



UvA-DARE (Digital Academic Repository)

De Delphi techniek: de invloed van feedback en informatie op de kwaliteit van groepsbeslissingen

Steenland, E.; Rowe, G.; van der Pligt, J.

Publication date

1992

Published in

Sociale psychologie & beïnvloeding van intermenselijke en gezondheidsproblemen

[Link to publication](#)

Citation for published version (APA):

Steenland, E., Rowe, G., & van der Pligt, J. (1992). De Delphi techniek: de invloed van feedback en informatie op de kwaliteit van groepsbeslissingen. In R. W. Meertens, A. P. Buunk, & P. van Lange (Eds.), *Sociale psychologie & beïnvloeding van intermenselijke en gezondheidsproblemen* (pp. 1-13). (Reeks toegepaste sociale psychologie; No. 6). VUGA.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

WOORD VOORAF

Op 12 en 13 december 1991 werd op de Katholieke Universiteit Brabant te Tilburg voor de derde maal de jaarvergadering van de Associatie van Sociaal-Psychologische Onderzoekers (ASPO) gecombineerd met twee symposia, een over Toegepaste Sociale Psychologie en een over Fundamentele Sociale Psychologie. Deze bundel bevat dertien geselecteerde bijdragen aan het symposium Toegepaste Sociale Psychologie.

Evenals in de voorgaande bundels in deze serie komt een divers geheel van belangrijke toepassingsgebieden van de sociale psychologie aan de orde. De bundel begint met twee onderzoeken over het verbeteren van groepsprestaties. De volgende twee hoofdstukken betreffen vormen van sociale interactie op het werk. Hoofdstuk 5 en 6 behandelen interpersonële relaties. Daarna volgen vier hoofdstukken waarin sociale uitwisseling en sociale vergelijking een belangrijke rol spelen: bij het personeelsverloop in organisaties, bij "burnout", bij overgang naar de WAO en in relaties met vrienden. Hoofdstuk 11 behandelt de relatie tussen onrealistisch optimisme en gedrag. De laatste twee hoofdstukken betreffen vaardigheden en intenties bij bedreiging van de gezondheid.

De dertien hoofdstukken geven theoretisch gefundeerd sociaal-psychologisch inzicht in uiteenlopende problemen en zij leveren een bijdrage aan de oplossing ervan.

Tenslotte een woord van dank aan allen die hebben meegewerkt aan de totstandkoming van de bundel, in het bijzonder aan Tia van Houten voor het persklaar maken van alle teksten.

R.W. Meertens
A.P. Buunk
P. van Lange
B. Verplanken

DE DELPHI TECHNIEK: DE INVLOED VAN HET SOORT INFORMATIE OP DE KWALITEIT VAN GROEPS-BEOORDELINGEN

E. Steenland, G. Rowe & J. van der Pligt

Samenvatting

Dit onderzoek richt zich op de manier waarop verschillende vormen van feedback de kwaliteit van groepsbeoordelingen beïnvloeden. Een summier vorm van feedback blijkt een voldoende voorwaarde te zijn voor het verhogen van de kwaliteit. Een uitgebreidere, van argumenten voorziene, vorm van feedback blijkt in een latere fase van het beoordelingsproces een extra effect op de kwaliteit te sorteren.

I. Inleiding

Veel beoordelingen vinden in groepen plaats. Soms gebeurt dit uit noodzaak, soms kiest men bewust voor een beoordeling in een groep. De idee hierbij is dat uitwisseling van kennis en informatie de kwaliteit van groepsbeoordelingen ten goede komt (zie bijvoorbeeld Hill, 1982). Sociaal psychologisch onderzoek naar de invloed van groepsdynamische processen op de kwaliteit van groepsbeoordelingen heeft echter potentieel versturende factoren aan het licht gebracht (zie voor een overzicht Van de Ven & Delbecq, 1971; Riggs, 1983). Dergelijke nadelige factoren zijn door Steiner (1972) als *process loss* omschreven. Hierbij kan gedacht worden aan de onevenredig grote invloed van groepsleden die dominant en tegelijk weinig deskundig zijn. Dergelijke beïnvloedingsprocessen kunnen eveneens leiden tot extreem negatieve invloeden op de kwaliteit van groepsbeoordelingen, zoals beschreven in het werk van Janis (1982) over *groupthink*.

De behoefte om *process loss* te voorkomen en negatieve groepsdynamische processen uit te schakelen, leidde in de zestiger jaren tot de introductie van de *Delphi techniek* (Dalkey & Helmer, 1963). De Delphi techniek maakt gebruik van vragenlijsten, waarbij de *anonimiteit* van de (inbreng van de) leden wordt gewaarborgd. De techniek beoogt door het verstrekken van *feedback*, de uitwisseling van informatie en kennis te behouden. Karakteristiek voor de Delphi techniek is verder dat de proce-

dure schriftelijk is en dat een aantal opeenvolgende ronden of *iteraties* plaats vindt.

De Delphi techniek is in de Verenigde Staten ontwikkeld met het doel de kwaliteit van lange termijn voorspellingen op het gebied van technologische veranderingen te verhogen (zie Linstone & Turoff, 1975). Tegenwoordig wordt Delphi vaak gebruikt bij de voorspelling van maatschappelijke veranderingen en de bepaling van beleidsstrategieën. In Nederland is de techniek in 1981 door de Stichting der Technische Wetenschappen (STW) geadopteerd om de wetenschappelijke kwaliteit en de mogelijke toegepaste waarde (zogenaamde utilisatie) van onderzoeksvoorstellen objectief te kunnen vaststellen (Van den Beemt & Le Pair, 1991).

