



Stichting NIOC en de NIOC kennisbank

Stichting NIOC (www.nioc.nl) stelt zich conform zijn statuten tot doel: het realiseren van congressen over informatica onderwijs en voorts al hetgeen met een en ander rechtstreeks of zijdelings verband houdt of daartoe bevorderlijk kan zijn, alles in de ruimste zin des woords.

De stichting NIOC neemt de archivering van de resultaten van de congressen voor zijn rekening. De website www.nioc.nl ontsluit onder "Eerdere congressen" de gearchiveerde websites van eerdere congressen. De vele afzonderlijke congresbijdragen zijn opgenomen in een kennisbank die via dezelfde website onder "NIOC kennisbank" ontsloten wordt.

Op dit moment bevat de NIOC kennisbank alle bijdragen, incl. die van het laatste congres (NIOC2023, gehouden op donderdag 30 maart 2023 jl. en georganiseerd door NHL Stenden Hogeschool). Bij elkaar bijna 1500 bijdragen!

We roepen je op, na het lezen van het document dat door jou is gedownload, de auteur(s) feedback te geven. Dit kan door je te registreren als gebruiker van de NIOC kennisbank. Na registratie krijg je bericht hoe in te loggen op de NIOC kennisbank.

Het eerstvolgende NIOC vindt plaats op donderdag 27 maart 2025 in Zwolle en wordt dan georganiseerd door Hogeschool Windesheim. Kijk op www.nioc2025.nl voor meer informatie.

Wil je op de hoogte blijven van de ontwikkeling rond Stichting NIOC en de NIOC kennisbank, schrijf je dan in op de nieuwsbrief via

www.nioc.nl/nioc-kennisbank/aanmelden-nieuwsbrief

Reacties over de NIOC kennisbank en de inhoud daarvan kun je richten aan de beheerder:

R. Smedinga kennisbank@nioc.nl.

Vermeld bij reacties jouw naam en telefoonnummer voor nader contact.

De feilbare assessor Over assessment van competenties

Dr. Wouter Schoonman, Lector Saxion Hogescholen

Samenvatting

Bij assessment van competenties is de assessor de zwakste schakel. In een klein experiment tijdens NIOC 2004 werd dit opnieuw duidelijk. Wat te doen aan de feilbaarheid van de menselijke beoordelaar?

Keywords

Assessment, competenties, assessor, betrouwbaarheid

Copyright

Artikel aangeboden aan: Onderzoek van Onderwijs, januari 2005. Uitgever: Koninklijke Van Gorcum, Postbus 43, 9400AA Assen. www.vangorcum.nl

Inleiding

In het hoger onderwijs en in het bedrijfsleven is assessment een populaire bezigheid. In alle gevallen gaat het om het beoordelen van prestaties van een persoon ten einde een beslissing over die persoon te kunnen nemen. Van die beslissing kan veel afhangen: de student ontvangt studiepunten, de sollicitant krijgt de begeerde baan of de deelnemer ontvangt feedback over verbeterpunten. Het meest kwetsbare onderdeel binnen assessment is de menselijke beoordelaar. De mate waarin hij^j een juist oordeel velt, bepaalt de waarde van het assessment. Hoe werkt dat?

Competenties

Wie assessment zegt, zegt competenties. Hoewel er talloze definities in omloop zijn, is een gangbare omschrijving van een competentie: de mix van kennis, vaardigheden en houding die succesvol beroepsgedrag mogelijk maakt (Schoonman, 2004a). Er zijn honderden competenties geformuleerd, waarbij in het Hoger onderwijs

vaak een onderscheid gemaakt wordt tussen generieke en beroepsspecifieke competenties. Een voorbeeld van een beroepsspecifieke competentie is het kunnen programmeren in een bepaalde computertaal, een generieke competentie is bijvoorbeeld het voeren van een klantgesprek. Voordat een assessment kan plaatsvinden moeten de te meten competenties vastgesteld zijn en moeten gedragsindicatoren bekend zijn: op welk gedrag moet de assessor letten en hoe moet bepaald gedrag gewaardeerd worden. Daarna wordt een levensechte beroepssituatie gecreëerd, waar de kandidaat het bedoelde gedrag kan laten zien.

