is dat degene die de taaldaad verricht daartoe institutionele bevoegdheid bezit. In een hiërarchische arbeidsorganisatie is daarom beoordelen vaak verbonden aan een organieke positie als afdelingsmanager. Beoordelen komt daarmee in een bedrijfskundige context van de balans tussen bevoegdheid en verantwoordelijkheid. In deze bijdrage zal ik, conform ook mijn leeropdracht, beoordelen daarom zowel psychologisch als bedrijfskundig beschouwen. Zo zal ik de vraag stellen: betekent het einde van beoordelen omdat het psychologisch slecht werkt, ook dat er geen bedrijfskundige aanleiding meer is om uitspraken te doen over arbeidsgedrag?

Bedrijfspsychologie richt zich op menselijk gedrag als samenwerken, beïnvloeden, ontwikkelen, organiseren in de context van een organisatie. Gedrag is wat wij met onze zintuigen ervaren, waarbij horen en zien de belangrijkste zijn. U neemt nooit gedrag in het algemeen waar. U ziet mij bijvoorbeeld nooit in het algemeen maar altijd in een specifieke (arbeids)situatie, zoals deze waarin ik mij vandaag bevind. Mensen leven in situaties. Meer precies is een situatie een in tijd besloten dagdeel waarin iets op het spel staat: je moet iets bereiken, je wil 'het goed doen'. Dat geldt bijvoorbeeld voor een onderhandeling of hoorcollege of werkvoortgangsgesprek, en zelfs voor een informeel gesprek op de gang. In de context van werkorganisaties kun je werkgedrag ordenen aan de hand van een beperkt aantal kritische werksituaties: situaties die je absoluut moet beheersen om je functie goed te kunnen vervullen. In het geval van een hoogleraar bijvoorbeeld een hoorcollege geven of een voortgangsgesprek voeren met een promovenda. In al die gevallen is er sprake van een criterium voor effectieve situatiebeheersing waardoor je achteraf kunt zeggen of je het goed hebt gedaan. Nog belangrijker: een situatie waarin anderen zeggen dat het goed is gegaan. Ik kom later terug op dit typisch bedrijfskundige klantkenmerk van een kritische werksituatie. Als actor moet je dat criterium uit de situatie lezen of, als dat je in bijvoorbeeld hiërarchische of professionele macht ligt, aan de situatie opleggen. Ik geef mijn studenten om die reden altijd mee om tijdens een sollicitatiegesprek aan de werkgever te vragen: *Wat moet ik hier doen om het goed te doen?*

Gelet hierop zou een beoordelingsuitspraak dus de volgende vorm moeten hebben: *In een kritische werksituatie van bijvoorbeeld een verloopadviesgesprek voeren doet U dit en dat en bent U daarom minder effectief gelet op het criterium van een tevreden klant.* Als het criterium verandert, bijvoorbeeld wat telt is een verkochte auto, krijgt ook de beoordeling een andere inhoud. De sociale bedrijfssituatie stelt de norm. Een ander voorbeeld: ik bevind vandaag mij in een specifieke en ook kritische werksituatie met een daarbij passende norm voor effectiviteit. Ik vertoon gedrag; of ik hier presteer is afhankelijk van de mate waarin ik voldoen aan het criterium voor effectiviteit. Lastig is dat dat criterium, d.w.z. de opvatting over wat de situatie eigenlijk behelst, kan verschillen. Sommigen komen vandaag voor de

inhoud, andere voor de uitstraling van een zekere waardigheid, en weer anderen om op andere gedachten te komen. Zo bezien kan ik het hier dus nooit helemaal goed doen.

We kunnen concluderen dat gedrag altijd situationeel is ingeperkt – U ziet nooit iemand iets in het algemeen doen-, en dat het combineren van geobserveerd gedrag met een criterium leidt tot een prestatie: $Prestatie = \text{gedrag} + \text{norm}$. Zoals gezegd: wil je kunnen presteren dan moet je de norm kennen (wat moet ik hier doen om het goed te doen?) en proberen eraan te voldoen. Die norm moet dan wel in gedragstermen worden vertaald, anders blijft het een vrijzwevende abstractie. Je moet dus gedrag maken van theoretische constructen als proactiviteit of klantgerichtheid; daardoor maak je die begrippen meetbaar. Naast voldoen aan de norm, kun je natuurlijk ook proberen die norm te beïnvloeden, bijvoorbeeld om die te verlagen.

Dit alles is klassiek onderdeel van het zogenoemde planningsgesprek waarin bij de start van een nieuwe beoordelingsperiode wordt afgesproken welke prestaties gerealiseerd moeten worden: welk gedrag tegen welk norm. Nu zou men kunnen tegenwerpen dat de gedachte van een planningsgesprek als start van een als het ware bevroren beoordelingsperiode met tevoren afgesproken prestaties en normen, een fictie is vanwege dat tegenwoordig alles permanent dynamisch is geworden. Ik kom hierop later terug omdat dit onderdeel is van de bedrijfskundige context van beoordelen – maar zelf relatieveer ik altijd die dynamiek. Ik gebruik altijd het voorbeeld van wat mijn oma (1903 – 1989) in haar leven heeft meegemaakt: grote kinderschare van wie een aantal al jong overleed (dat hoorde erbij), 2 wereldoorlogen, jaren dertig crisis van de vorige eeuw, de atoombom, opkomst van televisie, van de auto, de pil, vliegvakanties,

Je kunt ook proberen een algemene uitspraak over iemand te doen; dan laat je de situationele inbedding los. Omdat daarmee ook de norm verdwijnt doe je een uitspraak over gedrag die generaliseert over een aantal kritische werksituaties. In plaats van vaststellen of iemand in een bepaalde, kritische situatie voldoet aan een specifiek norm, telt nu hoe iemand zich doorgaans gedraagt. Van een specifieke uitspraak gericht op een bepaalde situatie met een daarbij horende norm, bijvoorbeeld '*Op tijdstip T vertoonde persoon V gedrag X in situatie S*', ga je over op een uitspraak over die persoon in het algemeen. Dergelijke elementaire gedragsobservaties kunnen ofwel bestudeerd worden vanuit de achtergrond van persoonsgebonden processen die zich over verschillende arbeidssituaties manifesteren, ofwel onderzocht worden vanuit het referentiekader van situatiegebonden condities die de verschillende proceselementen van dat gedrag tot één, vanuit de situatie in kwestie zinvol geheel smeden. Elementair gedrag kan opgevat worden als een uiting van bijvoorbeeld een geheugenproces of een taalproces, en als zodanig bestudeerd worden. In dat geval wordt geabstraheerd van de concrete arbeidssituatie waarin dat proces plaatsvond. En anderzijds kan ervoor gekozen worden hetzelfde elementaire gedragsvoorval te bezien vanuit het oogpunt van kritische kenmerken van de situationele context waaronder het tot stand kwam, bijvoorbeeld een reorganisatie doorvoeren. Zowel in de bedrijfspsychologie, maar ook in een meer praktische, bedrijfskundige context is het van belang beide perspectieven te kunnen hanteren. Het voorbeeld demonstreert ook dat het mogelijk is de beoordeling van in dit geval commercieel management te baseren op ofwel kenmerken van de persoon in kwestie: die moet in het algemeen goed kunnen denken en stevig zijn, ofwel kritische situaties in de functie: zo'n manager moet in die situaties goed kunnen presenteren en onderhandelen met het oog op de realisatie van gegeven (of zelf bedachte) criteria.

Ik heb jarenlang in mijn colleges Bedrijfspsychologie voor bedrijfskundig bedrijfseconomen getracht studenten op deze wijze te leren kijken naar gedrag (Jansen, 2002). Merk op dat door

te generaliseren over situaties heen de norm erg vaag wordt, zo niet verdampt. Het is inmiddels gelukkig een platitide te beweren dat beoordelingsgesprekken veel beter lopen als ze de inhoud hebben van: *Wat zie je als kritische werksituaties? Waarom? Wat staat daar op het spel? Hoe ging jou dat af? Geef een voorbeeld? Kortom: wat heb je gedaan zodat het gelet op welk criterium goed ging?* Dit wordt bevestigd door talloos onderzoek naar condities van effectieve feedback. Ik ben erg voor vasthouden aan een dergelijke situationele norm. Dus niet: zij is intelligent maar: Dat coachende gesprek met een medewerker pakte zij effectief aan gelet op het criterium van competentieverbetering, maar minder effectief met het oog op de norm van vergroten van authenticiteit. Maar U moet eens nagaan hoe vaak dergelijke context-verbijzonderingen voorkomen in artikels in psychologische tijdschriften. Die situationele inbedding komt meestal pas aan de orde bij de paragraaf '*Suggestions for practice*'. Terwijl de kern van mijn betoog is dat deze thuishoort in de sectie '*Contributions to theory*'.

