

*De genormaliseerde methode van werkclassificatie: vijftien jaar later**

door A. Hazewinkel**

1. Inleiding

Ongeveer vijftien jaar nadat de eerste officiële publikatie van de Deskundigencommissie voor Werkclassificatie [3] verscheen, zijn kort na elkaar twee proefschriften gepubliceerd, waarin verslag werd uitgebracht van het zolang verbeide wetenschappelijke onderzoek met betrekking tot de genormaliseerde methode van werkclassificatie [11, 9].***).

Het ligt voor de hand dat een artikel naar aanleiding van dit onderzoek verschijnt in *Mens en Onderneming*, waarin indertijd zó langdurig gediscussieerd is over werkclassificatie, dat de redactie zich tenslotte genoodzaakt zag de gedachtenwisseling te sluiten (zie bijvoorbeeld [15, 6, 10, 17, 12, 16, 5, 7]). Hier zal ze niet weer opgerakeld worden. Er zij slechts op gewezen dat in die discussie door velen de wenselijkheid van wetenschappelijk onderzoek naar voren is gebracht. Met dit onderzoek is nu een begin gemaakt door de vooral historische studie van Lijftogt [11] en de meettechnische studie van schrijver dezes [9]. Het is begrijpelijk dat hiermee niet alle vragen beantwoord zijn. De bedoeling van dit artikel is een kort overzicht te geven van de resultaten van het meettechnische onderzoek en hieraan enkele conclusies te verbinden die voor de praktijk wellicht van enig belang kunnen zijn.

2. Meettheoretische overwegingen

De genormaliseerde methode van werkclassificatie is een variant van een zogenaamde 'point system' [1]. Per *gezichtspunt* wordt nagegaan hoe zwaar de eisen zijn die een functie aan een werknemer stelt. *Kennis* is bij voorbeeld zo'n gezichtspunt of *bewegingsvaardigheid*, maar ook bezwarende omstandigheden (inconveniënten) als *werksfeer* en *inspannende houding*. Het oordeel omtrent deze functie-eisen wordt per gezichtspunt in een cijfer vastgelegd. Om tot één samenvattende einduitslag te komen worden vervolgens de *graderingscijfers* per gezichtspunt vermenigvuldigd met *afweegfactoren* die groot zijn voor belangrijke gezichtspunten (bij voorbeeld 'kennis' heeft een afweegfactor 5, 'zelfstandigheid' heeft een 6) en kleiner voor de minder belangrijke (zo heeft 'houding' een 1).

*) Dit onderzoek werd mede mogelijk gemaakt dank zij een subsidie van het ministerie van Economische Zaken, verstrekt op advies van de C.O.P., aan de Stichting Research Instituut voor de Toegepaste Psychologie aan de Universiteit van Amsterdam.

**) Wetenschappelijk medewerker van het Nederlands Instituut voor Praeventieve Geneeskunde TNO, afdeling Geestelijke Gezondheid.

***) Beide dissertaties zijn inmiddels ook als handelseditie verschenen, [11] als C.O.P.-uitgave, [9] bij Wolters, Groningen, onder de titel: 'Werkclassificatie: een wetenschappelijk instrument?'

Men kan het 'point system' beschouwen als een meetinstrument in de zin waarin een test dat is: bij toepassing verkrijgt men een aantal getallen die een indicatie geven van de eisen die functies aan de werknemer stellen. Het ligt voor de hand zich in zo'n geval af te vragen welke betekenis wij aan deze getallen mogen toekennen. Welke eigenschappen van deze getallen komen overeen met eigenschappen van de gemeten (geclassificeerde) functies? Komen alleen de rangordeverhoudingen van de getallen overeen met die tussen de functies of zijn ook de verschillen tussen de scores (de gegeven puntentotalen) zinvolle gegevens?

Stel twee functies A en B hebben respectievelijk een 7 en een 4 voor het gezichtspunt 'kennis'. Als de graderingscijfers slechts een rangorde weergeven mogen we hieruit concluderen dat functie A *meer kennis* vraagt dan B. We kunnen niet zeggen dat functie A meer verschilt van B dan B van bij voorbeeld een functie C met $3\frac{1}{2}$ voor 'kennis'. De 7 van A had ook een $4\frac{1}{2}$ kunnen zijn en de $3\frac{1}{2}$ voor C een 1; de *rangorde* van de kenniscijfers van de functies A, B en C was dan toch dezelfde gebleven. De deskundigencommissie voor werkclassificatie heeft zich steeds op het standpunt gesteld, dat de graderingscijfers per gezichtspunt inderdaad slechts een rangorde van functies geven (zie [3] en [4]).

Als ook de *verschillen tussen de graderingscijfers* zinvolle gegevens zijn kan men bijvoorbeeld uit de cijfers 7, 4 en $3\frac{1}{2}$ voor respectievelijk A, B en C concluderen, dat het verschil tussen A en B zes maal zo groot is als dat tussen B en C. Dit gegeven kan dan als argument voor loonverschillen gebruikt worden. Men spreekt in zo'n geval van een intervalschaal. Gezien de bewerkingen die uitgevoerd moeten worden op de graderingscijfers per gezichtspunt (vermenigvuldiging met afweegfactoren en optelling der zo verkregen produkten) is deze kwestie vooral van belang wanneer het om de betekenis van de gezichtspuntgradering gaat.

