

Personeelsbeoordeling

Effect van extern gegeven feedback op prestatie en emotionele beleving van de individuele werknemer

Paul Jansen

SAMENVATTING Bij personeelsbeoordeling geeft in het algemeen de directe manager feedback aan een individuele medewerker over diens werk. Onderzoek wijst uit dat feedback niet altijd positieve gevolgen heeft voor werkprestatie en werkbeleving. Er wordt een overzicht gegeven van onder welke condities feedback werkt, bijvoorbeeld hoe er wordt omgegaan met fouten. Zo blijkt erg specifieke feedback niet bevorderlijk voor leren en blijken managers in het algemeen minder last te hebben van beoordelaarsfouten. Iedere feedbackgever blijkt op een eigen, onder andere met de werkpositie samenhangende wijze te beoordelen. Vandaar dat steeds vaker gewerkt wordt met meerdere beoordelaars; dit is zelfs de kern van de *multi-rater* feedbackmethode. Tot besluit worden daarom verschillende beoordelingsmethoden (reguliere beoordeling, multi-rater feedback, assessment center) systematisch met elkaar vergeleken.

RELEVANTIE VOOR DE PRAKTIJK Expliciete feedback geven in de vorm van individuele personeelsbeoordelingen door managers is een kernactiviteit van prestatie management in organisaties. Er wordt een overzicht gegeven van wat wel en wat niet te doen bij het organiseren van individuele feedback.

1 Inleiding

Een belangrijk oogmerk van prestatie management is organiseren van feedback naar de medewerker. Dit gebeurt in het algemeen door middel van personeelsbeoordeling: dat wil zeggen door het doen van uitspraken over een persoon ('beoordeelde') door een of meerdere andere personen ('beoordelaars'). Deze uitspraken zijn:

- **Expliciet:** wat gezegd en opgeschreven is, kan niet meer ongedaan worden gemaakt.
- **Evaluatief:** de uitspraak omvat beschrijving en evaluatie. Beoordelen betekent gedrag voorzien van een weging op grond van een norm.
- **Sociaal:** de uitspraak wordt niet privé gedaan maar in de context van een 'publiek', bestaande uit ten minste medewerker(s) en collega-managers.
- **Consequentieel:** de uitspraak wordt gevolgd door positieve en/of negatieve acties van medewerker of manager.

Deze kenmerken maken dat feedback geven door middel van personeelsbeoordeling niet de meest gemakkelijk uit te voeren personele activiteit is. Beoordeling is 'given by someone who does not want to give it to someone who does not want to get it' (Bowman, 1999, p. 557). Van belang is dat het hierbij gaat om door personen gegeven, 'externe' feedback. Tegenover deze feedback die als het ware van buiten komt, staat 'interne' feedback die door de medewerker zelf wordt verzameld op basis van hoe hij/zij het werk doet. Daarbij komt dan geen andere persoon te pas. Als iemand bijvoorbeeld een tekst schrijft, corrigeert men zichzelf voortdurend op basis van teruglezen. Dit artikel richt zich uitsluitend op externe feedback.

Een belangrijk onderdeel van prestatie management is het geven van feedback aan individuele medewerkers door middel van beoordelingen. Het uitgangspunt daarbij is dat mensen leren van feedback in de zin dat de prestatie zal verbeteren na de feedback. Dit is de reden waarom medewerkers in organisaties periodiek worden beoordeeld. Het beleid is dat medewerkers behoefte hebben aan en dus recht hebben op feedback, want alleen dan kunnen zij hun werkgedrag verbeteren. Maar is dat wel zo? Onder welke condities werkt feedback? Is bijvoorbeeld specifieke feedback altijd beter dan algemene feedback, of maakt het wat uit met welk oogmerk de feedback wordt gegeven? En wat doen personen als zij door andere personen worden gewezen op fouten in hun gedrag? Hoe gaan medewerkers in het algemeen om met fouten, dat wil zeggen onder welke condities wordt er geleerd van fouten? Wordt er bijvoorbeeld meer geleerd als de feedback specifiek is, in plaats van vaag en algemeen? In paragraaf 3 wordt nagegaan welk antwoord de onderzoeksliteratuur geeft op deze vragen. Daarbij gaat speciaal de aandacht uit naar de rol van de externe feedbackgever, dat wil zeggen de 'beoordelaar'. Is het bijvoorbeeld zo dat feedback door meerdere beoordelaars meer betrouwbaar en valide is? Een in de praktijk veel gebruikt feedbackinstrument, waarin sprake is van meerdere beoordelaars, is de *360-gradenmethode*. In paragraaf 4 wordt aan de hand van de literatuur besproken wat we weten van deze 'multi-rater'

feedbackmethode. Daarbij wordt een vergelijking gemaakt met twee andere veelgebruikte feedbackmethoden, het assessment center en de traditionele personeelsbeoordeling. In de afsluitende paragraaf 5 trekken we enige conclusies.

2 Feedback

2.1 Feedback geven

Externe feedback in de zin van personeelsbeoordeling door bijvoorbeeld een manager kan worden gegeven op:

- **Dimensionele criteria:** voor de persoon typerende eigenschappen als flexibiliteit of sociabiliteit. Persoonsdimensies zijn min of meer stabiele kenmerken die terugkeren over een groot bereik van werksituaties heen. Dimensies kunnen worden gezien als de input van een persoon in het werk.
- **Situationele criteria:** vaardigheden of competenties als plannen, onderhandelen of controleren, die verbonden zijn met de feitelijke werksituatie. Het gaat om situatiespecifieke, in tijd en plaats beperkte activiteiten (vandaar het gebruik van werkwoorden) met begin- en eindmoment. Situationele criteria verwijzen naar *throughput*, naar hoe het werk feitelijk wordt gedaan, naar de wijze waarop een werkopdracht met behulp van persoonlijke input wordt omgezet in output.
- **Resultaatcriteria:** individuele werkoutput als bijvoorbeeld kwaliteit en kwantiteit. In dit geval gaat het om het effect van wat de persoon doet in zijn werk.