De psychologie loopt over van theoretische begrippen voor persoonsdimensies. We hebben een uitgebreide personentaal, te uitgebreid naar mijn smaak hoewel er de laatste decennia m.b.v. bijvoorbeeld de Big Five taxonomie wat meer ordening plaatsvindt. Vraagstukken worden nog te vaak op scholastieke wijze opgelost door er nieuwe begrippen voor te bedenken (Jansen & Eijkelenburg, 1998). Dit alles is in overeenstemming met het idee dat psychologie over mensen gaat. Maar weer verwijzend naar mijn hoorcollege reeks Bedrijfspsychologie: Ik heb studenten steevast onderwezen dat psychologie over situaties gaat. De vraag '*Waarom doet iemand zo?*' wordt dan geherformuleerd tot de vraag: Wat in deze situatie maakt dat dat gedrag in de ogen van deze persoon effectief is, of simpeler gezegd: persoonlijk belonend is? Wat is dus diens persoonlijke criterium voor effectiviteit? *Wat is hier lonend?* Ik heb meer aan een beschrijving van de werksituatie inclusief criterium van effectiviteit om het gedrag van iemand te voorspellen, en dus te beïnvloeden, dan aan een psychologisch rapport. Merk ook op dat de (sociale) situatie de norm definieert. U weet wat U hier van mij kunt verwachten omdat U ongeveer weet wat een afscheidscollege inhoudt. U kunt dus mijn te verwachten gedrag beter voorspellen op grond van deze werksituatie dan op grond van een psychologisch rapport over mij.

Deze wellicht bijna manisch nadruk op de situatie hangt denk ik sterk samen met het gegeven dat ik van meet af aan (sedert maart 1992) heb samengewerkt met economen, bedrijfseconomen, bedrijfskundigen en ook, en met veel plezier, met accountants. In die wetenschappen ligt vanzelf de nadruk op de context omdat men in de economie en bedrijfskunde personen over een kam scheert. *Mensen gaan voor maximalisatie van persoonlijk gewin*. Merk op dat dit niet eens zo veel lijkt te verschillen van mijn diagnosevraag: *wat is hier lonend?*; maar in de psychologie is het gewin sterk individueel en subjectief bepaald. Dat mensen bijvoorbeeld in het onderwijs prestatiebeloning afwijzen omdat financiële beloning geen primair drijfveer is in hun werk, is voor economen een wetenschappelijk resultaat, en voor bedrijfspsychologen een trivialiteit.

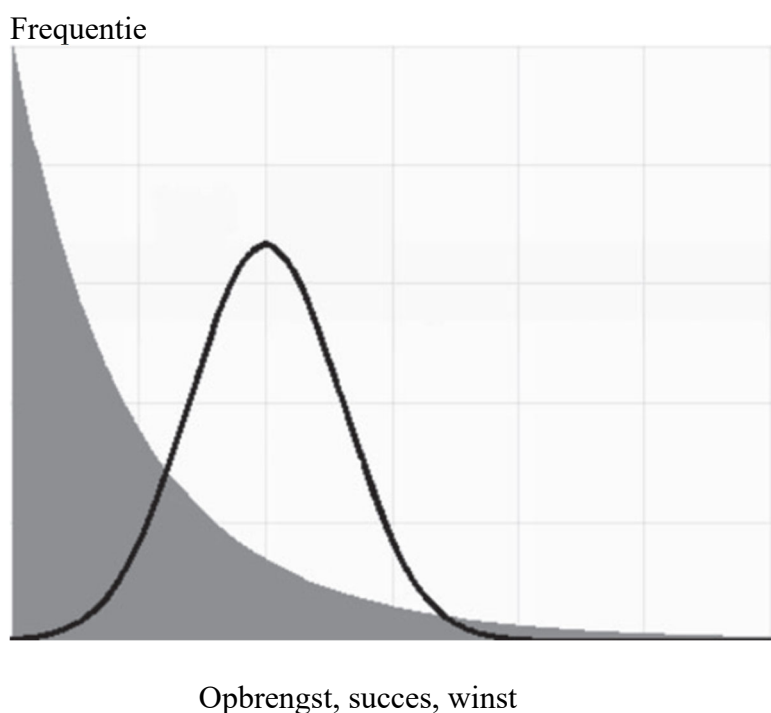
Ik concentreer vandaag op de situationele context van beoordelen en richt mij daarbij op twee situationele categorieën:

a. Gaat het om de interne prestatie binnen de organisatie (bijvoorbeeld een effectief hoorcollege geven) of om het externe resultaat als gevolg van die prestatie in de zin van opbrengst op de een of andere markt van de organisatie? Dat kan de consumentenmarkt zijn, of de arbeidsmarkt maar ook de financiële of die van de politiek. Het zal duidelijk zijn dat de laatste wijdere benadering van personeelsbeoordeling bij uitstek vraagt om een bedrijfskundige blik.

b. Gaat het om een absolute prestatie of een relatieve prestatie?. In het eerste geval worden de prestaties van een medewerker afgezet tegen een absoluut criterium, bijvoorbeeld zoveel artikels gepubliceerd of een kwaliteitsbeoordeling hoger dan 4 op een 5-puntsschaal. In het tweede geval worden prestaties van medewerkers onderling vergeleken: wie deed het het beste, wie was het meest effectief in de situatie?
Ik werk beide kenmerken van de beoordeling uit.

2. Verdeling van prestatie, resultaten of uitkomsten

In 2012 publiceerden O'Boyle en Aguinis (2012) een onderzoek naar de verdeling van individuele prestaties. Zij bekritiseerden daarin de aanname van veel bedrijfspsychologische theorie en praktijk dat prestaties normaal verdeeld zijn. In de praktijk blijkt dat idee niet te kloppen stellen zij. In plaats van een normale verdeling is er nogal eens sprake van een zogenoemde Paretoverdeling. Die is sterk rechts-scheef, d.w.z. een groot deel van de massa van de verdeling ligt aan de linkerkant op de X-as (Figuur 1).



Figuur 1 *Normale verdeling (vet) en Pareto verdeling (grijs; O'Boyle & Aguinis, 2012).*

O'Boyle en Aguinis vinden Paretoverdelingen voor onder andere de frequentie van tijdschriftpublicaties (veel mensen hebben weinig publicaties, weinig mensen hebben er veel), de frequentie van Emmy nominaties (veel mensen hebben weinig nominaties, weinig mensen hebben er meerdere), en bijvoorbeeld carrièresucces in base ball. Oorspronkelijk gebruikte Pareto deze verdeling als model voor de verdeling van rijkdom. Het Pareto-principe, bekend als de "80-20"-regel, zegt dat 20% van de bevolking 80% van de rijkdom bezit. De Paretoverdeling is geschikt om zo'n situatie te beschrijven. Het principe is ook op andere situaties van toepassing, zoals:

- De frequentie van de woorden in tekst (slechts een klein aantal woorden vormt het grootste deel van de tekst)
- De grootte van bedrijven (een klein aantal grote bedrijven, een groot aantal kleine).
- De omzetverdeling van boeken. Relatief weinig titels, de ‘bestsellers’, zorgen voor relatief veel omzet (Chin-A-Fo & Jaeger, 2020).

O’Boyle en Aguinis (2012, p. 79): *“We revisit a long-held assumption in human resource management, organizational behavior, and industrial and organizational psychology that individual performance follows a Gaussian (normal) distribution. We conducted 5 studies involving 198 samples including 633,263 researchers, entertainers, politicians, and amateur and professional athletes. Results are remarkably consistent across industries, types of jobs, types of performance measures, and time frames and indicate that individual performance is not normally distributed—instead, it follows a Paretian (power law) distribution. Assuming normality of individual performance can lead to miss specified theories and misleading practices”*.

O’Boyle en Aguinis hebben absoluut gelijk ons erop te wijzen dat niet alles normaal is verdeeld. Maar zij vallen in de valkuil dat een Pareto-verdeling ‘natuurlijk’ is. Het is nu eenmaal zo. Ik werk dat uit. In het citaat hebben zij het consequent over de individuele prestatie. Maar waarvoor staat in hun studies eigenlijk het begrip prestatie? En hoe individueel is die? Hoe komen prestaties als publiceren in een toptijdschrift of het verkrijgen van een Emmy-nominatie tot stand? Wat is de aard van het achterliggende *productiemechanisme van de prestatie?*