Klantgesprek

Binnen Saxion worden hulpmiddelen en instrumenten op het gebied van Assessment ontwikkeld. De Kenniskring en het Lectoraat Assessment richt zich daarbij in eerste instantie op gedragsproeven gericht op generieke hbo-competenties. Een van de hulpmiddelen is de simulatie van een Klantgesprek. Het is een sterk veren-

voudigde oefening – dus ook bruikbaar bij lagerejaars studenten) waarbij drie competenties worden gemeten: Klantgerichtheid, Gespreksvaardigheden en Commerciële vaardigheden. Deze competenties zijn in dit geval ontleend aan het curriculum van de opleiding Informatica binnen Saxion, maar komen in diverse curricula ook voor. Voor de eenvoud zijn per competentie slechts drie gedragingen beschreven. De deelnemers (assessoren) worden geïnstrueerd om als volgt te werk te gaan.

- Bestudeer het formulier en maak je het bedoelde gedrag eigen
- Kijk naar de filmpjes en noteer wanneer je bedoeld gedrag ziet
- Maak nog geen oordeel over de kwaliteit van het gedrag
- Beoordeel na de oefening het waargenomen gedrag als +, 0 of –
- Tel daarna het totaal op

Het onderscheid tussen Observeren en Evalueren is van groot belang (Altink et al, 2004). Zonder dit onderscheid is feitelijk geen sprake van assessment (in de zin van gedragsproeven). Het ontwikkelde formulier bij deze gedragsproef heeft de volgende vorm:

Indicatoren	Gezien	Kwaliteit
		+ 0 -
Klantgerichtheid		
Verplaatst zich in de positie van de klant		
Staat open voor de problemen van de klant		
Formuleert het probleem in termen van de klant		
Gespreksvaardigheden		
Heeft luisterende houding		
Maakt samenvattingen		

Stelt open vragen		
Commerciële vaardigheden		
Vertaalt probleem naar een oplossing		
Doet voorstellen voor verdere samenwerking		
Herhaalt gemaakte afspraken		
Totaal		

Figuur 1. Observatie- en beoordelingsformulier Klantgesprek

In de twee filmpjes (Schoonman, 2004b) is een ICT-er en een klant te zien. Zij voeren een gesprek over een opgeleverde applicatie door de



Figuur 2. Fragment uit 'Klantgesprek'

organisatie van de ICT-er. De klant heeft een aantal kritiekpunten zoals die in de afgelopen periode boven water zijn gekomen. In het eerste gesprek reageert de ICT-er afwijzend en verdedigend en lijkt niet geïnteresseerd in de problemen van de klant. In het tweede gesprek gaat het een stuk beter en worden constructieve en positieve oplossingen voor (dezelfde) problemen bereikt. De beide filmpjes zijn bedoeld als oefen- en demonstratiemateriaal.

Workshop

Tijdens het tweejaarlijkse NIOC congres van 2004 (een organisatie van informatica docenten in het hoger onderwijs) is een kleine assessment workshop gehouden. De oefening Klantgesprek werd vertoond en aan de deelnemers werd de mogelijkheid geboden het gedrag van de ICT-er via het formulier uit Figuur 1 te beoordelen. Dit na de instructie zoals die bij deze (en elke) assessment oefening hoort. De deelname was anoniem en

vrijwillig. Uiteindelijk werden negen volledige beoordelingsformulieren verkregen.

Observaties

Als eerste kijken we hoe de negen beoordelaars (beo1 tot en met beo9) de negen gedragingen (indicatoren) in beide gesprekken al dan niet hebben waargenomen (zonder oordeel). De keuze was wel of niet waargenomen (1 of 0). In onderstaande tabellen staan de resultaten:

Gesprek 1	beo1	beo2	beo3	beo4	beo5	beo6	beo7	beo8	beo9
verpl pos	0	0	0	0	0	0	0	0	0
staat open	1	0	0	0	0	0	1	0	1
form klant	0	0	0	0	0	0	0	0	0
luistert	1	0	1	1	1	0	0	1	0
samenvat	0	0	0	0	0	0	0	0	0
open vr	0	0	0	0	0	0	0	0	0
probl opl	1	1	1	1	1	1	1	1	1
samenw	0	0	0	0	0	0	0	0	0
herh afspr	0	0	0	0	0	0	0	0	0
totaal	3	1	2	2	2	1	2	2	2

Tabel 1. Observaties van negen beoordelaars van negen gedragingen (Gesprek 1)

Gesprek 2	beo1	beo2	beo3	beo4	beo5	beo6	beo7	beo8	beo9
verpl pos	0	1	1	1	1	1	0	0	1
staat open	1	1	1	1	1	1	0	1	1
form klant	1	1	1	1	1	1	1	1	1
luistert	1	1	1	1	1	1	1	1	1
samenvat	0	0	0	0	1	0	0	0	0
open vr	1	1	1	1	1	0	1	1	1
probl opl	1	1	1	1	1	1	1	1	1
samenw	1	1	1	1	1	1	1	1	1
herh afspr	0	0	0	0	1	0	0	0	0
totaal	6	7	7	7	9	6	5	6	7

Tabel 2. Observaties van negen beoordelaars van negen gedragingen (Gesprek 2)

In de rijen staan de negen gedragsindicatoren uit het formulier van Figuur 1. In de kolommen staan de beoordelaars. Een eerste conclusie is dat de negen beoordelaars het *globaal* eens zijn over welke gedragingen te zien waren in gesprek 1 en gesprek 2. In gesprek 1 worden door elke assessor zo'n twee gedragingen geobserveerd, in gesprek 2 zijn dat er ruim zes. In gesprek 1 is de overeenstemming deels artificieel. Het gesprek was zo gemaakt dat het bedoelde

gedrag niet of nauwelijks voorkwam. Bij de observaties van gesprek 2 – waar het bedoelde gedrag wel zichtbaar was – bestaan grotere verschillen tussen de beoordelaars. De waargenomen gedragingen lopen uiteen van 5 tot 9 (van de maximaal 9). Het vervelende is dat wanneer een assessor bepaald gedrag niet heeft waargenomen, dat er dan ook geen waardering van dat gedrag kan plaatsvinden. Iets dat er niet is kun je niet waarderen.

Gesprek 1	beo1	beo2	beo3	beo4	beo5	beo6	beo7	beo8	beo9
verpl pos									
staat open	0						-1		-1
form klant									
luistert	1		-1	-1				0	
samenvat									
open vr									
probl opl	0		-1	-1				1	-1
samenw	-1						0		
herh afspr									
totaal	0	0	-2	-2	0	0	-1	1	-2

Tabel 3. Evaluatie van negen beoordelaars van gedrag (Gesprek 1)

Gesprek 2	beo1	beo2	beo3	beo4	beo5	beo6	beo7	beo8	beo9
verpl pos		1	1	1	1	1			1
staat open	1	1	1	1	1	1		1	1
form klant	1	1	1	1	1	1	0	0	1
luistert	1	1	1	1	1	1	1	1	1
samenvat					1				
open vr	1	0	1	1	1		0	1	1
probl opl	1	1	1	1	1	1	1	1	1
samenw	1	0	1		1	1	0	1	1
herh afspr					1				
totaal	6	5	7	6	9	6	2	5	7

Tabel 4. Evaluatie van negen beoordelaars van gedrag (Gesprek 2)

Evaluatie

Conform de instructie is dit assessment proces in twee delen gesplitst: observatie en evaluatie. In het formulier (zie Figuur 1) zijn de rechter kolommen na elkaar en niet gelijktijdig ingevuld. Bij de evaluatie hadden de assessoren de keuze uit +, 0 of -. De betekenis was: positief, neutraal en negatief. Anders gezegd: het zichtbare, geobserveerde gedrag wordt van een waardeoordeel voorzien. In onderstaande tabellen wordt de waardering van de assessoren van *het waargenomen gedrag* weergegeven. Hierbij is een positieve waardering een +1, een neutrale waardering een 0 en een negatieve waardering een -1. In de kolommen staan evaluaties van de negen beoordelaars.