Een meettheoretische analyse heeft aangetoond dat de stelling door de deskundigencommissie geponoerd, als zouden de graderingscijfers niet meer dan een rangorde van functies geven, inhoudt dat de puntentotalen (de sommen van de met afweegfactoren gewogen graderingscijfers) willekeurige getallen zijn die met functie-eisen niets te maken hebben (zie [9] Hoofdstuk II). Willen deze puntentotalen enige betekenis hebben — ten minste iets zeggen over de rangorde der functies — dan moeten de graderingscijfers waaruit ze opgebouwd zijn niet alleen intervalschaalkarakter hebben (zoals de uitkomsten van temperatuurmetingen met een Celsiusthermometer), maar ook nog *gelijksoortige grootheden zijn*. Dit laatste betekent dat de verschillende gezichtspunt-beoordelingen alle uitgedrukt zijn in één en dezelfde dimensie, zoals 'zwaarte', 'geldwaarde' of iets dergelijks.

Als de graderingscijfers voor de verschillende gezichtspunten inderdaad

gelijksoortig zijn en intervalschaalkarakter bezitten (dus per gezichtspunt niet slechts een rangorde van functies geven) zou de situatie analoog worden aan die waarbij men geldbedragen, die uitgedrukt zijn in verschillende valuta's, wil optellen. Men kan ook niet zonder meer een gulden en een dollar optellen, maar als de koersverhoudingen bekend zijn (dollar : gulden = 3,6 : 1) kan men elk bedrag omrekenen in een willekeurig gekozen, eventueel niet bestaande valuta. Daarna zijn de bedragen uitgedrukt in dezelfde munteenheid en is optelling zinvol. Zo zou men 'gelijksoortige' graderingscijfers die slechts in 'munteenheid' verschillen op kunnen tellen na vermenigvuldiging met afweegfactoren, als de verhoudingen tussen die munteenheden redelijk goed geschat kunnen worden.

De vraag of de gezichtspuntgraderingen gelijksoortige grootheden zijn is moeilijk te beantwoorden. Als het antwoord positief zou uitvallen blijft toch nog het bepalen van de afweegfactoren een vrij willekeurige zaak die wetenschappelijk niet oplosbaar is (men zie Hazewinkel [9]). Maar deze problemen worden pas relevant als men kan aannemen dat de gezichtspuntgraderingen intervalschaalkarakter hebben. Moet men in deze de *stelling* van de deskundigencommissie ernstig nemen of moet men op de *praktijk* van de werkclassificatie afgaan, waar de graderingscijfers steeds gehanteerd worden alsof zij minstens intervalgegevens zijn? De verschillen tussen graderingscijfers worden immers beschouwd als zinvolle gegevens: zij corresponderen namelijk met centen uurloon.

De instructie is altijd geweest om per gezichtspunt een rangorde van functies op te stellen, gebruik makend van voorbeeldfuncties. Er werd nooit expliciet bij gezegd: denk er om, binnen één gezichtspunt moet elk punt verschil steeds hetzelfde betekenen (intervalschaal) want de konsekwenties van deze verschillen zijn steeds dezelfde, namelijk een bepaald constant verschil in centen (afhankelijk van de afweegfactor). In de praktijk was natuurlijk iedere taakanalist zich van deze konsekwenties voortdurend bewust, hoewel hij in feite niet met een rechte loonlijn te maken had, waar ieder puntenverschil in het loon tot uitdrukking kwam, maar met een trapvormige loonlijn. Voor de argumentatie maakt dit echter geen verschil.

Moeten we derhalve, gezien de praktijk, van intervalschalen spreken? Ik meen van niet.

De deskundigencommissie heeft in haar oorspronkelijke opzet 22 functies gegradeerd (zie Lijftogt [11]). Zij heeft daarbij zeer waarschijnlijk zonder zich dat te realiseren, althans geprobeerd om de cijfers zodanig toe te kennen dat ze niet alleen een rangorde aangaven. De discussie over bij voorbeeld een kennisgradering kan men zich als volgt voorstellen: Als de functie A 2 punten krijgt en de functie B 3, dan moet functie C toch zeker 5 punten hebben. Dit wijst reeds op meer dan een rangorde. De 5 betekent

hier niet alleen meer dan 3, maar meer boven B (namelijk 2 punten) dan B boven A (1 punt).

Na deze cijfertoekenning per gezichtspunt heeft de commissie gepoogd om met behulp van afweegfactoren tot totaalpunten te komen, die redelijk overeenstemden met de uitkomsten van andere in Nederland gebruikte werkclassificatie-systemen. Voor wat betreft deze 22 functies is zij hierin goed geslaagd. Het was echter zeer onwaarschijnlijk dat toepassing van dezelfde afweegfactoren op andere dan deze 22 functies zou blijven leiden tot een goede overeenkomst met de andere systemen. Toch is juist dit geschied. De verklaring hiervoor werd al door Lijftogt gesuggereerd (op. cit., p. 51) de graderingscijfers zouden wel eens kunnen fungeren ten dienste van de aanpassing aan praktijkverhoudingen. Dit betekent echter dat de graderingscijfers niet meer uitsluitend uitdrukking zouden zijn van de geschatte functie-eisen. In het gunstigste geval zou de rangorde nog wel in overeenstemming kunnen zijn met deze eisen, maar de afstanden tussen de schaalpunten zouden dan toch gewijzigd moeten zijn ten behoeve van deze aanpassing.