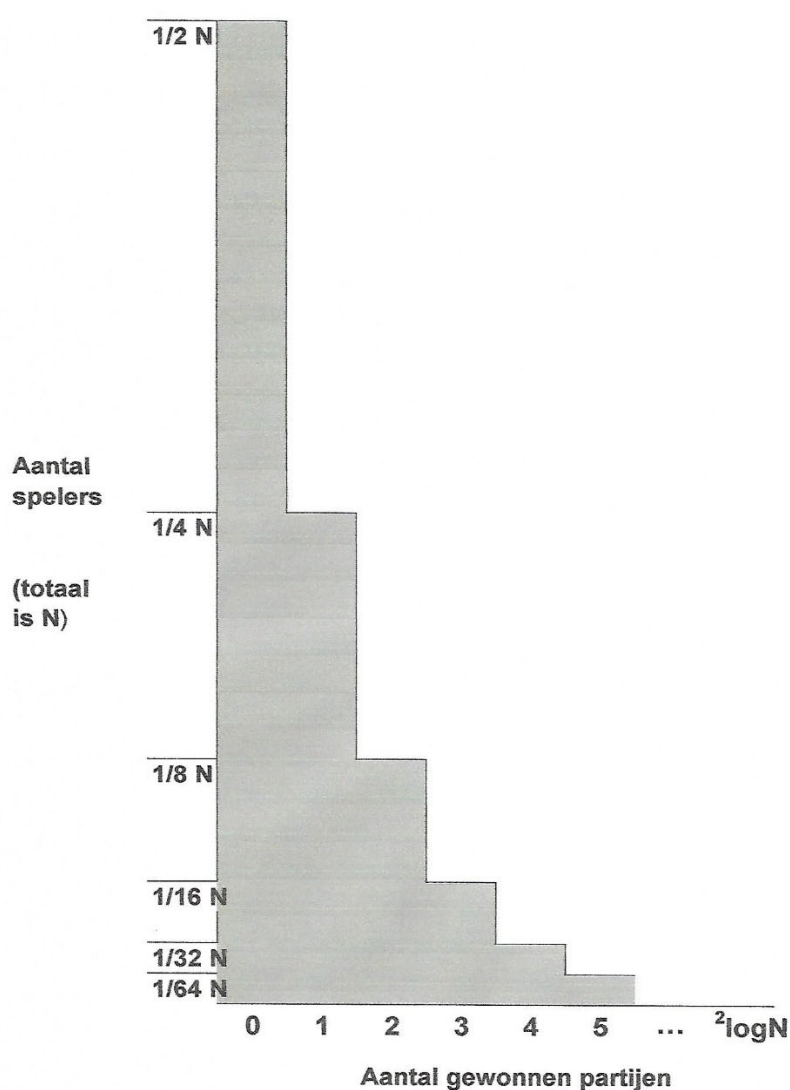
In latere publicaties zijn Aguinis en collega’s hierover meer genuanceerd. Zij hanteren nu het begrip resultaat in plaats van prestatie. Er is sprake van een procesketen: medewerkers presteren individueel in hun werk en dat leidt vervolgens tot een, al dan niet collectief resultaat zoals beoordeeld door de markt: *“O’Boyle and Aguinis (2012) and Aguinis and O’Boyle (2014) adopted the results-based definition of performance because “a focus on results rather than behaviors is most appropriate when (a) workers are skilled in the needed behaviors, (b) behaviors and results are obviously related, and (c) there are many ways to do the job right” (Aguinis & O’Boyle, 2014, p. 316)”* (Aguinis, O’Boyle, Gonzalez-Mulé, & Joo, 2016).

Er is niets tegen om van individuele prestatie over te gaan op behaald resultaat in wat ik aanduid als een markt: Een verzameling externe partijen of klanten die zitten te wachten op de prestatie, die daar iets mee doen en die er daarom iets van vinden. Bij de beoordeling van universitair medewerkers kijk je niet alleen naar hoe deze bijvoorbeeld college geeft maar ook wat het resultaat is van college-geven in termen van proximaal kennisvermeerdering bij studenten (bijvoorbeeld gemeten door tentamencijfers) en distaal succes in het latere werk (bijvoorbeeld als snelheid van salarisgroei). Business-schools hanteren het laatste als een van hun succescriteria.

Echter: Het is niet moeilijk aan te tonen dat door over te gaan van individuele prestatie op resultaat in een externe markt (bijvoorbeeld die van redacties van wetenschappelijk tijdschriften of van Emmy-jury’s) de Paretoverdeling juist ontstaat. Ik leg dit uit aan de hand van een eenvoudig voorbeeld. Laten we aannemen dat schaken als individuele competentie normaal verdeeld is. D.w.z., er zijn heel weinig topschakers, ook heel weinig beroerde schakers en er is een grote groep middelmatige schakers. Zou je de schaakprestatie individueel meten, bijvoorbeeld in de vorm van het oplossen van schaakvraagstukken zoals

die dagelijks in de krant staan, dan vind je dus een normale verdeling. Maar wat als je de schaakprestatie meet in de vorm van succes in een schaaktoernooi?

Stel N personen spelen steeds in tweetallen tegen elkaar. De winnaar gaat door naar de volgende ronde, de verliezer doet verder niet meer mee. Hoe zit dan de verdeling van gewonnen partijen eruit? De helft van de spelers, dus $\frac{1}{2}N$, heeft 0 gewonnen partijen. Van de overgebleven schakers valt weer de helft, dus $\frac{1}{4}N$, af in de 2^e ronde. Dus een kwart van de spelers, dus $\frac{1}{4}N$, heeft 1 gewonnen partij. Enzovoorts. Zetten we, net als Aguinis en collega's, het resultaat, in dit geval het aantal gewonnen partijen (vergelijk: aantal gepubliceerde tijdschriftartikels of aantal Emmy-nominaties), op de X-as en de frequentie op de Y-as dan ontstaat een Pareto-verdeling (Figuur 2).



Figuur 2. Verdeling van aantal gewonnen partijen bij een schaaktoernooi met N spelers.

Dus: door schaakcompetentie te meten als het resultaat van een toernooistructuur hebben we de prestatieverdeling veranderd van een normale verdeling (dat was tenslotte de aanname) naar een Paretoverdeling. Merk overigens op dat die Paretoverdeling ontstaat onafhankelijk van de onderliggende verdeling van de individuele schaakvaardigheid. Door de keuze van een toernooistructuur hebben we een Paretoverdeling *geforceerd*.

Dat roept de vraag op: In hoeverre was er in de studies van Aguinis en collega's sprake van een dergelijke toernooistructuur? Door de organisatie van een toernooi zijn we overgegaan van absolute meting van competenties als schaken of zingen of verkopen of college geven op relatieve meting. In plaats van dat de prestatie wordt bepaald aan de hand van een competentie- of functienorm, wordt die de uitkomst van vergelijking van personen onderling. Bij absolute meting kan iedereen een 10 krijgen, bij relatieve meting en beoordeling in een toernooi gaat de winst van de een ten koste van de ander. *Winner takes all*. Het gaat bijvoorbeeld om het aantal publicaties in toptijdschriften, aantal nominaties voor beste acteur of voor beste roman. Inderdaad zijn de prestaties in de onderzoeken van Aguinis en collega's in hoge mate gebaseerd op resultaten van interpersoonlijk competities, bijvoorbeeld in wedstrijdverband of in de vorm van een concours. Een belangrijk kenmerk daarvan is schaarste: niet iedereen kan een Emmy krijgen en maar een klein deel kan publiceren in toptijdschriften (redacties melden trots hun 'rejection rates').

Vinkenburg, Jansen, Dries en Pepermans (2014) betogen hoe deze structurering van beoordelen en op grond daarvan kiezen van een winnaar, door hen aangeduid als arena, een verklaring geeft voor de schaarste aan vrouwen op topposities in arbeidsorganisaties. Immers, ook het aantal topmanagementposities of hoogleraarsposities is schaars.

Ook andere wetenschappers hebben verder onderzoek gedaan naar de 'ware aard' van resultaatverdelingen teneinde te komen tot "*theory regarding individual output distributions by offering greater precision beyond the normal versus nonnormal dichotomy adopted in prior studies*" (Joo, Aguinis & Bradley, 2017, p. 1039). Terecht gaat men daarbij in op de generatieve mechanismen die tot de waargenomen outputverdeling leiden – vergelijk mijn concours- of arenavoorbeeld. Ik heb niets tegen dit soort onderzoek. Een Paretoverdeling van uitkomsten is als je van het generatieve mechanisme weet hebt en als dat mechanisme in de betreffende context algemeen aanvaard is, even veel of weinig 'natuurlijk' als een normale verdeling.

Er ontstaat evenwel een probleem als de Paretoverdeling, en daarmee ook het generatieve mechanisme, als natuurlijk wordt beschouwd, d.w.z. als iets wat 'nu eenmaal zo is' - , en daaraan vervolgens praktische consequenties worden verbonden door bijvoorbeeld alle resources in te zetten op de 'toppers'. Vergelijk bijvoorbeeld: "*Too generate greater overall output, one recommendation is to heavily invest in individuals with higher output accumulation rates than others*" (Joo, Aguinis & Bradley, 2017, p. 1027). In een dergelijk naïef-positivistische aanpak wordt een 'natuurlijke Paretoverdeling' van prestaties als onderbouwing gebruikt voor een Paretoverdeling in belonen. Het schaakvoorbeeld laat evenwel zien dat een Paretoverdeling van prestaties of resultaten helemaal niet zo een natuurlijk gegeven is. Een extra laatste voorbeeld. Bij een relatieve tentamenbeoordeling, bijvoorbeeld door een rangorde te maken, krijgt maar één student een 10 en ook één student een 1, bij absolute beoordeling kan iedereen een 10 krijgen (of een 1).

Een toernooimechanisme voor de bepaling van winnaars heeft nog een ander gevolg. Lengte en sprongkracht en ervaring bieden voordeel voor basketballen als absolute prestatie. Maar in

een top league waarin alle basketballers lang zijn, goed kunnen springen en veel ervaring hebben, zijn dit geen bepalende factoren meer voor de relatieve prestatie van winnen. Omdat alle hoogleraren slim zijn is intelligentie geen verklaring voor het carrièresucces van hoogleraren. Selectie op extremen (de top van de curve) leidt tot irrelevante predictoren want de relatieve toernooiprestatie wordt meer en meer een kwestie van onbelangrijke zaken – en dat biedt overigens ook de ruimte voor allerlei biases (Vinkenburg et al., 2014). Het beste is dan te loten. NWO is daar nog niet achter: Men krijgt in de laatste rondes allemaal topvoorstellen ter beoordeling, met scores van 4.90, 4.91, 4.92 etc. op een 5-puntsschaal. Grote kans dat de eindkeuze dan plaatsvindt op emergente en dus irrelevante zaken; en logisch dat de oordelen onderling weinig consistent blijken te zijn (Merkelbach, 2020).