Feedback op situatiegebonden competenties kan zonder al te veel moeite worden gegeven door iemand die de taak in kwestie goed beheerst. Daarentegen vereist dimensionele, persoonsfeedback waarin een uitspraak wordt gedaan over hoe iemand in het algemeen is in zijn werk, speciale scholing. Taakgebonden situationele feedback doet het beter dan persoonsfeedback ten aanzien van informatierijkdom voor medewerker en organisatie, acceptatie door de medewerker, overeenstemming tussen verschillende beoordelaars en bevordering van leren en ontwikkelen. Bovendien is het aantal beoordelingsfouten, zoals te weinig onderscheidend oordelen, lager (zie paragraaf 3.4). Nochtans blijft dimensionele, persoonsgerichte feedback noodzakelijk indien uitspraken moeten worden gedaan die uitgaan boven een concrete situatie. Dat is bijvoorbeeld het geval als op grond van het werkgedrag van een stafmedewerker een voorspelling moet worden gedaan over de succeskans als lijnmanager. Bij een dergelijke ‘sprongmutatie’ moet worden geabstraheerd van concrete situatiegebonden vaardigheden (‘Hij/zij doet ...’) naar meer abstracte, aan de persoon gebonden eigenschappen (‘Hij/zij is ...’).

2.2 Feedback organiseren

Omdat feedback geven in de zin van personeelsbeoordeling neerkomt op het doen van uitspraken over personen, kunnen de volgende vormen van feedback geven onderscheiden worden:

- expliciet door andere persoon in een hiërarchische context (bijvoorbeeld manager);
- expliciet door andere persoon in een niet-hiërarchische context (bijvoorbeeld collega);
- expliciet door meerdere andere personen (bijvoorbeeld manager, collega, medewerker en klant).

In het algemeen zijn in het economisch verkeer twee vragen leidend voor feedbackgevers (Groot en Jansen, 2010):

1. Kreeg ik wat ik verlangde? Werd ik behandeld zoals ik had verwacht?
2. Op basis van (1): ben ik blij of tevreden?

Dus werkfeedback is altijd zowel informatieel (1) als ook emotioneel (2):

1. Beschrijvende gedragsfeedback: ‘U deed ...’
2. Evaluatieve gevoelsfeedback: ‘Ik ben zeer tevreden/blij ... over wat u deed/leverde ...’

In een werkrelatie wordt feedback primair gegeven vanuit het gezichtspunt van wat de feedbackgever in de rol van bijvoorbeeld klant, manager, medewerkers of collega wil hebben, in plaats van op grond van wat behulpzaam is voor de feedbackontvanger met het oog op prestatieverbetering. Feedback organiseren met het oog op prestatieverbetering richt zich daarom op de vragen:

1. Wat deed jij?
2. Welk (ook emotioneel) effect had dat op je rolpartner?
3. Wat te doen opdat de rolpartner de volgende keer wel tevreden is?

3 Het effect van feedback

3.1 Positieve en negatieve weerslag van feedback op prestatie

Algemeen wordt aangenomen dat mensen leren van feedback. Maar de meta-analyse door Kluger en DeNisi (1996) laat zien dat in 38% van de onderzochte studies het werkgedrag na feedback slechter was dan voor feedback. Feedback werd in dit onderzoek gedefinieerd als: ‘actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one’s task performance’ (Kluger en deNisi, 1996, p. 255). Feedback werd beperkt tot situationele criteria en resultaatcriteria en was zowel beschrijvend (kennis van resultaten) als ook beoordelend (evaluatie volgens een norm). Verklaringen voor het negatieve effect van feedback op werkgedrag zijn:

- Mensen reageren positief op feedback die hun positieve zelfbeeld bevestigt en negatief op negatieve feedback (Woo, Sims, Rupp en Gibbons, 2008).
- Feedback (ook positieve) kan het gevoel van ‘externe controle’ bevestigen, wat de (intrinsieke) werkmotivatie verlaagt. Personen met een hoge externe controle hebben het gevoel dat de dingen in het leven hen overkomen en dat zij daarom weinig tot geen invloed hebben op hun werkprestatie. Als gevolg daarvan is er weinig reden zich extra in te spannen.
- Feedback kan de aandacht verschuiven van concrete taakprocessen naar ‘metataak’ emotionele processen betrek-

king hebbend op het omgaan met schaamte of angst, of 'metataak' werkstrategieën gericht op prestatieverbetering. De persoon gaat dan piekeren over of hij/zij het wel goed doet en twijfelen aan zichzelf in plaats van werken aan verbetering. In beide gevallen zal de prestatie (initieel) dalen: in het eerste geval omdat de persoon afgeleid is, in het tweede omdat de persoon van een prestatiehouding overgaat op een leerhouding (Dweck en Legget, 1988) waarin verschillende taakaanpakken worden geprobeerd en minder moeite wordt gedaan fouten te vermijden.

- Herhaalde negatieve feedback kan leiden tot 'vaandelvlucht' in de zin dat de medewerker er geheel van afziet de norm nog te willen halen.