Als een en ander zich zo ontwikkeld heeft, blijft er van de oorspronkelijke intervalschaal, indien deze al bij de eerste 22 functies benaderd werd, niet veel meer over.

Het veronderstelde gebruik van de graderingscijfers ten dienste van de aanpassing aan de praktijk heeft dus als konsekventie, dat de punten-totalen — meettheoretisch gesproken — nauwelijks enige betekenis meer hebben. Ze zijn dan immers sommen van graderingscijfers die — in het gunstigste geval — een rangorde van functies per gezichtspunt weergeven en zulke rangorde-getallen zijn niet optelbaar (zie Hazewinkel [9] Hfdst. II, voor een nadere toelichting). Tegelijkertijd verklaart dit gebruik van graderingscijfers hoe het mogelijk is dat deze willekeurige getallen zo goed overeenkomen met de praktijkverhoudingen. Deze overeenkomst wordt namelijk opzettelijk en dankzij het variabele karakter der graderingscijfers, met succes nagestreefd en kan dan ook niet meer als bewijs voor de waarde van de methode of de juistheid van de praktijkverhoudingen aangevoerd worden.

Deze praktijkverhoudingen zijn loonstructuren, die slechts ten dele bepaald worden door de aard van het werk als zodanig (in principe te beschrijven in termen van functie-eisen). Voor een groot deel zijn deze verhoudingen in hun historische groei ontstaan mede onder invloed van machts- en statusverhoudingen, schaarste van bepaalde categorieën van arbeidskrachten of van bepaalde soorten vakkennis e.d.

Theoretisch kan men dus twee groepen factoren onderscheiden, die het loon bepalen: functie-eisen zoals gedefinieerd door de deskundigencom-

missie (zie de omschrijvingen van de gezichtspunten in [3] en [4]) en alle overige factoren, die bij het ontstaan van loonverhoudingen een rol spelen of gespeeld hebben. Deze tweede groep van factoren wordt hier verder aangeduid met *externe factoren*.

Een werkclassificatiemethode die bedoeld is als instrument om functie-eisen te meten, zou uitkomsten moeten geven die door andere factoren dan functie-eisen, bijvoorbeeld de zoëven genoemde externe factoren, niet beïnvloed zijn. Dit geldt te meer wanneer deze factoren tot een waardering leiden, welke afwijkt van die gebaseerd op functie-eisen alleen. Stel bijvoorbeeld dat een functie, die in feite weinig eisen stelt, hoog betaald wordt omdat om een of andere irrelevante reden de functionaris, die deze functie uitoefende, altijd een zekere status had. Men denke aan een bode die bij een hooggeplaatst ambtenaar in en uit loopt, of aan een chauffeur van een minister. In zo'n geval is er sprake van een discrepantie tussen functie-eisen (niet zwaarder dan die van een 'gewone' bode-functie of van een chauffeur van een vuilniswagen) en externe factoren. Als de uitkomst van een werkclassificatie dit hoge loon nu 'waar maakt' kan men spreken van een meettechnisch ongewenste invloed van externe factoren op het gradingsproces.

Lijftogt noemt dit aanpassing aan de praktijkverhoudingen en stelt terecht, dat zij zich meestal aan het oog zal onttrekken. Het is de bedoeling van het in de volgende paragrafen te beschrijven experiment, om iets van de sluier op te lichten die tot nu toe om dit aanpassingsproces gehangen heeft.

3. Het experimentele onderzoek

De hypothese, die in dit experiment getoetst werd, kan als volgt geformuleerd worden. Ervaren taakanalisten, die functies gaan graderen, beschikken over informatie omtrent functie-eisen — interne, meettechnisch gesproken legitieme, informatie — en over informatie omtrent externe factoren. Zij zullen bij het graderen niet alleen rekening houden met de legitieme factoren maar ook informatie omtrent de externe factoren in hun overwegingen betrekken.

Het was niet mogelijk om deze veronderstelling te toetsen aan praktijk-situaties. Daar beschikt immers iedere taakanalist altijd over zowel legitieme als externe informatie (informatie omtrent functie-eisen en omtrent externe factoren). Het is niet mogelijk om sommige taakanalisten normaal te laten werken en aan een controle-groep de informatie omtrent de externe factoren te onthouden. Om deze reden is gekozen voor toetsing van de hypothese door middel van een laboratorium-experiment, waarbij de praktijksituatie zoveel mogelijk nagebootst werd. Ervaren taakanalisten (ver-

der proefpersonen genoemd) ontvingen beschrijvingen van bestaande functies, waaraan in de experimentele conditie gefingeerde informatie omtrent het niveau was toegevoegd. Deze gefingeerde informatie diende als nabootsing van een soort externe informatie. De proefpersonen in de controleconditie kregen deze informatie niet. Op deze wijze was het toch mogelijk om de invloed van externe factoren, zij het dan gefingeerde, los te maken van de invloed van de informatie betreffende de functie-eisen.

3.1 De experimentele opzet

Het experiment is uitgevoerd in drie fasen met twee categorieën proefpersonen en twee groepen functies.

In fase I namen vier werkclassificatiedeskundigen van de onderafdeling Taakanalyse van de afdeling Formatiezaken van het ministerie van Binnenlandse Zaken als proefpersoon aan het experiment deel.

In fase II waren de vier proefpersonen deskundigen van de drie grote vakcentrales. De functies waren voor de helft andere dan in fase I.