Samenvattend: Door de keuze van het resultaatmechanisme verandert een normale verdeling in een Paretoverdeling waarbij weinigen heel veel krijgen en velen heel weinig. Dat mechanisme komt vaak neer op relatief beoordelen in plaats van absoluut. Dus niet: ik en mijn werk, maar: *ik en jij*, of preciezer: *ik of jij*. Ik kan hierbij verwijzen naar een discussie die jaren geleden speelde binnen onze faculteit over dat beoordelingsscores normaal verdeeld moesten zijn, d.w.z. de meeste scores moesten in het midden van een 5-puntschaal liggen. Ook in dat geval is er sprake van een geforceerde normaalverdeling waarin het oordeel ‘zeer goed’ schaars is geworden en je als manager gedwongen wordt van een absolute beoordeling over te gaan op een relatieve. Mijn vraag was toen: geldt dit dan ook voor tentamenresultaten? Met andere woorden, is het evenzo ‘natuurlijk’ dat de helft van de studenten een onvoldoende scoort? Mijn eigen antwoord was toen (verwoord in een bericht aan de decaan): zowel voor studenten als voor medewerkers moet het streven zijn: iedereen een tien! Terecht benoemen Markouilli et al. (2017) in hun veelomvattend overzicht dat HRM-onderzoek naar ‘Assessing People’ profijt kan hebben van het loslaten van een naïef positivisme waarin er sprake is van één ware prestatie. In plaats daarvan moet er aandacht zijn voor wat ik heb aangeduid als het uitkomstmechanisme van de beoordeling.

3. Extern resultaat als gevolg van de interne prestatie

We zagen hoe belangrijk de keuze van een uitkomstmechanisme is voor de omzetting van een individuele prestatie in een resultaat op een externe ‘markt’. En aan die keuze is, niettegenstaande alle tegengeluiden, niets natuurlijks. Er is bijvoorbeeld een verschil tussen het gedrag, en bijgevolg de persoonlijk prestatie van een topmanager en het bedrijfsresultaat gemeten als bijvoorbeeld de beurskoers. Hetzelfde geldt voor het verschil tussen individueel pianospelen of optreden in een concours, of een onderzoeksplan maken en dat moeten zien geaccepteerd te krijgen in NWO-rondes. Wetenschappers worden daarom voor dat laatste speciaal getraind. *Je hebt een bedrijfskundige redenering nodig om het verband te leggen tussen individuele prestatie en resultaat op een externe ‘markt’*. En op dezelfde wijze is een bedrijfskundige redenering nodig om een relatie te leggen tussen de mate waarin een hoogleraar college geeft, en het percentage geslaagde studenten voor het desbetreffende vak. In beide gevallen beweeg je je van een interne, absolute prestatie naar uitkomsttransacties in externe markten waar de beoordeling vaak relatief van aard is. Prestatie en resultaat hebben beide hun plek, maar je moet ze niet verwarren.

Nu kun je zeggen: Om succes te hebben in bijvoorbeeld schaken moet je niet alleen je best doen in de zin van een hoge schaakcompetentie bezitten, maar ook in staat zijn effectief te zijn in het outcome-mechanisme dat hoort bij hoe de markt van schaken gemeenlijk is vormgegeven. Dus niet alleen presteren maar daarmee ook resultaat boeken. Je kunt wel heel goed zijn in je werk en veel publiceren, maar als je het *reputatiemechanisme* van de

wetenschap niet beheerst, word je nooit hoogleraar. Kennis en beheersing van het resultaatmechanisme horen erbij. Dus niet alleen: *Wat moet ik hier doen om het goed te doen?* maar ook: *Hoe word je hier eigenlijk topmanager* (of hoogleraar)?

Dit klopt, maar het blijft zo dat het een organisatiekeuze is een bijdrage te meten als individuele prestatie of als resultaat in een markt van onderlinge competitie. En kritisch psychologen (onder andere Bal, 2020) hebben gelijk dat gewoonlijk dat mechanisme niet ter discussie wordt gesteld terwijl dat juist nodig kan zijn vanwege bijvoorbeeld uitsluiting van bepaalde groepen. Is een pianist die het in de studio geweldig doet maar in een competitieve concoursomgeving faalt, een slechte pianist? Het tweede is veel meer een organisatiekeuze dan het eerste. Maar bovendien biedt juist de tweede beoordelingsetting de ruimte aan de nadelen van een ‘beautycontext’, bijvoorbeeld de grotere kans dat willekeurige, emergente criteria als kleur of geslacht gaan domineren (Vinkenburg et al., 2014). Niets voor niets dat in onderzoeken naar de lange termijn predictieve validiteit van assessment centers (en dan heb ik het over meer dan 10 jaar) juist een competentie als dominantie, je stevig positioneren belangrijker blijkt voor objectief carrièresucces in de zin van de snelheid van stijgen in de hiërarchie, dan intelligentie (Jansen & Stoop, 2001; Jansen & Vinkenburg, 2006). Niet omdat intelligentie niet nodig is, maar omdat bij relatieve selectie in zulk een concourscontext niet intelligentie meer doorslaggevend is, maar stevigheid en assertiviteit. Ik herhaal: In latere loopbaanfasen is het dan niet meer *ik en mijn werk*, maar: *ik en jij*, of preciezer: *ik of jij*.

Bij de overgang van individuele absolute prestatie naar het relatieve succes van een externe marktbijdrage, maken we dus de overgang van bedrijfspsychologie naar bedrijfskunde. Je prestatie moet tot iets leiden, je moet niet alleen presteren maar ook effect hebben. Je moet een antwoord hebben op de vraag, *‘Wie zit er op jouw prestatie te wachten?’* Dus: *‘Voor wie werk je eigenlijk?’* En dan gaat de bedrijfsmatige of politieke of culturele prestatiecontext tellen. In geval van *absoluut beoordelen* wordt functiegedrag gewogen aan de hand van een norm: hoe goed schaakt zij of kookt hij? De beoordelaar moet verstand hebben van dit ‘echte werk’. Ik mag graag kijken naar televisieprogramma’s waarin vakkundige koks of muzikanten uitleggen hoe zij de prestaties van anderen beoordelen. Je leert daarvan bijvoorbeeld beter naar klassieke muziek te luisteren.

Het bijzondere van relatief beoordelen is dat je geen vakdeskundige hoeft te zijn. U hoeft niet gepromoveerd te zijn om te beoordelen of dit een van de meest leuke, boeiende of saaie dan wel onbegrijpelijk lezingen is die u recentelijk hebt gehoord. Als klant in de externe markt van de organisatie beoordeelt U niet de functieprestatie op zich maar de mate waarin deze voldoet aan het door U gehanteerde criterium. Uw beoordeling van de bijdrage van de prestatie is niet alleen buitengewoon subjectief *maar mag dat ook zijn*: “Dit gerecht is lekkerder dan dat, dat muziekstuk klonk beter dan dat”. Op de vraag naar het waarom van de beoordeling is de reactie terecht: *Het was gewoon lekkerder, of het klonk gewoon beter*. Klanten zijn geen professionals in beoordelen, biases bij beoordelen zijn in deze context irrelevant. Die had je als leverancier van de prestatie maar moeten meenemen in de mechanisme waarmee je de prestatie omzet in je marktbijdrage. Bedrijven zien het laatste zelfs als de kern van hun marketing.

In een bedrijfskundige benadering van beoordeling is er dus uitdrukkelijk plaats voor deze klanthouding! Men kan weliswaar overwegen personeelsbeoordeling weg te doen omdat het een psychologisch slecht werkend proces zou zijn. Maar vanuit het oogpunt van managementcontrol zal men niet gauw beoordelen afschaffen in een bedrijfsorganisatie. Ik weet niet hoe het met U is, maar ik krijg na iedere transactie via internet stevast het

vriendelijke verzoek mee te werken aan een korte enquête om de dienst te *beoordelen*. In de bedrijfsvoering levert een medewerker een product of dienst waarvoor de afnemer betaalt. Die laatste mag daarvan iets vinden en hoeft ook niet getraind te zijn in beoordelen. Dit essentiële verschil tussen een bedrijfskundige en een bedrijfspsychologische visie op beoordelen heb ik lang geleden samen met collega Vloeberghs uitgewerkt bij het contrasteren van 360-graden feedback met assessment center feedback (Jansen & Vloeberghs, 1999). Een variant daarop geeft Figuur 3.

<i>Positie van beoordelaar</i>	<i>Object van de beoordeling</i>	<i>Relatie tussen beoordelaar en beoordeelde</i>	<i>Typische toepassing</i>
<i>Functie-expert</i>	competentie in functioneren	afstand	assessment center
<i>Klant</i>	mate van bevredigen van klantbehoeften	betrokkenheid (service verlenen)	klantsurvey

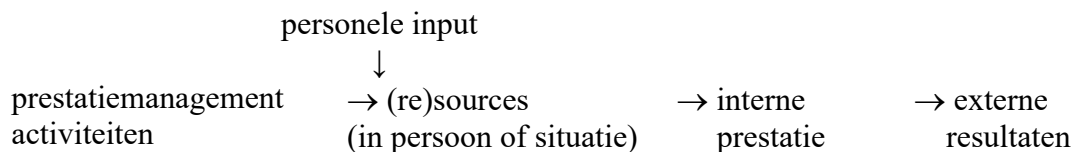
Figuur 3. Twee perspectieven bij de beoordeling van functioneren.

