

Een visie op en onderzoek naar de 360°-feedbackmethode

P.G.W. Jansen en A.M.L. de Jong

In werkorganisaties wordt in toenemende mate de beoordelingsmethode 360°-feedback toegepast. Daarbij wordt aan nauw bij het werk van de te beoordelen medewerker (de 'focale persoon') betrokken personen ('omstanders') gevraagd aan de hand van een vragenlijst met 'gedragsitems' (bijvoorbeeld: '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' en 'Hij/zij houdt voet bij stuk in lastige kwesties') hun inschatting van deze medewerker te geven. In deze vorm van 'multi-source feedback' (London en Smither, 1995) beoordelen dus relevante omstanders in welke mate specifieke gedragingen voorkomen bij de focale persoon. De beoordeelde persoon – het middelpunt ('focus') van de cirkel van omstanders – vult zelf ook de lijst in. Scores van omstanders worden gemiddeld en afgezet tegen die van de beoordeelde persoon zelf en het geheel wordt teruggerapporteerd aan de focale persoon.

Gepubliceerd empirisch onderzoek naar de 360°-feedbackmethode is vrijwel uitsluitend afkomstig uit de Verenigde Staten, zoals blijkt uit de referenties in de volgende paragraaf. In Nederland zijn inmiddels enkele op de toepassing gerichte boeken verschenen (Van den Berg, Van den Broek en Pijs, 1997; Van der Togt en Kemp, 1997). Daarnaast zijn Jansen en Vloeberghs (1998) in een conceptuele beschouwing ingegaan op de positie van dit type feedback temidden van andere feedbackmethoden als de kwaliteitsenquête en het assessmentcenter. Verslagen van empirisch Nederlands onderzoek zijn echter schaars (Breeuwer, Van der Zee en Smit, 1998; Jellema, 2000; Yperen en Baar, 2000); ze zijn voornamelijk te vinden in (een handvol) doctoraalscripties, zoals Van der Woude (1995), Rietveld (1997) en De Jong (1998).

Dit artikel geeft een visie op de waarde van de 360°-feedbackmethode vanuit de theorie, geïllustreerd aan de hand van een psychometrisch onderzoek naar een Nederlandse 360°-vragenlijst. Daarnaast wordt ingegaan op het toenemend gebruik ervan in de praktijk. Enkele belangwekkende en deels nog onopgeloste theoretische kwesties rond de 360°-

Prof. dr. P.G.W. Jansen is als hoogleraar Bedrijfspsychologie verbonden aan de afdeling Management & Organisatiekunde van de Faculteit der Economische Wetenschappen en Econometrie van de Vrije Universiteit Amsterdam. Mw. drs. A.M.L. de Jong is bedrijfskundig econoom en als organisatieadviseur werkzaam bij DTO te Den Haag.

methode worden aangesneden. Die worden voorbereid in de volgende paragraaf.

Onderzoek naar de 360°-feedbackmethode

Van belang bij het gebruik van 360°-feedback zijn ten eerste de mate waarin omstanders in hun beoordeling van dezelfde focale persoon overeenkomen, ten tweede de mate waarin dit instrument samenhangt met meer ingeburgerde feedbackmethoden als vragenlijsten en het assessmentcenter en ten derde de mate waarin de focale persoon de feedback accepteert en er daadwerkelijk iets mee doet. Deze vragen zijn stevast leidraad geweest voor onderzoek, maar het antwoord is nog geenszins helder (Waldman, Atwater en Antonioni, 1998).

Correlaties tussen beoordelaars

Jansen en Vloeberghs (1998) concluderen dat de 360°-feedbackmethode moeilijk te vergelijken is met bijvoorbeeld een assessmentcenter-beoordeling. Bij de 360°-methode gaat het om een inventarisatie van situatiegebonden gedrag en wel via relevante anderen, dat wil zeggen via 'betrokken omstanders'. Beoordelaars hebben zich een indruk moeten kunnen vormen, dus in redelijke mate moeten interacteren met de persoon in kwestie. Daarentegen geldt dat in het assessmentcenter assessoren in het algemeen uitdrukkelijk niet in interactie zijn met de kandidaat. Daarnaast zijn ze getraind in de hantering van een duidelijke norm van meer of minder effectief gedrag, in de vorm van expliciet geformuleerde 'gedragsdimensies'. Bovendien krijgen ze relevant gedrag voorgeschoteld doordat het in gecontroleerde omstandigheden wordt opgeroepen en zijn ze in staat met elkaar te overleggen omdat er meerdere assessoren zijn die een gemeenschappelijke waarnemingsbasis bezitten.

Dit verschil in posities van de beoordelaars komt terug in de betrouwbaarheid van de 360°-beoordelingen. Waar interbeoordelaarsbetrouwbaarheden bij assessmentcenters in het algemeen acceptabel tot goed zijn (gemiddeld .75, zie Jansen, 1993), zijn ze bij de 360°-methode consistent laag (Van der Woude, 1995, pagina 14); ze variëren tussen .15 en .45. Hieruit blijkt dat omstanders nauwelijks correleren in hun beoordelingen en dat geldt ook voor de correlaties tussen omstander en focale persoon.

In de toegepaste literatuur wordt van deze nood een deugd gemaakt (zie bijvoorbeeld van den Berg et al., 1997, pagina 44) door erop te wijzen dat dit alleen maar bevestigt dat er sprake is van verschillende 'per-

spectieven' op de beoordeelde. Maar lage overeenstemming tussen beoordelaars impliceert dat optellen over deze personen teneinde een 'genuanceerd beeld' van het functioneren te krijgen, geen meettechnische basis heeft. Anders geformuleerd: middelen van de oordelen over de beoordelaars teneinde de foutvariantie te verkleinen (Van der Woude, 1995, pagina 11) levert geen juist beeld op. Men kan ook besluiten de feedback te beperken tot wat individuele beoordelaars melden. Onderzoek van Mount et al. (1998) geeft aan dat uitkomsten meer afhankelijk zijn van de individuele beoordelaar dan van de positie die een beoordelaar ten opzichte van de betrokkene heeft. Dit heeft als nadeel dat de betrouwbaarheid van dergelijke eenzijdige beoordelingen in het algemeen erg laag is.