Iles en Judge (2005) onderzochten de invloed van feedback op positieve emoties zoals vreugde en trots en negatieve emoties zoals verdriet en woede, en op de prestatie. Zij vonden dat positieve feedback leidde tot positieve emoties; die resulteerden weer in het stellen van hogere doelen. Als er een klein verschil was tussen gesteld en behaald doel leidde de negatieve feedback weliswaar tot negatieve emoties, maar bleek de prestatie de volgende keer desondanks te stijgen. Bij grote discrepantie werd negatieve feedback echter gevolgd door neerwaartse doelaanpassing. De verwachting dat men in staat zal zijn het gestelde doel de volgende keer wel te halen, speelt dus een belangrijke rol bij het effect van negatieve feedback op prestatie. De auteurs merken op dat dit pleit tegen het gebruik van een gedwongen, meestal normaalverdeling bij beoordelingen. In dat geval krijgt een tevoren vastgesteld percentage van de medewerkers een negatieve beoordeling en zal dat deel als gevolg daarvan, ook als het absolute verschil in prestatie met positief beoordeelde collega's klein is, negatieve emoties ervaren. Bovendien leidt gedwongen verdeling van beoordelingsuitspraken tot verlaging van de gemiddelde output omdat men niet zijn uiterste best zal doen, maar het net even iets beter zal gaan doen dan een collega.

3.2 Leren van fouten

Het voorgaande suggereert dat het effect van feedback op gedragsverandering wordt bepaald door hoe de persoon in kwestie de feedback zowel verstandelijk opvat als ook emotioneel verwerkt. Auteurs als Kluger en DeNisi (1996), Keith en Frese (2008) en Kinicki, Prussia, Wu en McKee-Ryan (2004) hebben modellen voorgesteld van het feedback → prestatieproces. Zo opperen Kinicki et al. (2004) dat de geloofwaardigheid van de feedbackbron, dus van de 'klant' van de medewerker, de reactie op feedback en daarmee de uiteindelijke prestatieverbetering bepaalt (zie daarover verder paragraaf 4). Een belangrijk proceskenmerk van het succes van feedback is hoe wordt omgegaan met fouten. Dienen fouten als triggers voor exploratie en leren, of als signalen van wanprestatie die zal worden bestraft?

In de zogenoemde 'fouten-managementtheorie' worden fouten gezien als een natuurlijk gevolg van leren en daarom positief beoordeeld (Keith en Frese, 2008). Fouten horen bij een leeroriëntatie op werkgedrag, het zijn geenszins storende factoren die afleiden van de taak maar een wezenlijk onderdeel van actieve exploratie. De theorie achter deze benadering is dat door het maken van fouten en het op die wijze verkrijgen van feedback op verschillende gedragingen, medewerkers in staat zijn een 'actiegeoriënteerd', situationeel mentaal model van de specifieke werktak (zoals gebruik van een gecompliceerd apparaat) te maken. Door fouten te maken, krijgt men een gedragsmatig overzicht dat niet neerkomt op een abstracte beschrijving maar is opgebouwd uit actie-reactiepatronen. Zo hebben ervaren chauffeurs niet een mentaal overzicht van autobesturen gekregen door bestudering van een werktekening van de motor en een tabel met effecten van stuurwieloperaties, maar door ervaring in verschillende rijssituaties.

Het voorgaande impliceert voor prestatie management, en specifiek voor personeelsbeoordeling, dat feedback zo moet worden georganiseerd dat medewerkers in staat zijn actiegeoriënteerde, mentale modellen van kritische werksituaties te maken. Actieve exploratie, foutenmanagement en feedbackzoekend gedrag zijn wezenlijke bestanddelen van deze benadering van prestatie management (Keith en Frese, 2008). Mensen zoeken echter feedback om verschillende redenen (Dweck en Legget, 1988):

- **Zelfverbetering:** betere zelfinschatting maken en daardoor zowel de onzekerheid over 'Hoe doe ik het?' verminderen als ook meer effectief worden in concreet taakgedrag als het voeren van een verkoopgesprek (taakcompetitie).
- **Zelfverheffing:** het zelfbeeld vergroten door nieuwe informatie te zoeken die het al positieve zelfbeeld bevestigt ten opzichte van anderen (persoonscompetitie).

Negatieve feedback heeft bij zelfverbetering een ander effect dan bij zelfverheffing. In het eerste geval is de medewerker in competitie met een taak, in het tweede met anderen. Prestatieoriëntatie begint bij zelfverbetering maar kan, afhankelijk van de wijze waarop de feedback is georganiseerd (bijvoorbeeld collectieve feedback en leren in groepsverband), uitlopen op zelfverheffing. Vooral wanneer collega's fungeren als rolmodel kan feedback aanleiding geven tot persoonscompetitiviteit vanuit drang tot zelfverheffing.

3.3 Feedbackspecificiteit

Het is inmiddels genoegzaam bekend dat feedback tijdig, specifiek (dat wil zeggen geformuleerd in termen van concreet gedrag), en constructief moet zijn; alleen in dat geval is er sprake van een omgeving die 'feedback-rich' is (Kinicki et al., 2004). In lijn met het eerder gemaakte onderscheid tussen specifiek (situationeel) versus algemeen (trans-

situationeel, dimensioneel), werken we in deze bijdrage alleen feedbackspecificiteit uit. Feedbackspecificiteit verwijst naar de informatiedichtheid van de (meestal verbale) beoordelingsuitspraak. Specifieke, feitelijke, 'objectieve' feedback die correspondeert met het prestatiedoel (dat dus ook specifiek moet zijn gesteld) leidt tot betere toekomstige prestaties in relatie tot dat doel (Goodman, Wood en Hendrickx, 2004). Specifieke feedback die de aandacht richt op concreet taakgedrag leidt tot directe prestatieverbetering. Dit is bekend uit een lange historie van experimenten en theorievorming in conditioneringsonderzoek (Rietdijk, 2009). Maar specificiteit heeft ook nadelen.