In fase III waren de twee proefpersonen eveneens deskundigen uit de vakcentrales. De functies waren echter nu weer dezelfde als in fase I.

Elke proefpersoon kreeg 32 functiebeschrijvingen die ontleend waren aan 'Voorbeelden van gradering ten behoeve van de classificatie van gemeentelijke functies' [2]. De opdracht was de beschreven functies te graderen volgens de genormaliseerde methode zonder gebruik te maken van voorbeeldfuncties zoals gegeven in de publikaties van de deskundigencommissie [3, 4] of van eigen archiefmateriaal.

In iedere fase kreeg de helft van de proefpersonen functiebeschrijvingen waarop een codeletter — A, B, C of D — was aangebracht. Deze proefpersonen kregen de volgende bijkomende instructie:

'Wel wil ik U enig idee geven van het niveau van de functie in kwestie. Tenslotte heeft U ook in de praktijk vrijwel altijd al enig idee van het niveau van een functie voor U met de analyse begint. U zult daarom bij elke functiebeschrijving een klasse-indeling vinden. Hier is een overzichtje van deze klassen met hun betekenissen'.

Zij ontvingen dan het volgende tabelletje:

| Klasse | Totaalscore |
|--------|-------------|
| A | 70—84 |
| B | 55—69 |
| C | 40—54 |
| D | 25—39 |

met de mededeling:

'Dit niveau is uiteraard niet bindend voor U. Wij willen Uw oordeel over deze functies'.

De proefpersonen die functiebeschrijvingen met codeletters en extra in-

structie ontvingen worden verder aangeduid met *experimentele proefpersonen* om ze te onderscheiden van de *controle-proefpersonen* die deze informatie niet kregen.

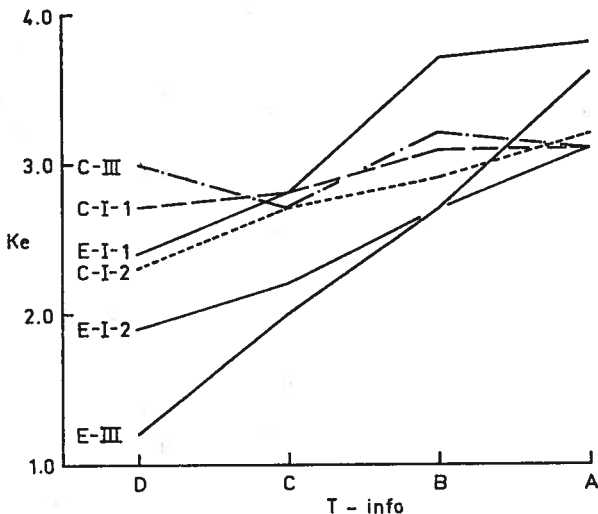
Deze niveau-informatie, verder T-informatie genoemd, was zo gekozen dat er geen enkel verband bestond met de in de praktijk gegeven punten-totalen, waarmee hier bedoeld worden de totaalscores zoals vermeld in de bijlagen van de eerder genoemde publikatie van de Commissie Werkclassificatie Gemeenten [2].

De bedoeling van deze experimentele opzet was om externe factoren, zoals hierboven omschreven, na te bootsen. De invloed van de gefingeerde T-informatie op de graderingen van de experimentele proefpersonen kan nu nagegaan worden door deze graderingen te vergelijken met die van de controle-proefpersonen. Deze vergelijking kan op vele manieren geschieden. Hier is gekozen voor een vergelijking aan de hand van een grafische voorstelling van de resultaten, aan de hand van de resultaten van een variantie-analyse en tenslotte aan de hand van de resultaten van een covariantie-analyse.

3.2 De resultaten

3.2.1 Weergave van de resultaten in grafieken

In de figuren 1 en 2 zijn tegen elkaar uitgezet de T-informatie (lopend van D = laag tot A = hoog) en de gemiddelde Kennisgraderingen van de



proefpersonen van fasen I en III (fig. 1) of de gemiddelde T-scores van de proefpersonen van fase II (fig. 2). De getrokken lijnen hebben betrekking op de graderingen van de experimentele proefpersonen, de onderbroken lijnen op die van de controle-proefpersonen. Voor de motieven om in fasen I en III de invloed van T-informatie op de Kennisgraderingen te bepalen en in fase II die op de T-score raadplege men het oorspronkelijke verslag van het onderzoek [9].

In fig. 1 lopen de experimentele curven (E-I-1, E-I-2 en E-III) alle steil omhoog. Twee van de controle-curven (C-I-1 en C-III) daarentegen lopen vrijwel horizontaal (voor wat betreft de enigszins afwijkende curve van C-I-2 zie men [9], p. 80-81). Op grond van dit duidelijk positieve verband tussen T-informatie en de graderingen in de experimentele conditie en de afwezigheid daarvan in de controle-conditie waar T-informatie niet verschaft werd, kan men voorlopig concluderen dat de T-informatie de graderingen in fase I en III beïnvloed heeft.