Zoals figuur 3 laat zien is bij het assessment center de beoordelaar een expert in de functie (vaak een hogere manager) die op afstand en getraind en vaak ook nog geregisseerd door een psycholoog, het gedrag van de kandidaat beoordeelt op expliciete criteria aan de hand van diens gedrag in gestandaardiseerde simulaties van kritische situaties uit het werk. Maar in geval van een klantsurvey vullen afnemers van het product of de dienst een vragenlijst in over hoe deze persoon zich vanuit hun ervaring heeft gedragen op een aantal criteria. Deze beoordelaars zijn niet getraind, hebben in werkinteractie gestaan met de beoordeelde en gaan bij hun beoordeling primair af op hoe zij tevreden zij zijn over het product of de dienst.

Tot nu toe hebben we de externe marktbijdrage sterk in het perspectief van de klant geplaatst en de interne prestatie vooral vanuit een psychologische invalshoek bezin. Maar figuur 3 laat zien dat ook in geval van intern functioneren het nuttig is een onderscheid te maken tussen een neutraal functioneringsperspectief en een uitdrukkelijk betrokken klantperspectief.

4. Bedrijfskundig beoordelen

Vanzelfsprekend staan werkprestatie en externe bijdrage met elkaar in verband. Vanuit de HRM-literatuur zijn veel zogenoemde prestatie-managementmodellen beschikbaar, vaak combinaties van denkschema's uit extern gericht strategisch management of intern gericht managementcontrol. Kernmerk is dat een keten van in- en externe factoren de overgang van interne functieprestatie naar externe marktbijdrage bepaalt. Ook worden meerdere soorten prestaties en meerdere typen externe markten onderscheiden. En tot slot wordt verondersteld dat observeerbare prestaties voortvloeien uit 'organisatie-intern kapitaal'. Voorbeelden van die bronnen of resources zijn intelligentie, motivatie of ervaring als het om de medewerker gaat, en een coachende manager of taakverrijking of ondersteunende collega's als het om de organisatie gaat. Dit leidt tot het model in figuur 4.



Figuur 4. *Keten van prestatiebesturing in een organisatie.*

Externe marktresultaten hebben niet alleen betrekking op de klantenmarkt met criteria als prijs en kwaliteit, of op de financiële markt met criteria als transparantie en profijt, of op de concurrentenmarkt met criteria als productiviteit en efficiency, maar omvat ook de ‘personele markt’ met criteria als gezondheid, tevredenheid en employability (De Vos, van der Heijden & Akkermans, 2020). Ik zie medewerkerstevredenheid als een autonoom doel van de onderneming. Niet omdat tevreden medewerkers bijdragen aan meer verkoop of betere kwaliteit (want dan worden zij behandeld als middel, d.w.z. als een te benutten resource) – hoewel daar op zichzelf niets tegen is -, maar vooral omdat tevredenheid (of bevlogenheid of betrokkenheid, d.w.z. positieve emoties i.h.a.) een eigenstandig en evenwaardig doel is. Laat dus wat winst liggen als je mensen daardoor meer tevreden zijn. Het actuele ideaal van ‘sustainable’ prestatie management betekent niet uitputten van resources maar ze zodanig benutten dat ze groeien.

In het vak Performance Management binnen de masterspecialisatie HRM, wordt studenten gevraagd bij bestaande organisaties systemen en praktijken van prestatie management, en vrijwel altijd daarmee van beoordelen, in kaart te brengen en te analyseren en daarover met het management in gesprek te gaan. Figuur 5 laat zien dat ze bij hun diagnose het prestatie model van rechts naar links, d.w.z. van buiten (markt) naar binnen (de organisatie) doorlopen.

Wie zijn onze ‘klanten’? Wat willen die van ons? Productiviteit (effectiviteit en efficiëntie)? Flexibiliteit (leren en innovatie)? Legitimiteit (maatschappelijke acceptatie)? Hoe kunnen we hen tevreden stellen? Wat zijn dus onze markten of resultaatgebieden?



In welke mate zijn deze externe organisatiedoelen gerealiseerd? Welke externe omstandigheden werkten daarbij belemmerend of bevorderend?



Op welke interne organisatiedoelen moeten we ons dus richten? Hoe zijn die verbonden aan interne personele prestaties?



Wat is de relatieve bijdrage van interne ‘harde’ of ‘zachte’ personeelsprestaties?



Welke prestatie management activiteiten of interventies zouden derhalve moeten worden toegepast? Hebben ze gewerkt als beoogd?



Als ze niet goed hebben gewerkt: aan welke persoonsfactoren of situatiefactoren lag dat?



Hoe zit het met de effectiviteit, efficiency, en acceptatie van de prestatie management-activiteiten of interventies?

Figuur 5. *Vragen voor diagnose van prestatie management*

Niet alleen kan dit schema gebruikt worden om de praktijk van beoordelen te beoordelen aan de hand van wetenschappelijke begrippen en modellen. Maar daarnaast maakt het aanschouwelijk dat een organisatie nogal wat invalshoeken heeft bij beoordeling, bijvoorbeeld: Hoezeer zijn klanten tevreden over de externe resultaten? In welke mate is de interne prestatie aan de maat? In welke mate zijn er resources gemaakt? In welke mate hebben managers of collega's daaraan bijgedragen? Tot op welke hoogte verhindert de wijze van organiseren, bijvoorbeeld de inrichting van taken of van processen van samenwerking, medewerkers om tot prestaties te komen? Hoe zit het met de kosten, de efficiëntie en de acceptatie van het beoordelingssysteem? En de praktijk? Hier is sprake van wat ik noem bedrijfskundig beoordelen omdat niet alleen de persoon wordt beoordeeld maar de totale keten waarin deze functioneert. Bovendien kan op elk beoordelingsmoment een zo veel mogelijk neutraal functioneringsperspectief of een uitdrukkelijk betrokken klantperspectief worden gekozen.

Dit schema dient ook als basis voor een beschouwing van de effectiviteit van beoordelen: bijvoorbeeld in welke mate draagt beoordelen bij aan de interne prestatie of aan de persoonsgebonden of situatiegebonden resources daarvoor, of aan resultaten in welke externe markten en tegen welke criteria? Bevordert de wijze van beoordelen bijvoorbeeld de attractiviteit van het bedrijf voor de relevante arbeidsmarkt, of medewerkerstevredenheid (zoals eerder gemeld een autonoom doel van de onderneming)? Het vele onderzoek naar de effecten van feedback geven concentreert zich vrijwel uitsluitend op wat ik heb aangeduid als de interne prestatie: in welke mate wordt die beter van feedback en welke feedbackcondities zijn daarop van invloed? Zo presenteren Gruman en Saks (2011) prestatie management als een belangrijke bron voor medewerker-engagement.

Wat betreft de impact van beoordelen op de prestatie: Nog steeds staat de conclusie uit de meta-analyse van Kluger en DeNisi (1996) dat in bijna 2/5 van de gevallen de prestatie na feedback verslechterde, zelfs als die positief was. Redenen zijn onder andere dat feedback geven het gevoel van extern gecontroleerd worden bevordert (het “*goed gedaan jochie*” – effect). Ook bij het volledig ontbreken van zo'n controle-sfeer, bijvoorbeeld bij relatieve beoordeling waarbij de beoordeling afhankelijk is van hoe collega's het deden, kan feedback leiden tot frustratie met als gevolg het dan maar opgeven (Iles & Judge, 2005). Dit alles geldt ook voor 360 graden feedback (zie bijvoorbeeld DeNisi & Kluger, 2000; Jansen & Vloeberghs, 1999). Er zijn in de loop verschillende modellen gepresenteerd om dit feedback-prestatie proces beter te begrijpen (onder andere, Keith & Frese, 2008, en Kinicki, Prussia, Wu & McKee-Ryan, 2004). Mooi en belangrijk allemaal want de interne prestatie is tenslotte een voorwaarde voor extern bijdrage, maar toch mis ik in dit alles beoordelen als proces van *integrale* managementcontrol (niet controle!), dus als een instrument voor prestatie sturing en beheersing gericht op een extern resultaat.