Ten slotte blijken er verschillen te bestaan in de aan de beoordeling ten grondslag liggende factorstructuren tussen de zelfbeoordeling en de beoordeling door relevante anderen. Dat maakt vergelijkbaarheid lastig. Zo kan een zelfbeoordeling meer genuanceerd zijn (en is er dus sprake van meer onderliggende factoren) dan een omstanderbeoordeling (waar men zich als ongetrainde beoordelaar al gauw laat verleiden tot één onderliggende 'good guy – bad guy'-dimensie). Ter vergelijking: uit assessmentcenteronderzoek komt als robuust feit naar voren dat zelfs getrainde managerassessoren in het algemeen feitelijk slechts een beperkt aantal dimensies hanteren (Sagie en Magnezy, 1997).

Een vergelijking met meer traditionele beoordelingsmethoden, zoals de beoordeling door de direct leidinggevende, ontbreekt in dit artikel, omdat het lastig bleek hierover gegevens te verzamelen.

Relatie tot andere methoden

Bij de assessmentcentermethode worden persoonlijke appreciaties van de assessoren zo veel mogelijk voorkomen door training in beoordeling van gedrag en door standaardisatie van het assessmentcenter. Bij omstanders in een 360°-feedbacktoepassing is dit niet het geval. Men is niet getraind, heeft nadrukkelijk een werkrelatie met de beoordeelde en bepaalt zelf op grond van welke ervaringen tot de beoordeling inkleurt. Factoren waarvan bekend is dat ze de beoordeling ten onrechte kunnen beïnvloeden, zoals intelligentie, aardig zijn en een goede prestatie in het verleden, kunnen vrijelijk meespelen bij de 360°-feedbackmethode (Borman et al., 1995; Mabe en West, 1982). Ook blijkt de beoordeling in het algemeen positiever, tenminste zolang deze niet anoniem is (Antonioni, 1994).

De belangrijkste reden waarom er bij de 360°-feedbackmethode meer storende factoren van invloed zijn, is dat deze feedbackmethode infor-

matie geeft over hoe in het dagelijks werk betrokken personen het werk van de beoordeelde ervaren. Er is sprake van 'klantfeedback', dat wil zeggen beoordeling door belanghebbenden. In een assessmentcenter is er daarentegen sprake van feedback van onbetrokken experts. Om die reden plaatsen Jansen en Vloeberghs (1998) 360°-feedback in de traditie van 'Total Quality Management', waar de evaluatie door de klant maatgevend is en diens ervaringen zonder meer voor waar worden gehouden. De 'beoordelaar als klant' hoeft zich op geen enkele wijze te verantwoorden voor het oordeel. Vanuit een strikt psychometrisch oogpunt ontstaat er hier een lastige kwestie: waar in het assessmentcenter onbetrouwbare oordelen worden genegeerd, zal vanuit kwaliteitsoogpunt een ontevreden klant nooit genegeerd worden.

Gebruik en effectiviteit van de feedback

Volgens Korman (1970) zijn personen geneigd zich zodanig te gedragen dat hun gedrag in overeenstemming is met hun zelfwaarneming. Maar die zelfperceptie wordt mede bepaald door evaluatie door omstanders; daarop sluit de 360°-methode aan. Een definiërend kenmerk van de 360°-feedbackmethode is immers dat de zelfbeoordeling wordt vergeleken met omstanderbeoordelingen. Op basis van de mate van overeenstemming tussen de beoordeling van de persoon zelf en de beoordeling door anderen (berekend als correlaties) is een onderscheid te maken in drie groepen (Atwater en Yammarino, 1992). De eerste groep is in-agreement, oftewel de 'overeenstemmers': de zelfbeoordeling is consequent gelijk aan de beoordeling door anderen. Er zijn 'goede' overeenstemmers (worden goed beoordeeld en beoordelen zichzelf goed) en 'zwakke' overeenstemmers (worden slecht beoordeeld en beoordelen zichzelf slecht). De tweede groep wordt gevormd door de over-estimators, oftewel de 'overschatters': de zelfbeoordeling is consequent positiever dan de beoordeling door anderen. En ten slotte zijn er nog de under-estimators, oftewel de 'onderschatters': de zelfbeoordeling is consequent negatiever dan de beoordeling door anderen.

Verschillen tussen de mate van overeenstemming van de zelfbeoordeling en beoordeling door anderen hebben consequenties voor de werkeffectiviteit. Yammarino en Atwater (1997) geven aan dat 'goede overeenstemmers' zeer positieve, 'zwakke overeenstemmers' negatieve, 'overschatters' zeer negatieve en 'onderschatters' gemengde werkprestaties leveren. Verklaring voor deze bevinding is dat overeenstemmers zich meer bewust zijn van het effect van hun handelen. Dat heeft tot gevolg dat zij hun eigen werkgedrag beter kunnen beoordelen, wat leidt tot effectiever functioneren (Atwater en Yammarino, 1992; Furnham en Stringfield, 1994).