Goodman, Wood en Hendrickx (2004) vergeleken het effect van specifieke feedback met dat van meer algemene feedback op zowel directe prestatieverbetering als ook langetermijnleren en generalisatie naar andere taken. Zij vonden:

- specifieke feedback is gunstig voor de directe prestatie maar ongunstig voor leren op de lange termijn; en
- meer algemene feedback is slecht voor de directe prestatie maar bevordert langetermijnleren en generaliseren naar andere maar wel vergelijkbare taken.

Het bleek dat specifieke feedback een prestatieoriëntatie oproept die schadelijk is voor leren. In overeenstemming met de conditioneringstheorie leidt specifieke, positieve feedback tot de neiging het oorspronkelijke gedrag in kwestie te herhalen. Vergroten van feedbackspecificiteit draagt dus wel bij aan de initiële prestatie maar ontmoedigt exploratie en leren op de langere termijn. Een voordeel van meer algemene feedback is dat medewerkers, doordat ze vaker fouten zullen maken, gestimuleerd worden op zoek te gaan naar alternatieve acties teneinde de juiste reactie te vinden en op die wijze strategieën ontwikkelen om fouten in de toekomst te vermijden (Goodman en Wood, 2004). Wat meer algemene feedback stimuleert dus een leerhouding. Specifieke feedback stimuleert oefenen van dezelfde taak, meer algemene feedback stimuleert modificatie van de taak.

3.4 De beoordelaar (feedbackgever)

Beoordelaarshalo

Onderzoek (bijvoorbeeld Viswesvaran, Schmidt en Ones, 2002) laat zien dat beoordelingen van werkprestatie een behoorlijke 'halo'-fout bevatten: er wordt te weinig onderscheid gemaakt tussen inhoudelijk sterk verschillende beoordelingscriteria als interpersoonlijke competentie, administratieve competentie, productiviteit, functiekennis en leiderschap, zodat de beoordeling uiteindelijk één enkel 'good guy-bad guy'-criterium representeert. Blijkbaar zijn beoordelaars weinig onderscheidend in hun beoordelingen. De meta-analyse van Viswesvaran et al. (2002) laat verder zien dat halo bij managers als beoordelaars twee keer kleiner is dan halo bij collega's als beoordelaars.

Halo is niet een gebrek aan betrouwbaarheid maar een probleem van validiteit. Een meting is valide als het meetinstrument (bijvoorbeeld een intelligentietest) inderdaad het 'begrip-als-bedoeld' (dus intelligentie in plaats van bijvoorbeeld stressbestendigheid) meet. Een meting is betrouwbaar als bijvoorbeeld een dag later dezelfde meting wordt verkregen. Betrouwbaarheid verwijst naar de herhaalbaarheid of systematiek van de beoordeling, validiteit verwijst naar de inhoud van de beoordeling. Beoordelingen kunnen systematisch zijn (bijvoorbeeld over tijd of tussen verschillende beoordelaars) maar nog steeds niet het begrip-als-bedoeld meten. Volgens Viswesvaran et al. (2002) staat halo voor een psychologisch proces waarin de beoordelaar zich, alvorens beoordelingen toe te kennen aan de afzonderlijke criteria, een algeheel beeld van de prestatie van de beoordeelde vormt. Die algemene indruk overlapt deels met de algemene indruk van andere beoordelaars, en is deels uniek voor de beoordelaar.

Om een aantal redenen geven medewerkers er de voorkeur aan te worden beoordeeld door hun manager (Gosselin, Werner en Halle, 1997). De manager:

- is meer accuraat, dat wil zeggen er is sprake van minder halo of andere beoordelaarsfouten;
- heeft meer bronnen ter beschikking waaruit kan worden geput na een goede of slechte beoordeling;
- kent in het algemeen de functie omdat hij/zij vaak deze bekleedde voordat hij/zij manager werd;
- is meer ervaren in beoordeling omdat personeelsbeoordeling deel uitmaakt van de functie van de manager;
- is geïmmiteerd aan het beoordelingsproces omdat de uitkomst (1) vaak moet worden bericht aan het hoger management, en (2) positieve (bijvoorbeeld promotie) of negatieve (bijvoorbeeld geen bonus) consequenties heeft;
- maakt deel uit van een al op voorhand geaccepteerde asymmetrie tussen manager en medewerker, terwijl feedback door een collega juist asymmetrie zal induceren waarvoorheen collegiale, informele gelijkheid bestond.

Meerdere beoordelaars

Variantie in beoordelingen ontstaat door variantie in beoordeelenden en door variantie in beoordelaars. Dat wil zeggen, beoordelingsverschillen worden veroorzaakt door prestatieverschillen van beoordeelenden of door persoonlijke verschillen tussen beoordelaars. Onderzoek wijst uit dat een groot percentage van de variantie in beoordelingen moet worden toegeschreven aan de beoordelaar, niet aan de beoordeelde. Zo vonden Scullen, Mount en Goff (2000) in twee grote studies ($N = 2350$ en $N = 2142$) dat 23 procent van de variantie in prestatiebeoordelingen kon worden toegeschreven aan de prestatie van de beoordeelenden en 58 procent aan beoordelings-tendenties van de beoordelaars als halo, 'leniency' (toegeeflijkheid in de zin van al gauw hoge beoordelingen geven) en 'recency' (laatste observaties geven de doorslag).