In zo'n bedrijfskundige benadering is het beslist niet zo dat beoordelingssysteem “*can be considered effective when its key stakeholders, i.e., appraisees, consider it useful*” (Ikramullah et al., 2016, p. 337). Beoordeelde vormen immers maar één van de afnemers. In hun toepassing van het Concurrerende Waarden model (Quinn & Rohrbaugh, 1983) op het ontwerp van een beoordelingssysteem, komt stakeholder-satisfactie alleen maar voor in een

Open System-visie en wordt de stakeholders bovendien weer intern ingevuld als ofwel beoordeelden ofwel manager-beoordelaars. In een, overigens erg mooie, studie naar de praktijk van beoordelen in een ziekenhuis wordt bijvoorbeeld de patiënt als beoordelaar niet genoemd (Redman et al., 2000). In hun overzicht van de laatste stand van zaken van onderzoek naar prestatie management, waarvan beoordelen volgen hen een ‘subset’ is (p. 2210), brengen Schleicher et al. (2018) in kaart welke onderdelen de effectiviteit bevorderen, bijvoorbeeld participatie van medewerkers of training van beoordelaars; - maar ook hier wordt over effectiviteit in de zin van externe markt bijdragen niets gezegd. In wat de auteurs aanduiden als een systeembenadering blijft deze ‘ultimate output’ beperkt tot bijvoorbeeld prestatiebeoordelingen en ontwikkelingsplannen. Hoewel de auteurs wel erop wijzen dat er de laatste jaren duidelijk meer onderzoek op organisatieniveau plaatsvindt, is van bedrijfskundig beoordelen dus nauwelijks sprake in deze, nota bene in het *Journal of Management* verschenen review. Wat we nodig hebben is een “*overall, holistic theory of PA system effectiveness*” (Ikramullah et al., 2016, p. 335). Een aanzet daartoe kan worden gevonden in de review van Levy en Williams (2004). Zij richtten zich expliciet op de bredere sociaal-organisatorische context van beoordelen. Voordeel van het door hen gepresenteerde ‘social context model’, met zowel distale als proximale proces- en structuurvariabelen is dat het ook kan dienen voor een veelomvattende evaluatie van beoordelen. In theorie is beoordelen dus van een geïsoleerde personele activiteit een *systeem* geworden. Maar ook zij constateren dat er in de praktijk sprake is van “*little systematic empirical work on the distal variables listed in Figure 1 other than a bit on culture, climate, and technology issues* (o.c., p. 885), en dat “*Appraisal Reactions is where there has been the most growth in the PA research since 1995 and also where practitioners see the most potential benefit*” (o.c., p. 899).

Ook wanneer je prestaties bepleit als eerlijkheid of bijdragen als grotere sociale cohesie of inclusiviteit of waardigheid in het werk (Bal, 2020), zul je moeten bedenken hoe die tot stand komen en moeten beoordelen in hoeverre je die realiseert. Je zult bijvoorbeeld inclusiviteit moeten meten en beoordelen aan de hand van een norm als waardigheid. Merk ook op dat dit pleidooi is gebaseerd op de *beoordeling* dat momenteel zulke prestaties niet worden gerealiseerd. Het aardige is dat expliciete inrichting van een prestatie managementsysteem de kans groter maakt dat dergelijk doelen worden gerealiseerd. Ook zul je moeten ingaan op de bedrijfskundige relatie tussen zulke prestaties en het leveren van waarde aan klanten, bijvoorbeeld goed brood of een adequate medische behandeling. Dit leidt tot nieuwe onderzoeksgebieden, bijvoorbeeld de mate waarin in bepaalde culturen wat wij zien als het verlenen van een service aan een klant, bijvoorbeeld als ober, als minderwaardig werk wordt ervaren.

Een dergelijke bedrijfskundige visie op beoordelen, waarbij de nadruk ligt op de externe markt bijdrage, is ook van belang bij het anders ‘*Erkennen en waarderen*’ van wetenschappers (Belleman & Flink, 2020). Merk op dat erkennen en waarderen een stuk sympathieker klinkt dan beoordelen. Evenwel: hoe dit er in de praktijk gaat uitzien “*is nog even onduidelijk*” (Belleman & Flink, 2020) maar wat ik begrijp is dat de universiteit ook andere taken als bijvoorbeeld leiderschap op waarde moet schatten. Onderzoekers moeten in een ‘narratief cv’ uitleggen “*waarom hun werk ertoe doet*”. Ik heb mijn twijfels over hoe zo’n narratief cv uitpakt in de beoordeling, maar het signaleert dat ook hier niet de interne prestatie maar de externe bijdrage grondslag wordt voor erkenning en waardering (d.w.z. voor beoordeling en beloning). Wel zal de prestatieketen van eigen interne functioneren naar extern marktsucces op meer moeten worden gebaseerd dan *een goed verhaal*.

Anders kijken naar bijdragen vraagt om een antwoord op de eerder gestelde strategische startvragen: *Wie zijn onze 'klanten'? Wat willen die van ons? Hoe kunnen we hen tevreden stellen? Wat zijn dus onze markten of resultaatgebieden?* En dus om een andere invulling van waartoe een universiteit dient. Wel is het zo dat universitair wetenschappers deels fundamenteel onderzoek doen en in die zin zelf hun markt maken, d.w.z. zelf innovatief voorop lopen en bepalen wat de klant zou moeten willen (net als Apple indertijd met de iPhone). In dat geval bestaat het werk eruit zelf een antwoord te geven op de vraag *“Wat moet ik hier doen om het goed te doen?”* – en dat geldt zowel voor het “*wat*” als voor het criterium waarom dat wat “*goed*” is. Ik heb dat persoonlijk ervaren toen ik hier in 1992 begon: niet alleen was mijn opdracht college geven en onderzoek doen, maar ook moest ik zelf inhoud en criteria daarvoor bedenken. Ik werd dus mijn eigen opdrachtgever, d.w.z. stakeholder en dus beoordelaar. En hetzelfde geldt voor leidinggeven. Voor mij ligt de waarde van *Erkennen en Waarderen* dat weer eens over dit soort zaken: strategische marktbijdragen en interne prestatieketens die daarmee verbonden zijn, wordt nagedacht – hopelijk zonder dat snelle oplossingen worden gezocht in onnavolgbare indicatorensets. Omdat in dat geval marktbijdragen op voorhand lastig te bepalen zijn en innovatie soms pas jaren later blijkt, zul je de beoordeling van dit type hoog-risico innovatie moeten baseren op ofwel interne prestaties waarvan je weet dat ze de kans op innovatie vergroten ofwel bijdragen in afgeleide markten als die van subsidieverleners of beroepsorganisaties (Balkin, Markman & Gomez-Mejia, 2000).

Concluderend: Ik heb dus mijn bedenkingen bij het afschaffen van het beoordelingsgesprek. Naast een psychologisch invalshoek, met bijvoorbeeld veel (naar mijn mening te veel) aandacht voor wat er allemaal kan misgaan door beoordelingsfouten, is er immers ook een bedrijfskundige invalshoek: Medewerkers krijgen betaald om te presteren en de afnemers van die prestatie mogen beoordelen of zij waar krijgen voor hun geld (en daar kan de organisatie ook van leren).

Als een concertbezoeker een pianorecital minder goed beoordeelt kan de pianist niet zeggen dat deze meer verstand heeft van pianospelen dan de luisteraar. Integendeel, de pianist zal moeten nagaan wat de luisteraar mist en wat hij daaraan kan doen. En als dat psychologisch pijn doet bij de pianist, moet dat maar. Negatieve feedback ontvangen vindt niemand leuk, ondanks alle vrolijke coachingsverhalen dat je er blij mee moet zijn en zelfs ervoor moet bedanken. Maar je zet je er maar overheen, tenslotte heb je zelf als pianist of hoogleraar ervoor gekozen je op de markt van luisteraars dan wel studenten of onderzoekers te begeven: *“performance is always evaluated in some manner, (...) too hard” is no excuse for industrial-organizational (I-O) psychology*” (Adler et al., 2016, p. 220). Die markt kan de directe afnemer zijn (bijvoorbeeld klanten in een winkel of studenten via onderwijsevaluaties) of als dat lastig is de ‘invisible hand’ van de manager die dergelijke feedback organiseert naar de medewerker. Beoordelen plaats ik dus in een businessmodel van transacties. Opvallend is dat uit een overzicht van ruim 12.000 onderzoeken op het gebied van HRM blijkt dat personeelsbeoordeling geplaatst wordt in het typisch bedrijfspsychologische cluster van ‘People Assessment’, d.w.z. persoonsdiagnostiek en veel minder is onderzocht in de context van ‘Strategic HRM’ (Markoulli et al., 2017).

Omdat managers steeds minder medewerkers feitelijk zien (dat was ook al voor corona het geval) dient de medewerker ervoor te zorgen dat deze beoordeelbaar is, bijvoorbeeld door prestaties bij te houden (zoals wij allen op de VU doen) volgens het schema: *Wat is tegen welk criterium door wie hoe beoordeeld en op welk nummer kan deze gebeld worden?* Maar ook: *Wie zijn de afnemers van wat ik doe, wie zit er dus op mij te wachten en wat precies*

verlangen die van mij? Beoordeelbaarheid betekent dat de medewerker zelf ervoor verantwoordelijk is dat de organisatie (de manager) in staat is diens bijdrage aan de organisatie te beoordelen. Maak je dus beoordeelbaar, ook in de zin dat jij zelf ervoor verantwoordelijk bent zodanige feedback te organiseren dat jij in staat bent jezelf te ontwikkelen. Dus niet alleen ‘craften’ van je werk maar ook van je beoordeling. In dit alles is de bedrijfskunde leidend en de bedrijfspsychologie volgend.