Momenteel bestaan er nog weinig data over de effectiviteit, dat wil zeggen over de mate waarin de uitkomsten van een 360°-feedbackronde van invloed zijn op de prestaties van de beoordeelde, **van toepassing van de 360°-feedbackmethode**. Er is sprake van een beperkt aantal deugdelijke, maar kortetermijnstudies (Atwater, Roush en Fishtal, 1995; Fleenor et al., 1996; Reilly, Smither en Vasilopoulos, 1996) met wisselende bevindingen en daarnaast zeer veel anekdotiek. In Jansen en Vloeberghs (1998) wordt een overzicht gegeven. Samenvattend lijkt het effect van omstanderfeedback dat bij overschatters het gedrag (zoals beoordeeld door omstanders) verbetert maar de zelfoverschatting blijft (Fleenor et al., 1996) of afneemt (Atwater et al., 1995) en bij onderschatters het gedrag niet verandert maar de onderschatting afneemt (Johnson en Ferstl, 1999). Uit een onderzoek van Atwater et al. (1998) komt echter naar voren dat de effectiviteit van onderschatters toeneemt en die van overschatters afneemt. Volgens Atwater et al. (2000) spelen voor de effectiviteit van feedback twee factoren mee, namelijk de mate van cynisme van de medewerker over de organisatie en de mate waarin iemand de waarde van de feedback herkent en ook gemotiveerd is hier nadere informatie over te vergaren. Conclusie is dat feedbackprocessen kunnen leiden tot gedragsverandering, prestatieverbetering en toegenomen zelfbewustzijn van individuen. Een additioneel effect van upward feedback is dat deze feedback de relatie tussen een medewerker en de leidinggevende en hun wederzijdse commitment kan verstevigen. Ook Walker (1999) constateert dat managers die slecht of middelmatig werden beoordeeld, na een periode van vijf jaar waarin ze participeerden in een opwaarts terugkoppelingsprogramma, duidelijk beter worden beoordeeld. Het meeste effect wordt gesorteerd als de resultaten worden besproken.



Een belangrijke voorwaarde voor de werkzaamheid van een methode is acceptatie. Veelzeggend is dat in de Verenigde Staten van de organisaties die 360° hadden ingevoerd in het kader van formele personeelsbeoordeling, inmiddels de helft daar weer van is afgestapt omdat zo'n toepassing te veel negatieve reacties opriep bij medewerkers. Bovendien begonnen omstanders al snel de neiging te vertonen eigenlijk alleen nog maar overdreven positief te beoordelen (Waldman, Atwater en Antonioni, 1998). Als deze methode wordt toegepast in het kader van personeelsontwikkeling, is de acceptatie van 360°-feedback groot (68% goed; 13% zeer goed) (Jellema, 2000).

De conclusie is dat nog geenszins duidelijk is wat 360°-feedback meet en wat op termijn de effecten ervan zijn op het (samen)werken (Waldman et al., 1998). In het voorbeeld wordt dit geïllustreerd aan de hand van een studie naar de psychometrische kwaliteit van een speci-

fiek, in Nederland toegepast 360°-feedbackinstrument. Gelet op hetgeen is gevonden in de Verenigde Staten verwachten we dat:

- het instrument voldoet aan de klassieke kenmerken als interne consistentie en factorstructuur;
- de correlaties tussen verschillende groepen beoordelaars laag zullen zijn;
- er discrepanties tussen zelf- en omstanderbeoordelingen zijn, waarbij zelfoverschatting domineert;
- er sprake zal zijn van lage correlaties tussen het 360°-instrument en andere feedbacktechnieken zoals assessmentcenter, intelligentietest en persoonlijkheidsvragenlijst;
- de acceptatie goed is zolang toepassing van het instrument beperkt blijft tot persoonlijke ontwikkeling.

Onderzoek 360°-feedbackinstrument van adviesbureau Construct Bedrijfspsychologie¹

In de periode 1997-1998 is een pilot uitgevoerd met het door Construct Bedrijfspsychologie ontwikkelde 360°-feedbackinstrument 360 CoRe. Dit instrument bestaat uit schalen voor de meting van 31 competenties die weer te clusteren zijn in zes competentiegebieden: Interpersoonlijke Vaardigheden, Management en Leiderschap, Werkinstelling, Stabiliteit, Intellectuele Capaciteiten en Communicatieve Vaardigheden. In de vragenlijst wordt eerst van elke competentie een definitie gegeven. Vervolgens wordt de competentie gemeten aan de hand van een schaal met items als:

- stemt activiteiten van zichzelf en van anderen op elkaar af, zodat doelen efficiënt kunnen worden gerealiseerd (competentie plannen en organiseren);
- blijft vriendelijk communiceren onder tijds- en werkdruk (competentie stressbestendigheid).

De items worden gescoord op een vijfpuntsschaal (zwak, matig, voldoet, sterk en uitzonderlijk); er is ook nog de optie 'niet van toepassing'. Het was niet mogelijk voor alle schalen voldoende gegevens te verkrijgen.

In totaal zijn de gegevens van 96 zelfbeoordelaars verkregen; 38 van hen zijn ook door (in totaal 333) mensen uit hun werkomgeving beoordeeld. Het gaat om leidinggevende(n), collega's, medewerkers, (externe) klanten en (externe) leveranciers. Het aantal omstanders per focale persoon varieert van 5 tot 22. Op deze groep van 38 zelfbeoordelaars worden analyses van de interbeoordelaarsovereenstemming en van de

overeenstemming tussen zelfbeoordeling en omstanderbeoordeling uitgevoerd.

De betrouwbaarheid

De waarde van alpha voor de zelfbeoordelingen varieerde van .77 tot .86. Bij de beoordelingen van omstanders lagen de waarden van alpha tussen de .76 en .94. Voor de alpha van het totaal waren de resultaten .78, .81, .82, .89 en .90. Over het algemeen is de alpha van de schalen relatief hoog. Ze liggen voor alle competenties ruim boven de acceptatiegrens van .60. Dit betekent dat de betrouwbaarheid van de schalen uitstekend is. Opvallend is dat met uitzondering van de schaal onderhandelen de alfa op basis van de beoordeling door anderen hoger is dan de op basis van de zelfbeoordelingen berekende alfa.