De evaluatie van de assessoren van Gesprek 1 (Tabel 3) is conform verwachting. Er is weinig wenselijk gedrag waargenomen en de beoordeling van wel waargenomen gedrag is overwegend negatief. De gemiddelde evaluatie is -1 wat als 'onvoldoende' vertaald kan worden. Bij de evaluatie van Gesprek 2 wordt veel gedrag positief gewaardeerd, maar de uiteindelijke waardering (laatste rij in Tabel 4) loopt sterk uiteen. Beoordelaar 7 komt tot 2 punten, terwijl beoordelaar 5 het maximum van 9 punten uitdeelt. De beoordelaars zijn het dus eens over zwak gedrag (het bedoelde gedrag wordt niet gezien en voor zover aanwezig negatief beoordeeld), maar verschillen nogal over wat als 'sterk gedrag' moet worden gezien. Voor de kandidaten over wie zo'n oordeel wordt uitgesproken, blijkt het oordeel van een toevallige individuele assessor af te hangen (zie ook Hofstee, 1999). Wanneer assessor 5 een oordeel uitspreekt scoort de kandidaat een '9', wanneer echter assessor 7 dit doet, wordt het een '2'. Hoewel het niet duidelijk is wat een '2' of een '9' betekent, is de discrepantie te hoog. Het mag niet zo zijn dat

een oordeel afhankelijk is van een toevallige assessor.

Analyse

De vraag is natuurlijk hoe het zit met de kwaliteit van deze assessment procedure bij de oefening 'Klantgesprek'. Omdat deze oefening is gebaseerd op vastgestelde competenties binnen een bestaande opleiding, lijkt het realiteitsgehalte niet het probleem. De vraag naar het realisme van de onderscheiden competenties en bijbehorende gedragsindicatoren, lijkt niet aan de orde. Wel is het zo dat er in dit experiment forse verschillen blijken op te treden tussen assessoren. De een ziet niet wat de ander ziet. In dit geval is er duidelijk sprake van een fors verschil tussen bijvoorbeeld beoordelaar5 en beoordelaar7. De gemiddelde correlatie tussen de beoordelaars wat betreft waargenomen gedrag is relatief hoog (een correlatie van ongeveer 0.7). De *evaluatie* door de diverse beoordelaars loopt echter sterk uiteen. Het lijkt erop dat zelfs een eenvoudige gedragsproef na instructie van de assessoren niet automatisch leidt tot een uniform oordeel. De individuele assessor lijkt een (te) belangrijke rol te hebben bij de totstandkoming van het uiteindelijke oordeel. Dit lijkt om diverse redenen onwenselijk.

Conclusie

Het is in dit (kleine) experiment opnieuw duidelijk geworden dat de individuele assessor de zwakste schakel is bij assessment. Er bestaan verschillen tussen assessoren, zowel wat betreft het *observeren* van gedrag als het *evalueren* daarvan. Dit leidt in de praktijk tot willekeurige en niet gestandaardiseerde beoordelingen. Het trainen en certificeren van assessoren lijkt daarom de aangewezen weg om tot eerlijke

en professionele beoordeling van beroepsprestaties te komen.

Referenties

Altink, W., Schoonman, W., Seegers, J. (2004). *Menselijk kapitaal. De ontwikkeling van mensen in organisaties*. Assen: Van Gorcum

Hofstee, W.K.B. (1999). *Principes van beoordeling. Methodiek en ethiek van selectie, examinering en evaluatie*. Lisse: Swets & Zeitlinger

Schoonman, W. (2004a). *Assessment voor en door iedereen. Lectorale rede*. Enschede: Saxion Hogescholen

Schoonman, W. (2004b). *Assessment oefening: het ICT klantgesprek*. DVD + formulier. Enschede: Saxion Hogescholen

Dank

Met dank aan Drs. Piet Hendriks – lid van de Kenniskring Assessment van Saxion – voor zijn commentaar op een eerdere versie van dit artikel.

Noten

¹ Leest u ook 'zij' waar 'hij' staat. Over de sexeverdeling bij assessoren is mij niets bekend.