Fig. 2 geeft een geheel ander beeld. De afwijkende curve van de experimentele proefpersoon E-II-1 is niet interessant: deze proefpersoon heeft indertijd aan het tot stand komen van de praktijkgraderingen meegewerkt

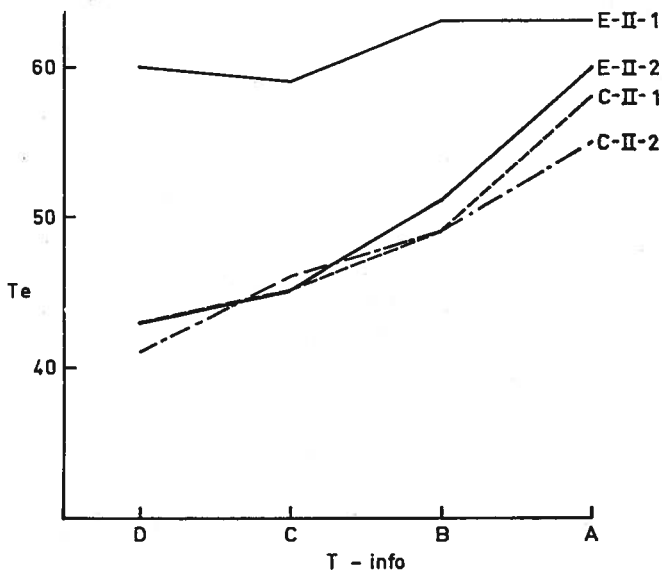


Fig. 2

en zijn graderingen correleren dan ook nog zeer hoog hiermee ($r = 0.93$). Het is begrijpelijk dat hij zich niet door de T-informatie heeft laten beïnvloeden. Veel interessanter is dat de twee controle-curven even steil oplopen als de curve van de tweede experimentele proefpersoon, E-II-2. Alleen afgaand op deze curve zou men geneigd zijn tot een invloed van de T-informatie te concluderen. De proefpersonen in de controle-conditie, die dus geen T-informatie hebben ontvangen, vertonen echter eenzelfde overeenkomst met de T-informatie! Het zou te ver voeren om dit verschijnsel hier in extenso te bespreken. Men zij wederom verwezen naar het volledige verslag [9]. Hier moge volstaan worden met de opmerking dat de T-informatie niet volkomen willekeurig (aselect) over de functiebeschrijvingen verdeeld werd. Alle opvallende discrepanties (een handlangersfunctie met een A bijvoorbeeld) werden vermeden. Men kan concluderen dat er een meningsverschil bestaat tussen de proefpersonen in fase II en de Commissie Werkclassificatie Gemeenten, waardoor een invloed van de T-informatie op de graderingen in deze fase niet vastgesteld kan worden.

3.2.2. De resultaten van een variantie-analyse

De variantie-analyse is een statistische methode om te bepalen of er een relatie bestaat tussen twee of meer variabelen. Hier is een vorm van deze techniek gekozen, de tweevoudige variantie-analyse, waarbij gelijktijdig de sterkte van de relatie geschat kan worden tussen T-informatie en de graderingen van de proefpersonen enerzijds en die tussen de in de praktijk gegeven graderingen en die van de proefpersonen anderzijds. Als men die praktijkgraderingen wil beschouwen als een maat voor de informatie die in de functiebeschrijvingen vervat is, dan kan men op deze wijze de relatie tussen de graderingen van de proefpersonen en de T-informatie (externe informatie) vergelijken met die tussen deze graderingen van de proefpersonen en de 'legitieme' informatie omtrent de functies *).

Het blijkt dat voor beide experimentele proefpersonen van fase I geldt dat het verband tussen de T-informatie en hun kennisgraderingen meer dan de helft zo sterk is als dat tussen de legitieme informatie en deze graderingen. Bij de experimentele proefpersoon van fase III is dit verband met T-informatie zelfs sterker dan dat met de praktijkcijfers voor Kennis. Zoals op grond van de resultaten, zoals weergegeven in fig. 1, te verwachten was, is er geen verband tussen T-informatie en de graderingen in de controle-conditie. Tabel 1 vat deze resultaten samen.

*) Opgemerkt zij hier nog dat in deze functiebeschrijvingen uiteraard ook reeds 'externe factoren' meespelen. Het is echter voor de proefpersonen wel heel moeilijk om deze van de legitieme te onderscheiden. Daarom wordt deze informatie die de functiebeschrijvingen bieden hier in haar geheel legitiem genoemd.

Tabel 1

De samenhang tussen praktijkcijfers voor Kennis en T-informatie enerzijds en de Kennisgraderingen van de proefpersonen in fasen I en III anderzijds, uitgedrukt als percentage 'verklaarde variantie'.

| | experimentele proefpersonen | | | controle-proefpersonen | |
|-----------------|-----------------------------|-----|----------|------------------------|----------|
| | fase I | | fase III | fase I | fase III |
| Praktijkcijfers | | | | | |
| Kennis | 48% | 43% | 21% | 68% | 65% |
| T-informatie | 23% | 29% | 38% | 1% | 0,2% |

De percentages hebben betrekking op wat wel 'verklaarde variante' genoemd wordt.*) Ze zijn vergelijkbaar met de kwadraten van correlatiecoëfficiënten, die ook aangeven welk percentage van de verschillen op een variabele voorspeld kunnen worden, op grond van kennis omtrent een andere, daarmee gecorreleerde variabele.**)

De resultaten van een variantie-analyse op de gegevens van fase II zijn samengevat in tabel 2.