De factoranalyse

Per schaal zijn aparte exploratieve factoranalyses verricht op de interitemcorrelaties voor zelfbeoordelingen en omstanderbeoordelingen. De beoordelingen door relevante anderen zijn samengevoegd, omdat er onvoldoende data konden worden verzameld om voor de diverse groepen omstanders aparte factoranalyses uit te voeren. Bij de uitgevoerde factoranalyses is de vuistregel gehanteerd dat het aantal waarnemingen ongeveer vijf keer groter moest zijn dan het aantal items. Voor de schalen contactvaardigheid, onderhandelen, resultaatgerichtheid, besluitvaardigheid en oplossingsvermogen was dat niet het geval.

Uit de factoranalyse komt naar voren dat de schalen, met uitzondering van taakgericht management, min of meer unifactorieel zijn. Wat opvalt is dat het aantal factoren bij de zelfbeoordeling consequent groter is dan of gelijk is aan dat bij de omstanders. De verdeling van het percentage verklaarde variantie over de factoren varieert bij de omstanders van 48%-70% (uitgezonderd taakgericht management: 35%). Het was in dit onderzoek door gebrek aan voldoende data niet mogelijk de samenhang tussen de schalen te bepalen. Geconcludeerd kan worden dat de betrouwbaarheid en constructvaliditeit na enige aanpassing van de schalen in orde is.

Relatie tot andere beoordelingsinstrumenten

De overige 58 zelfbeoordelingen waren onderdeel van een breder onderzoek dat in het kader van een grootschalig assessment als gevolg van een reorganisatie naast uit de 360°-zelfbeoordeling ook nog bestond uit:

- een intelligentietest: de DAT-subtests woordenlijst, analogieën, rekenvaardigheid en zinnen;
- Twee persoonlijkheidsvragenlijsten: de NPV en de 'Quest'.²

Op deze groep is de analyse van de soortgenootvaliditeit gebaseerd. De resultaten zijn dat er nauwelijks verbanden bestaan tussen het 360°-feedbackinstrument enerzijds en de overige persoonlijkheids- en intelligentiescores anderzijds. Een beperkt aantal correlaties is significant en in de goede richting: de competentie contactvaardigheid vertoont bijvoorbeeld een negatieve relatie tot sociale inadequaatheid van de NVP en een positieve tot het construct sociale redzaamheid, en de constructen stressbestendigheid correleren $.49^{**}$ ($p < .01$). Het merendeel is echter niet significant. Dit roept de vraag op wat er met de 360°-feedbackmethode eigenlijk wordt gemeten.

De acceptatie

Door middel van enquêtes die bij de uitgezette vragenlijsten en bij de feedbackrapporten zijn gevoegd, is de acceptatie van het 360°-feedbackinstrument gemeten. Dit is gedaan aan de hand van stellingen als:

- Met behulp van dit instrument wordt een eerlijk en volledig beeld verkregen van het werkgedrag van een persoon.
- Ik heb de indruk dat ik eerlijk en objectief ben beoordeeld.
- De uitkomsten zijn waardevol voor mij.

Van de ruim 200 personen die de evaluatie hebben ingevuld, vindt 42% dat de 360°-feedbackmethode een eerlijk en volledig beeld geeft van het werkgedrag van een persoon, 45% is neutraal en 10% is het er niet mee eens. Nochtans vindt 73% dat de uitkomsten voor hen waardevol zijn. Ten slotte meent 46% dat de uitslag overeenkomt met andere beoordelingen, 36% is neutraal en 18% is het hiermee oneens. Al met al blijkt de acceptatie van het 360°-feedbackinstrument redelijk, maar hebben respondenten toch ook hun twijfels. Zo is slechts 31% respectievelijk 26% het (zeer) eens met de stelling dat 360°-feedback kan worden gebruikt voor de selectie, respectievelijk beloning van personen.

Discrepantie tussen de zelf- en omgevingscore

Per focale persoon is de gemiddelde afwijking tussen zelfbeoordeling en corresponderende omstanderbeoordelingen berekend. Er zijn drie groepen te onderscheiden: een groep met negatieve afwijkingen (de 'onderschatters'), een groep rond het midden met geen of nauwelijks afwijkingen (de 'overeenstemmers') en een groep met positieve afwijkingen (de 'overschatters'). Vanuit de literatuur zou mogen worden verwacht dat individuen zich over het algemeen 'overschatten'. In dit onderzoek zijn in verhouding meer 'overeenstemmers' en 'onderschatters'. De gemiddelde afwijking is $-0,4$. Uit eenzelfde analyse op

schaalniveau blijkt dat er ook daar sprake is van 'onderschatting'. Alleen specifieke (mens-)managementvaardigheden, zoals motiveren, leidinggeven en mensgericht management, blijken overschat te worden.

Overeenstemming: correlaties tussen focale persoon en omstanders

Per schaal zijn de correlaties tussen de zelfbeoordelingen en het gemiddelde van de verschillende groepen omstanders berekend. De gemiddelde zelf/omstandercorrelaties per schaal zijn berekend. Opvallend hoge gemiddelde correlaties (in de orde van .50) vinden we bij de schalen samenwerken, motiveren, strategisch management en schriftelijke uitdrukkingsvaardigheid. De gemiddelde correlatie over alle schalen is .35.

In de onderste twee rijen van tabel 1 staan de gemiddelde zelf/omstandercorrelaties per type beoordelaar, berekend over alle schalen. Ze variëren van .12 (zelf/manager) tot .46 (zelf/medewerker).

Overeenstemming: correlaties tussen verschillende groepen omstanders

Daarnaast zijn de gemiddelde correlaties tussen de in dit onderzoek onderscheiden vijf typen beoordelaars (manager, collega, medewerker, klant, leverancier) per schaal opgenomen. Opvallende hoge correlaties (in de orde van .50) vinden we bij de schalen strategisch management, zelfstandigheid en schriftelijke uitdrukkingsvaardigheid. De gemiddelde correlatie tussen omstanders over alle schalen is .32.