Of dit, mede in het licht van blijvend vaker thuis werken door corona, leidt tot nieuwe, minder strakke wijzen van control is nog de vraag. Het werk zelf in de zin van wat wordt geleverd aan welke klant op welke wijze en tegen welke criteria, verandert niet wezenlijk door ‘distantiation’ (Sewell & Taskin, 2015, p. 1511). Ook belangrijke effectiviteitscondities voor prestatiebeoordeling als invloed (medewerkers moet prestatie kunnen beïnvloeden), duidelijkheid (medewerker moet weten welke prestaties leiden tot welke externe bijdragen), en transparantie (bepaling en verdeling van beloning moet inzichtelijk zijn) blijven geldig. De studie van Sewell en Taskin (o.c.) laat zien wel dat er een tendens is dat managementcontrol wordt gebaseerd op *zichtbare* deelname aan ingeroosterde virtuele meetings en i.h.a. leidt tot de ervaren druk om thuis te allen tijde via bijvoorbeeld de computer *beschikbaar* te zijn (zodat men bijvoorbeeld een expliciete afspraak moet maken met de manager over zoiets als een lunchpauze met de eigen kinderen). Ofschoon er daarmee sprake is van een ‘*new arrangement of space and time*’ (o.c., p. 1524) gaat de manager in de beoordeling ook in dit geval weer af op aanwezigheid in de zin van virtuele presentie en actief bezig zijn, in plaats van op reële organisatiebijdragen. Traditionele ‘*workplace norms*’ (l.c.) van zichtbaarheid, er-zijn en beschikbaarheid blijven dus prevaleren. Ik verwacht dat de kans daarop zelfs toeneemt omdat er sprake is van een beperkte channel capacity bij virtuele communicatie en informele op-de-gang contacten wegvallen.

Eerder heb ik het bezwaar genoemd dat jaarlijks beoordelen op tevoren besproken normen in de huidige tijd van permanente dynamiek niet meer goed mogelijk zou zijn. De permanente dynamiek aan de klantkant zou daarom gespiegeld moeten worden in continu spreken over hoe het gaat. Er is evenwel weinig evidentie voor de psychologische of bedrijfskundige werkzaamheid hiervan (Gorbatov & Lane, 2018). Externe turbulentie beheersen met behulp van intern turbulentie klinkt ook niet echt (psycho-)logisch. In plaats van een eenmalige jaarlijkse beoordeling elk transactie beoordelen omdat dat meer ‘objectieve informatie’ zou leveren, negeert dat verwerken van informatie organisatie vereist.

Duidelijk is de belangrijke rol van de manager bij deze organisatie van het informele en toevallige. Deze dient niet slechts door te geven wat diverse afnemers van de medewerker vinden maar dit te integreren, er dus zelf iets van te vinden, bijvoorbeeld gelet op ontwikkeling. Zo moet een leidinggevende aan de universiteit ervoor durven kiezen weliswaar te spreken met een medewerkers over een minder onderwijsevaluatie maar die eventueel niet op te nemen in het officiële verslag. Een manager is geen doorgeefluik maar een filter, een ‘lemen laag’ in de goede zins des woords. Door bijvoorbeeld een ontwikkelingsgericht perspectief in te brengen wordt van een resultaat weer een prestatie gemaakt, d.w.z. een globale, vaak alleen maar evaluatieve marktfeedback wordt voor de medewerker weer beïnvloedbaar en verbeterbaar. Professioneel management blijft dus nodig.

Dit beroep op de bedrijfskundige context van beoordelen onderstreept iets wat eerder bij de bespreking van het toernooimechanisme aan de orde kwam. Het is nodig attent te zijn op de aard van het stochastisch proces waardoor de interne prestatie of, en met name, de externe bijdrage tot stand komt. Waarnemingen worden zowel bepaald door hoe het gesteld is in ‘de

natuur' als door hoe wij die natuur waarnemen, d.w.z. door het meetinstrument, de observatiemechaniek, de buis waardoor wij kijken, het net waarin wij de ruwe data vangen en tot waarnemingen maken. Naast een inhoudelijke theorie hebben wij dus ook altijd een meettheorie nodig (Roskam, 1989). En in koor met postmoderne antipositivisten zou men kunnen zeggen: omdat niets in zich natuurlijk is, – immers: waar eindigt het natuurlijk proces en waar begin het meetproces? -, is alles meettheorie in de zin dat wij 'onze data maken'. Ik vind dit een nogal slordige gedachte, maar dat is nu niet van belang. Duidelijk is goed nagedacht moet worden over hoe data tot stand komen. En die datatheorie reikt verder dan de psychologische, interne prestatie.

Ik ben het eens met Pulakos en O'Leary (2011) dat er zowel aandacht moet zijn voor het microniveau van de dagelijkse praktijk van beoordelen (en daar komt de psychologie van pas) als voor de bedrijfskundige context waarbinnen dat alles plaatsvindt. Hoewel zowel in een arbeidsorganisatie als in een business-school, het laatste behoort te domineren, is het inderdaad in de praktijk niet zo makkelijk organisatiedoelen te vertalen naar individuele te behalen resultaten die ook nog eens SMART ('specific, measurable, attainable, relevant, time-bound') zijn. Met enige graagte heeft de bedrijfspsychologie zich genesteld in deze 'black box' tussen organisatie (bijvoorbeeld officiële HR praktijken) en realisatie in en door mensen. Met als gevolg bijvoorbeeld regelmatig terugkerende pleidooien om bijvoorbeeld de communicatie tussen manager en medewerker te verbeteren. Mijn pleidooi is voor wat meer *out-of-the (black) box* werk, d.w.z. meer bedrijfskundig. Kies dus 'de weg omhoog', zowel in de totale prestatieketen als in de bekende hiërarchie van micro (medewerker) via meso (groep) naar macro (organisatie).

5 Slotwoord

Het voorgaande betoog met steeds wisselende aandacht voor het microproces waarin een oordeel tot stand komt en voor verbinding omhoog met organisatie en strategie, komt overeen met mijn loopbaan. In 1979 afgestudeerd in de mathematische psychologie ben ik direct in de praktijk gaan werken bij onder andere de Rijks Psychologische Dienst en de PTT (later KPN). Maar daarnaast ben ik blijven publiceren op het gebied van psychometrie en heb ik een proefschrift geschreven over de toepassing van item-response theorie modellen op meting door middel van attitudevragenlijsten. In mijn praktijkwerk heb ik procedures ontworpen, uitgevoerd en geëvalueerd voor onder andere selectie en beoordeling. Ook dat heeft tot een aantal publicaties (onder andere praktijkboeken) geleid. Kern van mijn werk is altijd het *beoordelen van mensen* geweest, eerst in de vorm van assessment centers, later ook via methoden van personeelsbeoordeling. Bij beoordelen kwam later ook ontwikkeling en verandering van mensen, zowel op microniveau als macroniveau.

Vanaf maart 1992 heb ik gewerkt als hoogleraar Bedrijfspsychologie bij de Vrij Universiteit. Die 12 jaar praktijk ervoor had ik niet willen missen, maar ze waren natuurlijk niet goed voor mijn H-index. Van meet af aan heb ik ook veel aandacht besteed aan onderwijs in de methoden en technieken van onderzoek, vooral ook voor wat betreft de eigen aard van bedrijfskundig onderzoek. Daarover ging ook mijn entree uit 1992. Dit alles in samenwerking met de oud-collega's Mandy van de Velde (Universiteit van Utrecht) en Josje Dijkers (Hogeschool Utrecht). De laatste jaren hebben we ons specifiek gericht op onderzoek in het HBO, in samenwerking met Lisette Munneke (Hogeschool Utrecht).

Ik zeg altijd: als je "netjes werkt" dan levert onderzoek altijd iets op. Netjes houdt in: goede theorie, weldoordachte begrippen, deugdelijk design, aandacht voor meting en correcte

statistische analyses. Het leuke is dat studenten en promovendi merken dat dit inderdaad klopt! Ze gloeien als ze ‘iets’ vinden in de data: structuren, patronen, orde in de chaos. Niet alleen leidt netjes werken tot ontdekken van waarheid in de chaos, naar netjes werken houdt ons ook overeind in het leven: Fatsoenlijk, ordelijk, gewetensvol, met respect, aanspreekbaar, verantwoordelijk, welgemanierd, waardig. De huidige uitzonderlijke maatschappelijke situatie bewijst dit nog maar eens.

Ik heb het altijd als een groot geluk beschouwd dat ik competente en aardige mensen om mij heen had en nog heb met wie gezamenlijk heb kunnen bouwen aan een naar mijn mening uitstekende onderzoeks- en onderwijsgroep op het gebied van bedrijfspsychologie en HRM. Speciaal noem ik Svetlana Khapova, eerst medewerker, later hoogleraar en vervolgens ‘mijn baas’. Dank ook aan mijn promovendi; ik heb veel van hen geleerd en daarom ga ik graag even nog door met promotiebegeleiding. Dank ook aan de collega’s van de afdeling M&O, en speciaal de HRM-groep. Tot slot dank aan alle anderen van de School of Business & Economics, en ook VU-breed met wie ik in allerlei verbanden (onder andere i.h.k.v. het periodieke VU-brede werkbelevingsonderzoek) heb mogen samenwerken. Ook ben ik collega prof. Gilbert Cijntje van de ICUC University op Curaçao dankbaar voor onze lange en heel plezierige samenwerking. Graag noem ik ook collega Jaap Paauwe, emeritus hoogleraar van Tilburg University, en alle (oud-) medebestuurders van het Dutch HRM Network.