Tabel 2

De samenhang tussen in de praktijk gegeven T-score en T-informatie enerzijds en de T-scores van de proefpersonen anderzijds in fase II, uitgedrukt als percentage 'verklaarde variantie'

| | experimentele proefpersonen | | controle-proefpersonen | |
|--------------------|-----------------------------|--------|------------------------|--------|
| | E-II-1 | E-II-2 | C-II-1 | C-II-2 |
| T-score (praktijk) | 81% | 45% | 45% | 47% |
| T-informatie | 0% | 19% | 14% | 14% |

Het beeld is geheel in overeenstemming met fig. 2. Alleen blijkt hier de samenhang tussen T-informatie en de graderingen van de proefpersonen (hier T-scores) in de controle-conditie iets zwakker dan in de experimentele conditie. De vraag is of dit verschil significant is. Deze vraag doet zich, zij het minder dringend, ook voor met betrekking tot de resultaten van tabel 1. De daar in de experimentele conditie gevonden samenhangen ble-

*) Hays [8] spreekt van 'variance accounted for' of 'relative reduction in uncertainty'. Spitz [13] gebruikt de uitdrukking 'relatieve reductie der onbetrouwbaarheid'.

**) Zie [8], p. 325, 381 en 406 e.v. voor wat betreft de bij een variantie-analyse verklaarde variantie en [13], p. 232 e.v. voor de verklaarde variantie bij correlatiematen.

ken zeer significant te zijn en de samenhangen in de controle-conditie waren niet significant. Toch kan ook dan nog niet geconcludeerd worden dat de verschillen tussen experimentele en controle-conditie statistisch significant zijn. Dit kan wel nagegaan worden door middel van een covariantie-analyse.

3.2.3 De resultaten van een covariantie-analyse

De covariantie-analyse diende om de volgende vraag te beantwoorden: is het verband tussen de in de experimentele conditie gegeven T-informatie en de graderingen van de proefpersonen sterker dan verwacht mag worden op grond van de resultaten in de controle-conditie? Hiertoe werden steeds de gegevens van de experimentele proefpersonen vergeleken met die van zo veel mogelijk vergelijkbare controle-proefpersonen. De gegevens van de proefpersonen E-I-1 en E-I-2 van fase I werden vergeleken met die van de controle-proefpersoon C-I-1, eveneens van fase I. De gegevens van de experimentele proefpersoon E-III werden vergeleken met die van C-III en de gegevens van de experimentele proefpersoon E-II-2 met het gemiddelde van de controle-proefpersonen C-II-1 en C-II-2. Het bleek dat de relatie tussen T-informatie en Kennisgraderingen bij de drie experimentele proefpersonen van fasen I en III ook na correctie op grond van de gegevens in de controle-conditie nog significant waren. De relatie tussen T-informatie en T-score bij de tweede proefpersoon van fase II bleek na correctie niet meer significant te zijn.

3.3 Samenvatting van de resultaten

De onder 3. geformuleerde hypothese is door het experiment bevestigd: externe factoren, zoals door ons door middel van gefingeerde T-informatie nagebootst, beïnvloeden onder bepaalde omstandigheden de graderingen van ervaren taakanalisten op statistisch significante wijze, als men deze vergelijkt met graderingen van vergelijkbare taakanalisten die zonder deze informatie dezelfde functies graderen.

Strikt genomen zijn deze conclusies slechts ondubbelzinnig aangetoond voor enkele proefpersonen, maar dit waren *ervaren taakanalisten* en aange-toond is in ieder geval dat het effect bij zulke proefpersonen kan voorkomen en zeer groot kan zijn. De significantieschattingen hadden steeds betrekking op steekproeven van functies en niet op steekproeven van proefpersonen en generalisatie naar andere proefpersonen is dus op grond van de statistische gegevens niet verantwoord. De grote overeenkomst in de structuur van de resultaten bij zulke uiteenlopende proefpersonen als die in fasen I en III maakt het echter op zijn minst onwaarschijnlijk dat het fenomeen van de invloed van T-informatie specifiek is voor deze, willekeurig gekozen, proefpersonen.

Belangrijk is bovendien dat alle experimentele proefpersonen, zonder uitzondering, ten stelligste hebben beweerd dat zij de T-informatie bewust hebben genegeerd. Zij hebben dus niet gemerkt dat zij rekening hebben gehouden met deze informatie bij het bepalen van hun graderingscijfers. Hoeveel moeilijker zal het voor hen zijn om in de praktijk zich ervan bewust te zijn in hoeverre zij rekening houden met wat wij 'externe factoren' hebben genoemd. In de praktijk zijn de externe factoren immers gegevens die soms wel heel relevant zijn voor de loonvorming, ook al hebben ze niet altijd een direct verband met functie-eisen als zodanig.

4. Conclusies

Op grond van een theoretische analyse werd reeds geconstateerd dat de genormaliseerde methode onmogelijk kon voldoen en dat een overeenkomst met praktijkcijfers niet anders dan kunstmatig ontstaan kon zijn. De resultaten van het experiment sloten hierbij aan. Het bleek dat de taakanalist de methode in feite niet als analytisch meetinstrument gebruikt maar globaal gradeert of, als hij al de methode analytisch hanteert, dan toch door globale indrukken beïnvloed kan worden.