In de onderste twee rijen van tabel 1 staan de gemiddelde omstandercorrelaties per type beoordelaar, berekend over alle schalen. Ze variëren van -.02 (manager/leverancier) tot .56 (medewerker/leverancier). Opvallend is de relatief grote onderlinge overeenstemming (in de orde van .50) in de groep van collega/medewerker/klant/leverancier.

Tabel 1: Overeenstemming: vergelijking met eerder onderzoek

In tabel 1 zijn gemiddelde correlaties, berekend over alle schalen, tussen diverse typen beoordelaars (inclusief de focale persoon) uit dit onderzoek en uit eerdere studies opgenomen. De overeenstemming met de andere onderzoeken is redelijk groot; afwijkingen worden aangetroffen bij de correlaties voor zelf/manager (in dit onderzoek relatief laag), zelf/medewerker (in dit onderzoek relatief hoog) en manager/medewerker (in dit onderzoek zeer laag; dit hangt ook samen met beide voorafgaande bevindingen).

Onderzoek	Zelf- manager	Zelf- collega	Zelf- medewerker	Manager- collega	Manager- medewerker	Collega- medewerker
Fleenor et al. (1996)	.19		.24		.41	
Harris en Schaubroeck (1988)	.22	.24		.48		
Van der Woude (1995)	.24	.26	.26	.32	.30	.29
Wohlers en London (1993)	.16	.22	.24	.31	.17	.25
Dit onderzoek (1998)	.12	.26	.46	.26	.05	.48
Aanvullend in dit onderzoek (1998)	Zelf-klant	.16	Manager-klant	.21	Medewerker-klant	.50
	Zelf-leverancier	.33	Manager-leverancier	-.02	Medewerker-leverancier	.56
			Collega-klant	.53		
			Collega-leverancier	.48		

De correlatie tussen klant/leverancier uit het onderhavige onderzoek ontbreekt vanwege onvoldoende gegevens

Tabel 1. Gemiddelde correlaties tussen typen beoordelaars uit verschillende onderzoeken

Conclusie

Uit de factor- en betrouwbaarheidsanalyses in dit onderzoek komt (conform onze verwachting) naar voren dat de gebruikte schalen voldoen aan de psychometrische eisen. De alfacoefficienten voor beoordelingen door anderen zijn op een uitzondering na hoger dan de alfacoefficienten voor de zelfbeoordelingen. Hetzelfde vonden overigens Breeuwer et al. (1998). Uit de factoranalyse blijkt bovendien dat de focale persoon consequent meer dan of een gelijk aantal factoren onderscheidt als zijn/haar beoordelaars. Het lijkt erop dat zelfbeoordelingen genuanceerder zijn dan omstanderbeoordelingen. Daarvoor zijn nogal wat verklaringen denkbaar. Mogelijk is er bij de omstanders sprake van een halo-effect. Maar ook is het mogelijk dat zelfbeoordelaars meer stilstaan bij de oorzaken van hun gedrag en als gevolg hiervan meer aspecten meenemen, zoals intenties en de inspanning, terwijl omstanders alleen beoordelen op het gedrag dat ze waarnemen. Doordat de factorstructuur van de schalen afwijkt, moet in ieder voorzichtig worden omgegaan met de vergelijking van het zelfbeeld met het beeld van anderen.

Overeenstemming tussen beoordelaars

In het genoemde onderzoek bleek de discrepantie tussen de zelf- en omgevingsscore gering als wordt gekeken naar de gemiddelden per schaal. Dit is verrassend (en dan ook niet conform onze verwachting) gelet op resultaten uit eerder onderzoek. Een mogelijkheid is dat de verschillen in beoordeling door de diverse beoordelaarsgroepen elkaar compenseren, zodat per saldo het gemiddelde dicht bij de zelfscore komt te liggen. Bovendien bleken de 'onderschatters' de meerderheid te vormen. Een mogelijke verklaring hiervoor is dat ten eerste de steek-

proef vrij divers is. Er kunnen verschillende beoordelingsmethoden zijn gehanteerd op individueel, maar ook op organisatieniveau. Verschillen in strengheid van beoordelen kunnen zich ook voordoen afhankelijk van de cultuur van de organisatie. Ten tweede blijken vooral managers te lijden aan overschatting. Aangezien managers een relatief klein deel uitmaken van deze steekproef, kan hun invloed tenietgedaan zijn door de overige deelnemers.

Wel conform onze verwachting bleken de correlaties tussen de verschillende beoordelaarsgroepen nogal uiteen te lopen. De interbeoordelaars-correlatie varieert met de positie van de beoordelaars ten opzichte van de focale persoon. Maar die positie kan niet los worden gezien van zowel de hiërarchisch-formele als de sociale afstand. Doordat verschillende beoordelaars verschillende organieke posities innemen ten opzichte van de focale persoon, zullen zij hun oordeel baseren op een andere steekproef van gedragingen van die persoon (Greguras en Robie, 1998). Zo zullen bijvoorbeeld een collega en een leidinggevende anders aankijken tegen samenwerken. Een collega zal erop letten of de samenwerking als prettig wordt ervaren. Een leidinggevende zal het daarentegen belangrijk vinden dat door middel van samenwerking de gewenste resultaten worden behaald. Gezien het toenemende aantal mensen dat in projecten werkt waarbij de functionele baas niet dezelfde is als de hiërarchische leidinggevende, kan dit wellicht enigszins de populariteit van 360° verklaren. Daarnaast speelt de toegenomen behoefte aan persoonlijke ontwikkeling en het belang dat hier door met name professionals aan wordt gehecht, een rol.

Ook kan het type organisatie effect hebben op de overeenstemming tussen beoordelaars. Zo heeft bijvoorbeeld een leidinggevende in een sterk verplatte organisatie waar hij/zij deel uitmaakt van het team waarin ook de focale persoon werkt, een andere positie dan een leidinggevende in een meer hiërarchisch opgebouwde onderneming. Verwacht mag worden dat in het eerste geval de correlatie tussen beoordelingen van leidinggevende(n) en ondergeschikten hoger zal zijn dan in het tweede geval. Wellicht verklaart dit dat in het onderzoek van Breeuwer et al. (1998, pagina 52) de overeenstemming tussen de beoordelaars 'onverwacht hoog' bleek.