Enige bevindingen ten aanzien van beoordelaar-bias zijn (Heidemeier en Moser, 2009):

- Eigenschaps- (dimensionele) beoordelingen vertonen grotere 'leniency' dan vaardigheids- (situationele) beoordelingen.
- Beoordelaarsovereenstemming is lager in geval van persoonseigenschappen (dimensies).
- Medewerkers overschatten hun 'contextuele prestatie' (datgene wat men doet buiten de officiële taakbeschrijving, zoals collega's helpen en zich als goede burger gedragen in de organisatie) meer dan hun taakprestatie.
- Wanneer een zelfbeoordeling wordt gevolgd door externe validering, bijvoorbeeld in de vorm van tests of taakproeven, leidt dat tot een meer bescheiden zelfbeoordeling.
- Zowel het gemiddelde van de beoordelingen als de overeenstemming tussen beoordelingen is groter in meer collectivistische samenlevingen.

Werk situatie in plaats van persoon

Wellicht is het probleem bij inzet van meerdere beoordelaars niet dat zij als persoon verschillen, maar dat zij zich in verschillende werksituaties bevinden, dat wil zeggen dat ze verschillen in werkrelatie met de beoordeelde. Zij zijn bijvoorbeeld collega, manager, medewerker of klant. Volgens deze redenering zou valide beoordeling vereisen dat iedereen zich in dezelfde functiesituatie bevindt, dus een equivalente werkrol vervult ten opzichte van de beoordeelde. In dat geval zouden beoordelaars dezelfde informatie hebben en hetzelfde type criterium toepassen om taakeffectiviteit te beoordelen. Groepsbeoordeling vereist dus dat de taak zo is gestructureerd dat er maar één correcte oplossing is (Martell en Leavitt, 2002) en beoordelaars daarom onderling 'uitwisselbaar' zijn. Uitmiddelen van beoordelingsverschillen vereist dat de beoordelaars als uitwisselbaar kunnen worden beschouwd; het zijn 'realisaties' van een en dezelfde beoordelaar (net als meerdere worpen van een dobbelsteen nog steeds worpen zijn van dezelfde dobbelsteen). Mount, Judge, Scullen, Sytsma en Hezlett (1998) onderzochten of het inderdaad niet de persoon maar de organisatorische positie (rol) van de beoordelaar is die de beoordelingsvariantie veroorzaakt. Zullen beoordelaars in dezelfde positie dezelfde beoordelingen geven? Mount et al. (1998) antwoorden met een duidelijk 'nee!' op deze vraag. Beoordelaarsvariantie in prestatiebeoordelingen hangt sterker samen met persoonsverschillen tussen individuele beoordelaars dan met verschillen in organisatorische positie.

Het is mogelijk wat meer specifiek te zijn over 'dezelfde werksituatie' en 'uitwisselbaar zijn'. Viswesvaran et al. (2002) vonden in hun meta-analyse dat overeenstemming over prestatiebeoordelingen tussen collega's en manager afhankelijk was van het type criterium:

- Er was minder overeenstemming over abstracte criteria als leiderschap of probleemoplossend vermogen. Dit zijn

typisch dimensionele criteria waarvoor generalisatie over meerdere werksituaties vereist is.

- Er was meer overeenstemming over concrete criteria als (i) situationele criteria: inzet en functiekenis, of als (ii) outputcriteria: aantal ongelukken, kwaliteit en output-niveau. Dit zijn typisch criteria waarbij het erom gaat bekwaamheid, efficiency of effectiviteit te beoordelen met betrekking tot één specifieke, kritische functiesituatie.

Ofschoon het grootste deel van de beoordelaarsvariantie gebonden is aan de persoon van de beoordelaar, wordt een deel door de functie van de beoordelaar bepaald, dat wil zeggen door de werkrelatie tussen beoordeelde en beoordelaar. Het blijkt dat beoordelaars in die relatie meer zeker zijn:

- over wat zij ervaren in een duidelijk afgebakende situatie met beperkte duur (bijvoorbeeld vergadering of presentatie of verkoopgesprek);
- over wat zij ontvangen op een min of meer kwantificeerbare manier.

In de volgende paragraaf wordt nader ingegaan op feedback door meerdere beoordelaars.

4 Multi-rater feedback

4.1 De multi-rater methode

In toenemende mate worden zogenoemde multi-rater feedbackmethoden toegepast in werkorganisaties (Church en Bracken, 1997). Door middel van multi-rater feedback, ook wel aangeduid als multi-source feedback (London en Smither, 1996) of 360-gradenfeedback, worden bij het werk van de beoordeelde betrokken personen (zoals collega's, de eigen manager, ondergeschikten en klanten; in het vervolg aangeduid als 'omstanders') gevraagd aan de hand van een standaardlijst met gedragsitems aan te geven hoe de beoordeelde medewerker op hen overkomt. Voorbeelden van zulke gedragsitems zijn:

- hij/zij is iemand bij wie je gemakkelijk terecht kunt met je vragen;
- hij/zij test nieuwe ideeën zo veel mogelijk eerst uit bij medewerkers;
- hij/zij houdt voet bij stuk in lastige kwesties.

De beoordeelde persoon vult zelf ook de lijst in. Scores van omstanders worden gemiddeld en vergeleken met die van de beoordeelde persoon zelf. Aldus verkrijgt de beoordeelde informatie over hoe relevante rolpartners in het werk zijn/haar bijdrage of gedrag ervaren en beoordelen. De medewerker wordt gezien als het middelpunt van een 'cirkel' van beoordelaars die vanuit verschillende (invals)hoeken hem of haar ervaren.