Men kan zich afvragen of dit nu erg is. Dit brengt ons tot de vraag naar de doelstelling van toepassing van een werkclassificatiemethode. Een veel gehoord argument is dat men tegenwoordig niet meer uitkomt met de oude drie-deling ongeschoold-geoeffend-geschoold. Een werkclassificatiemethode zou de mogelijkheid bieden om fijnere onderscheidingen aan te brengen. Uiteraard moeten die onderscheidingen dan wel zinvol zijn en daarom zocht men naar een wetenschappelijk verantwoorde methode die op objectieve wijze fijnere onderscheidingen aan zou kunnen brengen (zie [14], p. 16).

Men heeft getracht deze nauwkeurigheid en deze objectiviteit te bereiken door een analytische methode te kiezen waarbij functies niet globaal maar per gezichtspunt vergeleken worden. Uit het besproken onderzoek is gebleken dat deze analytische methode, aanvechtbaar als ze theoretisch al is, bovendien in de praktijk nauwelijks als analytische meetprocedure fungeert.

Het lijkt of hiermee de methode is afgedaan. Zolang men de methode slechts als meetinstrument ziet is deze conclusie inderdaad juist. In feite wordt hij echter vrijwel niet (meer) als wetenschappelijk instrument beschouwd en toegepast. Er is in de praktijk een gezonde relativering opgetreden in de waardering van de methode. Deze relativiserende visie op de methode komt echter niet voldoende openlijk in de toepassing tot uitdrukking. Men zegt bijvoorbeeld niet openlijk: 'Deze functie komt tien punten te kort om in een hogere loongroep te komen waar deze om bepaalde redenen wel thuishoort, maar het is in dit geval — mede gezien de betrekke-

lijke waarde van de methode — beter om dit te kort maar te negeren'. Men lost zo'n probleem in de praktijk op door hier en daar een halve of hele punt toe te voegen, zodat de hogere loongroep in ieder geval waar gemaakt schijnt te zijn door de methode. Hierdoor steunt men de opvatting dat die punten toch wel erg belangrijk zijn en au sérieux genomen dienen te worden. Ons onderzoek is gericht tegen het gebruik van de methode *als ware het een wetenschappelijk instrument*. Het heeft geen zin of 'lip service' te plegen aan de wetenschappelijke kritiek als men op het ene moment de wetenschappelijke waarde relativeert, en op het andere moment de punten-totaleen toch weer zozeer ernstig neemt, dat plaatsing van een functie in een hogere loongroep alleen aanvaardbaar wordt door toevoeging van extra punten.

Ik ben niet al te optimistisch over het effect dat deze woorden zullen hebben. Reeds in 1950 schreef Van Susante: 'Wanneer een systeem in zijn grondslagen zwakheden vertoont, doet men goed zich bij de toepassing daarvan bewust te zijn. Men krijgt de indruk, dat sommige weinig psychologisch geschoolde technici dit bewustzijn niet bezitten' ([15], p. 269).

Het lijkt nu wel of alleen die 'weinig psychologisch geschoolde technici' de zondebokken moeten worden. Dit is althans door schrijver dezes niet bedoeld. Het is voor allen die in de praktijk beslissingen moeten nemen een moeilijke zaak om te aanvaarden dat er voor deze beslissingen onvoldoende gegevens voorhanden zijn. Een methode die zelfs maar de schijn van wetenschappelijkheid met zich draagt is in zo'n situatie verleidelijk. Ook psychologen kunnen voor een soortgelijke verleiding bezwijken als zij hun uitspraken baseren op de resultaten van al of niet wetenschappelijk verantwoorde methoden.

Het is waarschijnlijk iets gemakkelijker om de beperkingen van bijvoorbeeld een werkclassificatiemethode te aanvaarden als ook de positieve aspecten duidelijk zijn. Over deze positieve eigenschappen heeft weer Van Susante waardevolle opmerkingen gemaakt. Hij schreef: 'Indien men door een werkclassificatie meer dan tot dusver de maatstaven analyseert en op grond daarvan bij de loonvaststelling op deugdelijker materiaal kan onderhandelen, is dit van groot belang' ([15], p. 270). Inderdaad levert de genormaliseerde methode een aantal begrippen met behulp waarvan men beter dan voorheen maatstaven kan analyseren. De ervaring in de laatste vijftien jaar opgedaan met de methode heeft bovendien geleid tot een redelijk goede overeenstemming in de interpretatie van deze begrippen.

Door toepassing van de genormaliseerde methode krijgt men derhalve een gedifferentieerd beeld van de beschreven functies. Door graderingscijfers toe te kennen ontstaat echter op zijn best een rangorde van functies per gezichtspunt. Men zal moeten aanvaarden dat deze rangschikkingen weinig

meer informatie geven dan een globale rangschikking van de functies als geheel door ervaren taakanalisten, zeker zolang de graderingscijfers automatisch via een afweegprocedure die van te voren vaststaat tot een totaal leiden. Wil men per gezichtspunt informatie die min of meer onafhankelijk is van de totaalwaardering volgens de overige gezichtspunten, dan zal men de mechanische berekening van puntentotalen door middel van vaststaande afweegfactoren moeten laten vallen.