Ten slotte is de mate van overeenstemming ook afhankelijk van de observeerbaarheid van competenties. Zo waren focale persoon en omstanders het opvallend eens over de competenties strategisch management en schriftelijke uitdrukkingsvaardigheid. De laatste is goed observeerbaar en dus interpersoonlijk betrouwbaar in te schatten. Bovendien is het nauwelijks een competentie waarbij verschillende

typen klanten of rolpartners verschillende criteria zullen hanteren voor de beoordeling van hun tevredenheid. Voor de eensgezindheid over de competentie strategisch management hebben wij geen verklaring. In dit geval zou via interviews moeten worden vastgesteld welke concrete gedragingen leidraad waren bij en kritisch waren voor de beoordeling.

Relatie tot andere instrumenten

De analyse van de soortgenootvaliditeit bevestigt dat de 360°-methode een eigen positie inneemt ten opzicht van andere meetmethoden als de intelligentietest en de persoonlijkheidsvragenlijst. Bovendien blijkt uit ander onderzoek (bijvoorbeeld van der Woude, 1995) dat er eveneens nauwelijks samenhang is met assessmentcenterscores, zelfs als nominaal dezelfde dimensies worden gemeten. Dit is begrijpelijk: bij 360°-feedback zijn de beoordelaars 'klanten' en niet per se experts in de functie van de focale persoon. Bij een 360°-feedbackbeoordeling telt dus de mening van de omgeving; aan de wensen van de interne en externe 'rolpartners' moet worden voldaan om goed te kunnen functioneren in een organisatie. Bij de assessmentcentermethode daarentegen wordt juist getracht een zo genuanceerd mogelijk beeld te krijgen van de beoordeelde door gebruik te maken van onafhankelijke, getrainde beoordelaars die zijn ingevoerd in welke competenties noodzakelijk zijn voor het goed functioneren van de te beoordelen persoon. Het zijn experts in de beoordeling van de bij de focale persoon aanwezige competenties die relevant zijn voor zijn/haar functie.

Acceptatie

De acceptatie van het instrument bleek redelijk – maar ook niet meer dan dat. Ook internationaal blijkt er meer dan bij enig ander instrument (zoals het assessmentcenter) huiver te bestaan bij personen ten opzichte van de 360°-methode (vergelijk de titel van Waldman et al., 1998!). Niet onderzocht, maar zeer waarschijnlijk, is dat er een verschil bestaat in acceptatie tussen focale persoon en omstanders. De kans op daadwerkelijke verbetering in prestaties blijkt toe te nemen als de mogelijkheid bestaat de feedback met de betrokken beoordelaars te bespreken. Dit vergroot de acceptatie van de uitkomsten, maar vereist wel dat de anonimiteit van de beoordelaars wordt opgeheven. Voor effectiviteit van 360°-feedback is het daarom nodig beoordelaars 'accountable' te maken voor hun oordelen (Walker en Smither, 1999). Voldoende aandacht aan de follow-up van de resultaten schenken door bijvoorbeeld het opstellen van een persoonlijk ontwikkelplan, het volgen van een specifieke training en/of een gesprek met een deskundige zijn ook van belang (Walker en Smither, 1999; Jellema, 2000).

Eindconclusie

Geconcludeerd kan worden dat de psychometrische kwaliteit van het onderzochte instrument goed is. De interbeoordelaarsbetrouwbaarheid is laag, maar valt daarmee niet uit de toon ten opzichte van andere 360°-feedbackinstrumenten. De acceptatie van het 360°-feedbackinstrument is redelijk. Met betrekking tot de soortgenootvaliditeit komt naar voren dat de 360°-feedbackmethode een geheel eigen positie inneemt binnen al langer bestaande assessmentinstrumenten. Hierbij blijft het de vraag 'wat nu precies wordt gemeten'.

Ook uit recenter, zowel nationaal als internationaal, onderzoek (onder meer Atwater et al., 2000) komt hetzelfde beeld naar voren. Jellema (2000) concludeert na onderzoek naar zeven Nederlandse 360°-feedbackinstrumenten dat de instrumenten deels goed scoren, maar zeker ook hun tekortkomingen vertonen. Bovendien wordt de betrouwbaarheid en validiteit weinig onderzocht.

Ondertussen neemt de populariteit van de 360°-feedbackmethode nog steeds toe. Veel organisaties zijn er inmiddels mee bekend en het gebruik groeit gestaag. Ondanks de besproken tekortkomingen van de methodiek is de praktische waarde kennelijk groot. Dit past in de ontwikkeling richting een kenniseconomie, waar steeds meer sprake is van netwerkorganisaties waar individuen in tijdelijke werkverbanden functioneren. In (project)teams wordt men ingezet op basis van persoonlijke competenties die minder functiegebonden zijn. Zowel voor de organisatie als het individu is het belangrijk inzicht te hebben in de eigen talenten en hierin zelfsturend te opereren (Schoemaker, 2001).

Aan persoonlijke ontwikkeling wordt bovendien zowel door individuen als organisaties steeds meer waarde gehecht. De 360°-feedbackmethodiek sluit hier goed op aan door het individu een tool te geven waarmee een completer beeld van zijn/haar functioneren wordt verkregen. Het is in dit opzicht een aanvullend personeelsinstrument. Assessment- en developmentcenters kennen een andere insteek (experts in plaats van directe werkrelaties). De 360°-feedbackmethode is een toegankelijk instrument, dat zich direct vertaalt naar iemands functioneren. Het oordeel van een klant, collega en leidinggevende kan onbetrouwbaar zijn. De verkregen informatie is echter wel van belang en heeft betekenis voor de medewerker en daarmee waarde voor de beoordeelde. Dit in tegenstelling tot het AC, waar de uitslag vaak vrij abstract is en weinig verband heeft met de realiteit van iemands dagelijks werk. Bovendien heeft het vaak ook weinig gevolgen, waardoor de noodzaak voor gedragsverandering vrij gering is.