Noodzakelijke voorwaarde voor zinvolle toepassing van multi-rater feedback is dat beoordelaars zich een beeld hebben moeten kunnen vormen, dus in redelijke mate

moeten interacteren met de persoon in kwestie, dus deel moeten uitmaken van diens communicatiepatroon. Vandaar dat gebruik van multi-rater feedback niet los gezien kan worden van dat communicatiepatroon:

- Omstanders zijn geen onbevangen, ‘objectieve’ buitenstaanders maar belanghebbend; ze hebben iets met de beoordeelde persoon.
- Herhaald gebruik van multi-rater feedback heeft zelf ook weer effect op de samenwerking en communicatie tussen beoordeelden en beoordelaars.

Waar het assessment center resulteert in een inventarisatie van effectief beheerste situaties, dus in een specifiek competentieprofiel, leidt de multi-rater feedbackmethode tot een algemeen beeld van de beoordeelde persoon (vgl. Atkins en Wood, 2002). Multi-rater feedback schept de noodzaak betrokken te zijn bij de te beoordelen persoon. De beoordelaar is in die zin als het ware ‘klant’ van de beoordeelde. Het assessment center veronderstelt distantie tussen kandidaat en assessoren, multi-rater feedback vereist juist interactie. Om die distantie te waarborgen, zijn assessoren getraind in de beoordeling. Als gevolg daarvan is de overeenstemming tussen verschillende beoordelaars bij het assessment center in het algemeen acceptabel tot goed (Engelbrecht en Fischer, 1995) en bij de multi-rater methode laag tot afwezig (Heidemeier en Moser, 2009). Andere, ‘goede’ (dat wil zeggen, ze bepalen juist de kracht van de methode) redenen voor lage beoordelaarsovereenstemming bij de multi-rater methode komen hierna aan de orde.

Aard van de feedback: klant of deskundige?

Het volgende voorbeeld kan het verschil tussen beoordeling in multi-rater feedback en in een assessment center verduidelijken (zie ook tabel 1). Als een pianist een recital geeft, is de feedback van het aanwezige publiek (= diens ‘klanten’) gebaseerd op hun beleving van de muzikale prestatie, op het beeld dat zij hebben. Beeldfeedback is meer motiverend dan ‘droge’ assessment center-feedback, omdat het publiek tevreden stellen tenslotte het ultieme doel van de pianist is. Belangrijk is evenwel dat klantfeedback geen concrete aanwijzingen voor gedragsverbetering bevat. Daadwerkelijke gedragsverbetering vereist gedragsfeedback, dus terugkoppeling in termen van competenties, en die haalt de pianist niet bij zijn publiek maar bij een deskundig pianoleraar.

Multi-rater feedback en de personeelsbeoordeling

Bij de traditionele personeelsbeoordeling stelt de manager in samenspraak met de medewerker de kwantiteit en kwaliteit van het functioneren van die laatste vast, weegt beide aan de hand van tevoren gemaakte afspraken en doelstellingen en doet vervolgens eenzijdig een uitspraak in de zin van bijvoorbeeld ‘zeer goed’, ‘goed’, ‘voldoende’ of ‘verbetering is noodzakelijk’. Er is dus sprake van één beoordelaar (de manager) die bovendien vooral vanuit het oogpunt van de (afgesproken) bijdrage aan de realisering van organisatie-doelstellingen kijkt naar de medewerker. Het overzicht in tabel 2 maakt duidelijk dat multi-rater beoordeling en ‘gewone’ personeelsbeoordeling op nogal wat punten verschillen.

Het is dan ook in het algemeen onverstandig om beide typen beoordelingsinstrumenten aan elkaar te koppelen. In onderzoek bleek dat ruim 35 procent van de beoordelaars hun beoordeling in positieve zin aanpassen als die wordt meegenomen bij de vaststelling van de formele personeelsbeoordeling (Waldman, Atwater en Antonioni, 1998). Dit is begrijpelijk: mensen richten hun gedrag naar de verwachte positieve of negatieve effecten ervan. Bij complete multi-rater feedback heeft men elkaar in een ‘collectieve greep’, dat wil zeggen het ‘spel’ is: als jij mij nu negatief beoordeelt, zal ik dat later ook bij jou doen ... De in de praktijk onvermijdelijke uitkomst van dit ‘contract’ is inflatie van de beoordeling: iedereen beoordeelt elkaar na verloop van tijd ‘voldoende’. De kracht van de multi-rater feedbackmethode verdwijnt als sneeuw voor de zon wanneer men deze verbindt met de formele personeelsbeoordeling, en zeker wanneer men deze koppelt aan salarisverhogingen.

4.2 Onderzoekresultaten

Relatie tussen verschillende beoordelaars

Een zeer algemene uitkomst uit onderzoek is dat de beoordelingen van de verschillende omstanders niet overeenstemmen (Heidemeier en Moser, 2009). Redenen:

- beoordelaars zijn niet getraind in beoordelen;
- beoordelaars zijn (te) betrokken bij de beoordeelde persoon, houden dus onvoldoende afstand;
- beoordelaars kennen de beoordeelde persoon alleen maar in één bepaald opzicht.