“It seems to me that there is a special domain of very profound and strong human needs related to scientific research, which are similar in many ways to aesthetic and perhaps religious needs. And it also seems to me that the satisfaction of these needs should be considered an important task of research. Hence, I believe, the question of the value of any research cannot be adequately answered without taking into account the intellectual satisfaction which the results of that research bring to those who understand it and care for it. It may be unpopular and out-of-date to say-but I do not think that a scientific result which gives us a better understanding of the world and makes it more harmonious in our eyes should be held in lower esteem than, say, an invention which reduces the cost of paving roads, or improves household plumbing”

(Tarski, A. (1944). The semantic conception of truth: and the foundations of semantics. *Philosophy and Phenomenological Research*, 4 (no. 3), p. 370)

Tot slot *dank dank dank!* aan Marie-Thérèse, Geerten, Maarten, Wieneke en natuurlijk Willem. Freud schijnt gezegd te hebben dat het in een mensenleven maar om twee zaken gaat: ‘*Lieben und Arbeiten*’ (Elms, 2005). Hier zat hij ernaast. Wat uiteindelijk telt is ‘*Geliebt werden*’.

Ik heb gezegd.

Referenties

- Adler, S., Campion, M., Colquitt, A., Grubb, A., Murphy, K., Ollander-Krane, R., & Pulakos, E. D. (2016). Getting rid of performance ratings: Genius or folly? A debate. *Industrial and Organizational Psychology*, 9(2), 219-252.
- Aguinis H, & O'Boyle E. (2014). Star performers in twenty-first-century organizations. *Personnel Psychology*, 67, 313–350
- Aguinis, H., O'Boyle, E., Gonzalez-Mulé, E., & Joo, H. (2016). Cumulative advantage: Conductors and insulators of heavy-tailed productivity distributions and productivity stars. *Personnel Psychology*, 69), 3-66.
- Balkin, D.B., Markman, G.D., Gomez-Mejia, L.R. (2000). Is CEO pay in high technology firms related to innovation? *Academy of Management Journal*, 43, 1118-1129.
- Belleman, B., & Flink, E. (2020). Hoe gaan we wetenschappers straks 'erkennen en waarderen'? *Ad Valvas*, 24 september 2020.
- Bal, P. M. (2020). Why we should stop measuring performance and well-being. *Zeitschrift für Arbeits-und Organisationspsychologie*, 64, 196-200.
- Chin-A-Fo, H., & T. Jaeger (2020). De bestsellercode. *NRC*, vrijdag 6 maart 2020, p. C10.
- DeNisi, A. S., & Kluger, A. N. (2000). Feedback effectiveness: Can 360-degree appraisals be improved?. *Academy of Management Perspectives*, 14, 129-139.
- De Vos, A., Van der Heijden, B. I., & Akkermans, J. (2020). Sustainable careers: Towards a conceptual model. *Journal of Vocational Behavior*, 117, 103196.
- Elms, A. C. (2001). Apocryphal Freud: Sigmund Freud's most famous "quotations" and their actual sources. *Annual of Psychoanalysis*, 2001, 29, 83-104.
- Gorbatov, S., & Lane, A. (2018). Is HR missing the point on performance feedback? *MIT Sloan Management Review*, 59(4), 65-71.
- Gruman, J. A., & Saks, A. M. (2011). Performance management and employee engagement. *Human resource management review*, 21), 123-136.
- Ikramullah, M., Van Prooijen, J. W., Iqbal, M. Z., & Ul-Hassan, F. S. (2016). Effectiveness of performance appraisal. *Personnel Review*, 45, 334 – 352.
- Iles, R., & Judge, T.A. (2005). Goal regulation across time: The effects of feedback and affect. *Journal of Applied Psychology*, 90 (No. 3), 453-467.
- Jansen, P.G.W. (2002). *Organisatie en mensen. Inleiding in de bedrijfspsychologie voor economen en bedrijfskundigen* (3e herziene druk). Baarn: Nelissen.

Jansen, P.G.W., & Eijkelenburg, S. (1998). Onderzoek en accountancy: normatieve gidsen en inductieve speurders. In: A. Bac & K. Wildschut (Red.). *Reflecties op Bak. Liber Amicorum aangeboden bij het emeritaat van Prof. drs. G.G.M. Bak RA*, Tilburg University Press, 113-124.

Jansen, P.G.W., Stoop, L.A.M. (2001). The Dynamics of Assessment Center Validity: Results of a Seven Year Study. *Journal of Applied Psychology*, 86 (No.4, Aug), 741-753.

Jansen, P.G.W., & Vinkenburg, C.J. (2006). Predicting management career success from assessment center data: a longitudinal study. *Journal of Vocational Behavior*, 68, 253-266.

Jansen, P.G.W., & Vloeberghs, D.F.J. (1999). Multi-rater feedback methods: personal and organizational implications. *Journal of Managerial Psychology*, 14 (No.6), 455-476.

Joo, H., Aguinis, H., & Bradley, K.J. (2017) Not all nonnormal distributions are created equal: Improved theoretical and measurement precision. *Journal of Applied Psychology*, 102, 1022–1053.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology*, 93, 59-69.

Kinicki, A.J., Prussia, G., Wu, B., & McKee-Ryan, F.M. (2004). A covariance structure analysis of employees' response to performance feedback. *Journal of Applied Psychology*, 89, 1037-1069.

Kluger, A.N., & DeNisi, A. (1996). The effects of feedback interventions on performance. A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119 (No.2), 254-284.

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of management*, 30(6), 881-905.

Markoulli, M. P., Lee, C. I., Byington, E., & Felps, W. A. (2017). Mapping Human Resource Management: Reviewing the field and charting future directions. *Human Resource Management Review*, 27), 367-396.

Merkelbach, H. (2020). Verloot die onderzoekssubsidies. *NRC*, zaterdag 4 januari 2020, p. W2.

O'Boyle E.H., & Aguinis H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology*, 65, 79–119.

Pulakos, E. D., & O'Leary, R. S. (2011). Why is performance management broken? *Industrial and Organizational Psychology*, 4(2), 146-164.

Quinn, R., & Rohrbaugh, J. (1983). A spatial model of effectiveness criteria: towards a competing values approach to organizational analysis. *Management Science*, 29, 363-377

- Redman, T., Snape, E., Thompson, D., & Yan, F. K. C. (2000). Performance appraisal in an NHS hospital. *Human Resource Management Journal*, 10(1), 48 - 62.
- Roskam, E. E. (1989). Operationalization, a superfluous concept. *Quality and Quantity*, 23 (3), 237-275.
- Searle, J.R. (1998). *Mind, language and society. Philosophy in the real world*. Basic Books.
- Sewell, G., & Taskin, L. (2015). Out of sight, out of mind in a new world of work? Autonomy, control, and spatiotemporal scaling in telework. *Organization Studies*, 36(11), 1507-1529.
- Schleicher, D. J., Baumann, H. M., Sullivan, D. W., Levy, P. E., Hargrove, D. C., & Barros-Rivera, B. A. (2018). Putting the system into performance management systems: A review and agenda for performance management research. *Journal of Management*, 44(6), 2209-2245.
- Tarski, Alfred. (1944). The semantic conception of truth: and the foundations of semantics. *Philosophy and Phenomenological Research*, 4 (no. 3), 341-376.
- Vinkenburg, C.J., Jansen, P.G.W., Dries, N., & Pepermans, R. (2014). Arena: a critical conceptual framework of top management selection. *Group & Organization Management*, 39 (No.1), 33-68.

KORT CV Paul Jansen

Prof. dr. Paul G.W. Jansen (1954) was van 1 maart 1992 tot 29 maart 2021 hoogleraar Bedrijfspsychologie, School of Business and Economics, Vrije Universiteit Amsterdam. Studeerde (*cum laude*) Mathematische Psychologie aan de Katholieke Universiteit Nijmegen. In de periode 1979 – 1992 was hij werkzaam op het Ministerie van Binnenlandse zaken, bij de Rijks Psychologische Dienst en bij PTT / KPN. Van 1995-2001 was hij in deeltijd directeur van het *'Limperg Instituut': Interuniversitair Onderzoeksinstituut voor Accountancy*. Voorzitter (1997-2001) van de *'Nederlandse Stichting voor Managementontwikkeling' NFMD*, bestuurslid van het *Dutch HRM Network* (samenwerkingsverband van Nederlandse onderzoekers op het gebied van HRM). Van 2011-2017 programmadirecteur van de Masteropleiding Bedrijfskunde van de faculteit der Economische Wetenschappen en Bedrijfskunde, Vrije Universiteit Amsterdam. Sedert 2015 deeltijdaanstelling op de leerstoel *Bedrijfspsychologie & Onderzoeksmethoden* in de Center voor Management en Leadership van de *IC University of the Caribbean* te Curaçao. Adviseur van verschillende bureaus op het gebied van werving, selectie, assessment en prestatie management.

Onderwijs, onderzoek en advisering op het gebied van management development, loopbanen, assessment (bijv. assessment centers en 360-graden feedback) en prestatie management (bijv. beoordelen, belonen en beheersing van de personele prestatie- en ontwikkelingscyclus).

Paul Jansen was promotor bij 47 dissertaties. Hij publiceerde, deels in samenwerking met collega's, 76 artikels in internationale toptijdschriften, 22 boekhoofdstukken en 20 boeken over onder meer bedrijfspsychologie, werving en selectie, personeelsbeoordeling, managementontwikkeling, veranderingsmanagement en methoden en technieken van praktijkgericht onderzoek.

Voor zijn wetenschappelijke werk kreeg Paul Jansen de Dutch HRM Network award 2013 uitgereikt tijdens de 8^e International Biannual Conference “*H' versus 'R' in HRM*” te Leuven, (13-15 November 2013).

Zie verder: <https://research.vu.nl/en/persons/pgw-jansen>.