Daarnaast is het in de dagelijkse praktijk nog zeker geen gemeengoed om elkaar feedback te geven op elkaars functioneren. 360°-feedback kan een aanzet geven om een klimaat te scheppen waarin dit onderdeel wordt van de organisatiecultuur. Het instrument kan op deze manier worden gezien als een hulpmiddel om de communicatie te verbeteren. Daarnaast is feedback een onmisbaar instrument voor een lerende organisatie.

Het voornaamste gebruik van de methode is om een impuls en richting te geven aan persoonlijke ontwikkeling doordat het een beeld geeft van het functioneren zoals dat wordt waargenomen en beleefd door mensen uit de directe werkomgeving. Op een meer geaggregeerd niveau kan het voor team- en organisatieontwikkeling worden gebruikt. Toepassing voor beoordeling en beloning brengt de nodige risico's met zich mee. Een andere toepassing is evaluatie van opleidingen: het effect van een gevolgd opleiding kan worden gemeten.

Voor effectieve toepassing moet echter wel aan een aantal belangrijke voorwaarden worden voldaan, zoals voorbereiding, voldoende communicatie over het doel van inzet van het instrument en voorlichting over de werking van het instrument, waarborging van anonimiteit en vertrouwelijkheid, zorgen dat er constructieve feedback wordt gegeven bijvoorbeeld door training van de beoordelaars, vrijwillige deelname van de feedbackontvanger, zorgen voor de benodigde tijd en procesbegeleiding en het regelen van een follow-up in de vorm van bijvoorbeeld een persoonlijk ontwikkelplan (De Jong, 1998; Jellema, 2000).

Samenvatting

In arbeidsorganisaties worden in toenemende mate zogenoemde 360°-feedbackmethoden toegepast. Onderzoek ernaar is evenwel beperkt en als gevolg daarvan wordt slecht begrepen wat deze techniek meet en wat de effecten ervan zijn. Aan de hand van een empirisch onderzoek naar een 360°-vragenlijst wordt ingegaan op de relatie tussen verschillende groepen beoordelaars, de discrepantie tussen zelfbeoordeling en omstanderbeoordelingen, de structuur van de beoordelingen en de relatie tot persoonlijkheidsvragenlijsten en een capaciteitentest. Psychometrisch (factoranalyse, interne consistentie) blijkt het onderzochte instrument in orde. Er is nauwelijks verband tussen de beoordelaars onderling en tussen de beoordeelde en de beoordelaars. De acceptatie van het instrument blijkt redelijk. Er is echter nauwelijks een relatie tot andere diagnose-instrumenten. Daarnaast wordt ingegaan op de waarde van inzet van het instrument in de praktijk en op de relatie tussen 360°-feed-

back en feedback vanuit assessment centers of vanuit instrumenten in het kader van Total Quality Measurement.

S U M M A R Y

A V I S I O N O N A N D R E S E A R C H O F
T H E 3 6 0 - D E G R E E F E E D B A C K
M E T H O D

360-degree feedback methods are increasingly employed in organizations. However, research is limited and as a result it is poorly understood what this technique measures and what the effects are. On the basis of a psychometric analysis of a 360-degree questionnaire, we examine the relationship between different groups of raters, as well as the discrepancy between self-ratings and the ratings of relevant others, the structure of the ratings, and the relationship with such other instruments as a personality questionnaire and a test for mental ability. From a psychometric point of view (factor analysis, internal consistency) the instrument proves to be all right. However, there is hardly any relationship between the raters and between ratee and raters. The acceptance of the instrument by ratees is reasonable. However, a relationship with other assessment instruments can hardly be found. Furthermore, the practical value of the use of this instrument in a growing number of organizations is mentioned. Results are discussed with respect to literature on (self) ratings, 360-degree methods, and ratings in the context of both assessment centers and Total Quality Measurement.

Noten

- 1 Construct Bedrijfspsychologie te Almere is een psychologisch adviesbureau dat gespecialiseerd is in assessmentcenters, loopbaanadvisering, trainingen en personeelsinstrumenten.
- 2 De Quest is een door de betrokken organisatie gehanteerde persoonlijkheidsvragenlijst bestaande uit 147 items verdeeld over elf schalen (work locus of control, systematische aanpak, stressbestendigheid, omgang met anderen, sociale redzaamheid, omgang als manager, ambitie tot leidinggeven, prestatie-motivatie, sociale presentatie, flexibiliteit, assertiviteit). Personen geven op een driepuntsschaal ('helemaal niet mee eens, beetje mee eens, helemaal mee eens') hun gedragsvoorkeur aan. Uit een onderzoeksrapport van Roemer en Van der Maesen de Sombreff (1993) blijkt dat de Quest-schalen voldoende betrouwbaar zijn (alfa tussen .75 en .90; met uitzondering van omgang als manager: .62) en de predictieve validiteit rond de .20 ligt. Hierbij zijn criteria gehanteerd als de som van de beoordelingen van vier theoretische kennistoet-

sen, de som van een aantal docentenbeoordelingen en de som van een aantal beoordelingen tijdens een stage.