Tabel 1 Feedback op competenties of op het klantbeeld

voorwerp van feedback	effect van feedback	gever van feedback	relatie feedback gever-ontvanger	instrument van feedback
situationele competentie van focale persoon	informerend	deskundige in de functie	distantie	assessment center
tevredenheid van ‘klant’ over wat focale persoon levert	motiverend	klant van de functionaris	betrokkenheid	multi-rater feedback

Tabel 2 Multi-rater feedback vergeleken met de personeelsbeoordeling

	multi-rater beoordeling	personeelsbeoordeling
Beoordelaar	relevante omstanders of 'klanten': manager, medewerker, collega, externe klant	manager
Criterium	'klanttevredenheid'	organisatiebijdrage
Vervolg	- gedragsverandering - bijstelling van zelfbeeld	gedragsverandering, salarisbijstelling, mutatie, opleidingsadvies, 360° feedback (etc.)
'Eigenaar'	individuele beoordeelde	organisatie
Frequentie	indien nodig	jaarlijks

In het laatste geval zijn inconsistente beoordelingen vanuit de omgeving min of meer natuurlijk. Het wordt helemaal niet vreemd gevonden dat een manager bijvoorbeeld een positieve beoordeling op sensitiviteit krijgt van zijn/haar medewerkers, maar een negatieve beoordeling op sensitiviteit door zijn/haar hogere manager. Sensitiviteit heeft voor medewerkers een andere concrete gedragsvertaling (bijvoorbeeld in een *coachende* stijl naar hen luisteren, rekening houden met hun beweegredenen en gevoelens) dan voor de hogere manager (bijvoorbeeld in een meer *directieve* stijl naar medewerkers de hogere manager het gevoel geven dat gedaan wordt wat volgens degene de bedoeling is). Dit betekent dat de terugkoppeling en interpretatie van de feedback zeer specifiek, op het niveau van de afzonderlijke omstanders dient te geschieden. Optellen van feedback over verschillende typen omstanders is uit den boze.

Effect van multi-rater feedback

In het algemeen blijken personen zichzelf te overschatten als het gaat om positieve gedragingen of competenties als plannen, vergaderen en naar medewerkers luisteren. Mensen hebben een (te) rooskleurig beeld van zichzelf, zijn optimistisch over wat ze kennen en kunnen. Dan komt feedback van pas. De vraag is natuurlijk in welke mate de beoordeelde persoon omstanderfeedback verkregen uit een multi-rater toepassing, als 'objectief' waarneemt. Een andere belangrijke vraag is welk effect gebruik van multi-rater feedbackmethoden heeft op interacties in het werk, dus op het samenwerken.

Logischerwijs kunnen er drie groepen van beoordeelde personen onderscheiden worden:

- 'overschatters': de zelfbeoordeling is consequent hoger (positiever) dan die door de omstanders(s),
- 'onderschatters': de zelfbeoordeling is consequent lager (negatiever) dan die door de omstanders(s),

- 'overeenstemmers': de zelfbeoordeling is consequent gelijk aan die door de omstanders(s).

Overschatting blijkt de regel. Overeenstemmers blijken door omstanders meer effectief gevonden te worden dan 'overschatters'. De reden daarvoor is dat 'overeenstemmers' open staan voor reacties van anderen (van hun rolpartners of 'klanten'), daarmee meer rekening kunnen houden in hun handelen en dus zowel meer empathisch als meer sensitief zijn. Voor overschatters zonder omgevingsfeedback bestaat er geen enkele reden tot verandering: ze doen het in eigen ogen goed en dat is ook de enige informatie die ze hebben. En derhalve zullen ze zich ook niet verbeteren. Onderschatters zonder omgevingsfeedback hebben in hun eigen lage zelfbeoordeling een duidelijke aanleiding om te veranderen, maar ze hebben bij gebrek aan concrete gedragsfeedback geen aanknopingspunt voor de richting van de verandering.

Uitkomsten van onderzoek naar hoe overschatters en onderschatters omgaan met multi-rater feedback zijn niet eenduidig. Bij overschatters verbetert het gedrag (zoals beoordeeld door omstanders), maar de overschatting blijft of neemt af. De mate van overschatting is natuurlijk wel kleiner geworden, want men is volgens de omstanders tenslotte competentier geworden. Onderschatters blijken eerder geneigd hun eigen gedragingen positief aan te passen op grond van feedback van omstanders dan overschatters. Bovendien stijgt ook de zelfbeoordeling van onderschatters na positieve omgevingsfeedback. De zelfevaluatie wordt dus naar boven aangepast. Het voorafgaande wijst op relatieve onafhankelijkheid van zelfbeoordeling en gedrag zoals beoordeeld in de omstanderbeoordeling. Bovendien blijkt multi-rater feedback een nuttig instrument om over- en onderschatters indringend 'wakker te schudden'. Overschatters worden gewezen op het feit dat ze niet zo goed zijn als ze denken dat ze zijn, onderschatters wordt duidelijk gemaakt dat ze in de praktijk beter functioneren dan ze zelf denken.

5 Samenvatting en conclusies

Er bestaat een sterke bedrijfspsychologische invalshoek bij bestudering en advisering over feedback. Die is terecht, omdat feedback gedragsbeïnvloeding als doel heeft, maar het kan ook leiden tot een eenzijdige focus op feedback als meetvraagstuk, dus op betrouwbaarheid en validiteit. De bespreking van multi-rater feedback laat zien dat organisatorische inbedding van feedback niet verwaarloosd mag worden. Feedback is mede een organisatorisch, management- en HRM-thema (zie de inleiding in dit themanummer). Zo blijkt effectieve implementatie van (multi-rater) feedbacksystemen afhankelijk van onder meer de volgende organisatorische voorwaarden (Fleenor, Taylor en Chappelow, 2008; Lepsinger en Lucia, 1997):

- Het topmanagement is daadwerkelijk betrokken, zowel als beoordelaar als beoordeelde.