Geen enkele exacte combinatie van rangorde-getallen is theoretisch te verdedigen ([9], Hfdst. III), maar een combinatie met van te voren bekende en voor elke functie geldende afweegfactoren brengt bovendien met zich mee dat de taakanalist gemakkelijk kan toewerken naar een einduitkomst die hij zich los van de gezichtspuntwaarderingen als doel had gesteld. Een alternatief is moeilijk te geven, zeker een alternatief dat voor iedere bedrijfssituatie pasklaar zou zijn. Er zullen grote bedrijven zijn waar het eenvoudig om praktische redenen niet mogelijk zal zijn om anders dan langs mechanische weg tot eindwaarderingen te komen, nadat de graderingscijfers zijn vastgesteld. Dan nog is het echter van het grootste belang dat men weet wat men doet: men past *niet* een wetenschappelijk verantwoorde methode toe waarop geen beroep mogelijk zou moeten zijn. Als de uitkomsten in strijd blijken te zijn met de 'common sense' en het rechtsgevoel — ook van de werknemer wiens functie beoordeeld wordt! — moet het mogelijk zijn de werkclassificatie-uitkomst te negeren zonder te goochelen met halve puntjes.

Er zullen echter veel bedrijven zijn waar andere mogelijkheden wel uitvoerbaar zijn. Het is niet ondenkbaar dat een rangschikking van functies, bijvoorbeeld door middel van paarsgewijze vergelijkingen, betere resultaten geeft nadat een werkclassificatierapport is opgemaakt waarin wel graderingscijfers zijn toegekend, maar geen totaalpunten zijn berekend. Het werkclassificatierapport dient dan als neerslag van een grondige analyse van de functies op grond waarvan vergelijking beter mogelijk wordt en ook in de bespreking achteraf beter discutabel. Dit laatste blijft het belangrijkste aspect van elke loonbepaling. Bij loonbepaling gaat het niet alleen om toepassing van een of andere wetenschappelijke theorie of meetmethode. Loonbepaling is in eerste instantie een kwestie van personeelbeleid. Elke theorie is een abstractie en geen werknemer wordt graag als een abstractie behandeld wanneer het gaat om de waardering van zijn werk.

Het spreekt vanzelf en behoeft hier geen verder betoog dat men alleen in abstracties kan denken en spreken, zelfs wanneer het gaat om de beoordeling van dat zogenaamde unieke individu. Het maakt echter een groot verschil of men in zulke delicate zaken bereid is zijn theorie zoveel mogelijk aan te passen aan de te waardenen situatie of dat men met één *genormaliseerde* methode alle functies in Nederland denkt recht te kunnen doen.

Literatuur

1. *Benge, E. J., S. L. H. Burk en F. N. Hay*: Manual of job evaluation; procedures of job analysis and appraisal. New York; London, Harper & Brothers, 1941.
2. *Commissie Werkclassificatie Gemeenten*: Voorbeelden van gradering ten behoeve van de classificatie van gemeentelijke functies, dl. A, B en C. 's-Gravenhage, Vereniging van Nederlandse Gemeenten, z.j.
3. *Deskundigencommissie voor Werkclassificatie V3000*: Ontwerp genormaliseerde methode van werkclassificatie, dl. I: Gezichtspunten; dl. II: Graderingsschema's. Hoofdcommissie voor de Normalisatie in Nederland (HCNN) september 1952.
4. *Deskundigencommissie voor Werkclassificatie NEN 3000*: Genormaliseerde methode van beschrijving en gradering van werkzaamheden ten behoeve van werkclassificatie en andere doeleinden, dl. I: Gezichtspunten; dl. II: Graderingsschema's. Nederlands Normalisatie-Instituut (NEN), april 1959.
5. *Gevers Deynoot, R. A. Th.*: De werkclassificatie in het juiste licht gesteld? Mens en Onderneming 8 (1954) 277-87.
6. *Groot, A. D. de*: Genormaliseerde werkclassificatie. Mens en Onderneming 7 (1953) 401-15.
7. *Groot, A. D. de*: De discussie over werkclassificatie. Mens en Onderneming 8 (1954) 443-7.
8. *Hays, W. L.*: Statistics for psychologists. New York; Toronto; London, Holt, Rinehart and Winston, 1963.
9. *Hazewinkel, A.*: De genormaliseerde methode van werkclassificatie als meetinstrument; een theoretische en experimentele studie. Groningen, Wolters, 1967. (Dissertatie Amsterdam).
10. *Hornstra, L.*: De classificatie als denkfout. Mens en Onderneming 7 (1953) 49-56.
11. *Lijftogt, S. G.*: De genormaliseerde methode van werkclassificatie; waardering en kritiek. Den Haag, Drukkerij Albani, 1966. (Dissertatie Utrecht).
12. *Somermeyer, W. H.*: Enige opmerkingen naar aanleiding van het artikel van prof. A. D. de Groot: 'Genormaliseerde werkclassificatie'. Mens en Onderneming 8 (1954) 90-92.
13. *Spitz, J. C.*: Statistiek voor psychologen, pedagogen, sociologen. Amsterdam, N.V. Noord-Hollandsche Uitgevers Maatschappij, 1965.
14. *Stichting van den Arbeid*: Werkclassificatie als hulpmiddel bij de loonpolitiek. 's-Gravenhage, Stichting van den Arbeid, 1952.
15. *Susante, J. M. van*: Werkclassificatie, Mens en Onderneming 4 (1950) 260-71.
16. *Werkclassificatie*: Kritieken naar aanleiding van het artikel 'De Werkclassificatie als denkfout' van dr. L. Hornstra (Mens en Onderneming VIII-1; januari 1954). Mens en onderneming 8 (1954) 200-6.
17. *Wiegersma, S.*: Enkele problemen betreffende de genormaliseerde werkclassificatie. Mens en Onderneming 8 (1954) 81-9 en 193-9.