Literatuur

- Antonioni, D. (1994). 'The effects of feedback accountability on upward appraisal ratings'. *Personnel Psychology*, 47, pp. 349-356.
- Atwater, L.E. en F.J. Yammarino (1992). 'Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions?'. *Personnel Psychology*, 45, pp. 141-164.
- Atwater, L.E., P. Roush en A. Fischthal (1995). 'The influence of upward feedback on self- and follower ratings of leadership'. *Personnel Psychology*, 48, pp. 35-59.
- Atwater, L.E., C. Ostroff, F.J. Yammarino en J.W. Fleenor (1998). 'Self-other agreement: does it really matter?'. *Personnel Psychology*, 51, pp. 577-598.
- Atwater, L.E., D.A. Waldman, D. Atwater en P. Cartier (2000). 'An upward feedback field experiment: supervisors' cynicism, reactions, and commitment to subordinates'. *Personnel Psychology*, 53, pp. 275-297.
- Berg, J. van den, L. van den Broek en S. Pijs (1997). *360°-feedback als 'eye-opener'*. Deventer: Kluwer Bedrijfsinformatie.
- Borman, W.C., D.W. Dorsey en L.A. White (1995). 'Effects of ratee task performance and interpersonal factors on supervisor and peer performance ratings'. *Journal of Applied Psychology*, 80, nr. 1, pp. 168-177.
- Breeuwer, S., K.I. van der Zee en G.N. Smit (1998). 'Meerzijdige beoordeling van managers. De 360°-feedbackmethode'. *M&O*, 52, nr. 5, pp. 42-58.
- Carless, S.A., L. Mann en A.J. Wearing (1998). 'Leadership, managerial performance and 360-degree feedback'. *Applied Psychology: An International Review*, 47, nr. 4, pp. 481-496.
- Fleenor, J.W., C.D. McCauley en S. Brutus (1996). 'Self-other rating agreement and leader effectiveness'. *Leadership Quarterly*, 7, nr. 4, pp. 487-506.
- Furnham, A. en P. Stringfield (1994). 'Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation'. *Journal of Occupational and Organizational Psychology*, 67, pp. 57-67.
- Greguras, G.J. en C. Robie (1998). 'A new look at within-source interrater reliability of 360-degree feedback ratings'. *Journal of Applied Psychology*, 83, Nr. 6, pp. 960-968.
- Ghorpade, J. (2000). 'Vijf paradoxen van 360-graden feedback'. *Management Select*, nr. 3, pp. 34-44.
- Jansen, P.G.W. (1993). 'De werking van het assessment center'. *Gedrag en Organisatie*, 6, pp. 10-24.
- Jansen, P.G.W. en D.F.J. Vloeberghs (1998). '360-graden feedback methoden: consequenties voor persoon en organisatie'. *M&O*, 52, nr. 1, pp. 47-62.

- Jellema, F. (2000). 'De kwaliteit van 360-graden feedback instrumenten'. *M&O*, nr. 3, mei/juni, 23-35.
- Jellema, F. (2000). 'Toepassing van 360-graden feedback in Nederlandse organisaties'. *Opleiding & Ontwikkeling*, 7/8, pp. 21-25.
- Johnson, J.W. en K.L. Ferstl (1999). 'The effects of interrater and self-other agreement on performance improvement following upward feedback'. *Personnel Psychology*, 52, pp.271-303.
- Jong, A.M.L. de (1998). *Gebruik van een 360-graden feedback instrument*. Doctoraalscriptie Vakgroep Bedrijfskunde, Faculteit der Economische Wetenschappen en Econometrie Vrije Universiteit, Amsterdam.
- Korman, A.K. (1970). 'Toward an hypothesis of work behavior'. *Journal of Applied Psychology*, 54, pp. 31-41.
- London, M. en J.W. Smither (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research'. *Personnel Psychology*, 48, pp. 803-839.
- Mabe, P.A. en S.G. West (1982). 'Validity of self-evaluation of ability'. *Journal of Applied Psychology*, 67, nr. 3, pp. 280-296.
- 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*, 51, pp. 557-576.
- Reilly, R.R., J.W. Smither en N.L. Vasilopoulos (1996). 'A longitudinal study of upward feedback'. *Personnel Psychology*, 49, pp. 599-612.
- Rietveld, A.D.M. (1997) *De 360° methode als eye-opener*. Doctoraalscriptie, Katholieke Universiteit Brabant, Tilburg.
- Roemer, G. en P. van der Maesen de Sombreff (1993). *Onderzoeksrapport over Brain capaciteitentests en Quest persoonlijkheidsvragenlijst*. Meurs Personeelsadvies, Woerden.
- Sagie, A. en R. Magnezy (1997). 'Assessor type, number of distinguishable dimension categories, and assessment centre construct validity'. *Journal of Occupational and Organizational Psychology*, 70, pp.103-108.
- Schoemaker, M (2001). 'Identiteit in netwerkorganisaties'. *Tijdschrift voor HRM*, 1, pp. 77-94.
- Togt, J.J.W. van der en G.E.P.M. Kemp (1997). *Prestaties sturen door 360° feedback*. Deventer: Kluwer Bedrijfsinformatie.
- Waldman, D.A., L.E. Atwater, en D. Antonioni (1998). 'Has 360 degree feedback gone amok?' *Academy of Management Executive*, 12, Nr. 2, pp. 86-94.
- Walker, A.G. en J.W. Smither (1999). 'A five-year study of upward feedback: what managers do with their results matters'. *Personnel Psychology*, 52, pp. 393-423.
- Wohlens, A.J., M. Hall en M. London (1993). 'Subordinates rating managers: Organizational and demographic correlates of self/subordinate agreement. *Journal of Occupational and Organizational Psychology*, 66, pp. 263-275.

- Woude, M.J.L.C., van der (1995). *Beoordelingsprocessen binnen 360-graden feedback instrumenten*, Doctoraalscriptie Vakgroep Onderwijsstudies, Faculteit der Sociale Wetenschappen, Rijksuniversiteit Leiden.
- Yperen, N.W. en A.P. Baar (2000). 'Betrouwbaarheid en rechtvaardigheid van de 360° feedback beoordelingsmethode'. *M&O*, nr. 4, juli/augustus, 39-51.
- Yammarino, F.J. en L.E. Atwater (1997). 'Do managers see themselves as others see them? Implications of self-other rating agreement for human resources management'. *Organizational Dynamics*, spring, pp. 35-44.