- Het type beoordelingsdimensies is gekoppeld aan de aard van de organisatie en wordt dus niet ontleend aan bijvoorbeeld een standaardlijst van een extern adviesbureau.
- Er is transparantie over wat er met de gegevens gebeurt.
- Feedback gericht op persoonlijke ontwikkeling wordt niet gekoppeld aan feedback gericht op bijvoorbeeld prestatie-evaluatie en bonustoekenning.

Aandacht vragen voor dit type voorwaarden is niet overbodig. Zo bestaat er in de meerderheid van de organisaties de sterke neiging om multi-rater feedback te verbinden met prestatie-management en salarisverhoging (Bracken, 2009). ■

Prof. dr. P.G.W. Jansen is sedert 1992 hoogleraar Bedrijfspsychologie aan de Faculteit der Economische Wetenschappen en Bedrijfskunde, VU Amsterdam. Hij studeerde (*cum laude*) Mathematische Psychologie aan de Katholieke Universiteit Nijmegen. Hij was 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*.

Literatuur

- Atkins, P.W.B. en R.E. Wood (2002), Self-versus others' ratings as predictors of assessment center ratings: validation evidence for 360-degree feedback programs, *Personnel Psychology*, vol. 55, pp. 871-904.
- Bowman, J.S. (1999), Performance appraisal: Verisimilitude trumps veracity, *Public Personnel Management*, vol. 8, pp. 557-576.
- Bracken, D.W. (2009), Book review, *Personnel Psychology*, vol. 62, pp. 175-179.
- Church, A.H. en Bracken, D.W. (1997), Advancing the state of the art of 360-degree feedback: Guest editors' comment on the research and practice of multirater assessment methods, *Group & Organization Management*, vol. 22, pp. 149-161.
- Dweck, C.S. en E.L. Legget (1988), A social-cognitive approach to motivation and personality, *Psychological Review*, vol. 95, pp. 256-273.
- Engelbrecht, A.S. en A.H. Fischer (1995), The managerial performance implications of a developmental assessment center process, *Human Relations*, vol. 48, no. 4, pp. 387-404.
- Fleenor, J.W., S. Taylor en C. Chappelaw (2008), *Leveraging the impact of 360-degree feedback*, San Francisco: Pfeiffer.
- Goodman, J.S. en R.E. Wood (2004), Feedback specificity, learning opportunities, and learning, *Journal of Applied Psychology*, vol. 89, no. 5, pp. 809-821.
- Goodman, J.S., R.E. Wood en M. Hendrickx (2004), Feedback specificity, exploration, and learning, *Journal of Applied Psychology*, vol. 89, no. 2, pp. 248-262.
- Gosselin, A., J.M. Werner en N. Halle (1997), Ratee preferences concerning performance management and appraisal, *Human Resource Development Quarterly*, vol. 8, pp. 315-333.
- Groot, T.L.C.M. en P.G.W. Jansen (2010), *Performance management and control*, VU University of Amsterdam, Faculty of Economics and Business Administration.
- Heidemeier, H. en K. Moser (2009), Self-other agreement in job performance ratings: A meta-analytic test of a process model, *Journal of Applied Psychology*, vol. 94, no. 2, pp. 352-370.
- Ilies, R. en T.A. Judge (2005), Goal regulation across time: The effects of feedback and affect, *Journal of Applied Psychology*, vol. 90, pp. 453-467.
- Keith, N. en M. Frese (2008), Effectiveness of error management training: A meta-analysis, *Journal of Applied Psychology*, vol. 93, pp. 59-69.
- Kinicki, A.J., G. Prussia, B. Wu en F.M. McKee-Ryan (2004), A covariance structure analysis of employees' response to performance feedback, *Journal of Applied Psychology*, vol. 89, pp. 1037-1069.
- Kluger, A.N. en DeNisi, A. (1996), The effects of feedback interventions on performance. A historical review, a meta-analysis, and a preliminary feedback intervention theory, *Psychological Bulletin*, vol. 119, pp. 254-284.
- Lepsinger, R. en Lucia, A.D. (1997), *The art and science of 360° feedback*, San Francisco: Pfeiffer.
- London, M. en Smither, J.W. (1996), Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research, *Personnel Psychology*, vol. 48, pp. 803-839.
- Martell, R.F. en K.N. Leavitt (2002), Reducing the performance-cue bias in work behavior ratings: Can groups help?, *Journal of Applied Psychology*, vol. 87, pp. 1032-1041.
- Mount, M.K., T.A. Judge, S.E. Scullen, M.R. Sytsma en S.A. Hezlett (1998), Trait, rater and level effects in 360-degree performance ratings, *Personnel Psychology*, vol. 51, pp. 557-576.
- Rietdijk, M.M. (2009), *Organisaties conditioneren. De invloed van beloning en straf op werkprestaties*, proefschrift Vrije universiteit Amsterdam, zie: <http://hdl.handle.net/1871/15235>.
- Scullen, S.E., M.K. Mount en M. Goff (2000), Understanding the latent structure of job performance ratings, *Journal of Applied Psychology*, vol. 85, pp. 956-970.
- Viswesvaran, C., F.L. Schmidt en D.S. Ones (2002), The moderating effect of job performance dimensions on convergence of supervisory and peer ratings of job performance: Unconfounding construct-level convergence and rating difficulty, *Journal of Applied Psychology*, vol. 87, pp. 345-354.
- Waldman, D.A., L.E. Atwater en D. Antonioni (1998), Has 360 degree feedback gone amok? *Academy of Management Executive*, vol. 12, no. 2, pp. 86-94.
- Woo, S.E., C.S. Sims, D.E. Rupp en A.M. Gibbons (2008), Development engagement within and following developmental assessment centers: Considering feedback favorability and self-assessor agreement, *Personnel Psychology*, vol. 61, pp. 727-759.