De Delphi techniek kent verschillende vormen, die met elkaar gemeen hebben dat een aantal ronden plaats vindt waarin feedback wordt verstrekt. Het soort informatie dat als feedback wordt verstrekt en het totale aantal ronden dat plaats vindt, verschilt echter nogal. Een gebruikelijke manier om de Delphi techniek toe te passen, is als volgt. Stel dat een tiental onderzoeksvoorstellen moet worden beoordeeld op *maatschappelijke relevantie*, waarna uiteindelijk het meest relevante onderzoeksvoorstel zal worden gehonoreerd. In de eerste ronde wordt daartoe een viertal deelnemers gevraagd de onderzoeksvoorstellen op relevantie te beoordelen door de voorstellen aan de hand van een aantal criteria, die gerelateerd zijn aan maatschappelijke relevantie, te rangordenen van 1 t/m 10, waarbij rangnummer 1 dient te worden toegekend aan het meest relevante voorstel. In een tweede ronde wordt de deelnemers gevraagd nogmaals hun rangorde te geven, waarbij ze kunnen beschikken over de rangorde informatie van de andere deelnemers uit de ronde ervoor. Na twee ronden wordt besloten het iteratieve proces te stoppen en wordt de gemiddelde rangorde berekend. Het onderzoeksvoorstel met gemiddeld het laagste rangnummer wordt vervolgens uitgeroepen tot het meest relevante onderzoeksvoorstel. De afwijking tot het gemiddelde wordt hierbij als een indicatie voor de mate van consensus gezien. Op het moment dat geen verdere afname in de standaardafwijking meer optreedt, wordt in de Delphi techniek gesproken van een stabilisatie in het beoordelingsproces.

In de loop der jaren is de Delphi techniek regelmatig ter discussie gesteld. Sackman (1974) vestigt de aandacht op onvoldoende empirische steun voor de in zijn ogen ongebreidelde toepassing van de techniek. Rowe, Wright & Bolger (1991a) wijzen op het ontbreken van systematisch onderzoek naar de onderliggende mechanismen van de techniek en pleiten voor onderzoek naar de belangrijkste karakteristieken van de Delphi techniek, te weten de rol van feedback en iteratie. Het arbeidsintensieve karakter van het iteratieve proces lijkt er namelijk voor te zorgen dat het aantal ronden of iteraties veelal niet boven de twee uitkomt. Tevens wordt feedback vaak in een dusdanig summiere vorm verstrekt, dat

getwijfeld mag worden aan het informatieve gehalte en dientengevolge aan het vermogen substantieel bij te dragen aan een verbetering van de kwaliteit van de beoordeling.

In dit verband zijn twee onderzoeken vermeldenswaard. In een onderzoek van Parenté, Anderson, Myers & O'Brien (1984) werd de bijdrage van feedback aan het iteratieve proces onderzocht. Proefpersonen die aan de hand van zogenaamde consensus-informatie nogmaals werden gevraagd een voorspelling ten aanzien van het al dan niet optreden van een bepaalde gebeurtenis te doen, bleken niet nauwkeuriger te voorspellen dan proefpersonen die het zonder feedback moesten stellen. In een onderzoek van Boje & Murnighan (1982) bleek dat door feedback in de vorm van argumenten te verstrekken de kwaliteit van de uiteindelijke oplossing ook nadelig beïnvloed kon worden. Dit was vooral het geval bij quiz-achtige taken waarbij het moeilijk was om goede argumenten te verzinnen.

De resultaten van Parenté e.a. (1984) en Boje & Murnighan (1982) vormden voor Rowe, Bolger & Wright (1991b) aanleiding om de effectiviteit van iteratie te onderzoeken. In een onderzoek waarin proefpersonen een aantal gebeurtenissen in de nabije toekomst moesten voorspellen, onderzochten zij de invloed van een tweede ronde op de nauwkeurigheid van de voorspelling. Daarbij werd in twee van de drie condities feedback verstrekt die ofwel bestond uit statistische informatie in de vorm van het gemiddelde, de mediaan en de spreiding van de antwoorden ofwel uit inhoudelijke informatie in de vorm van feitelijke argumenten. Hun resultaten wezen uit dat niet het aantal ronden op zich, maar het verstrekken van feedback de nauwkeurigheid van de voorspellingen verbeterde, waarbij echter geen verschil tussen de beide feedback condities (statistische versus inhoudelijke) kon worden aangetoond. Als verklaring voor het feit dat feedback in de vorm van argumenten niet leidde tot een grotere nauwkeurigheid in de voorspellingen, voerden Rowe e.a. (1991b) aan dat de kwaliteit van de door de proefpersonen aangedragen argumenten wellicht onvoldoende was. Zij verklaarden de toename in de nauwkeurigheid van de voorspellingen door te wijzen op normatieve beïnvloedingsprocessen.

Volgens Deutsch & Gerard (1955) spelen normatieve beïnvloedingsprocessen (aanpassing aan de groepsnorm) met name een rol in situaties waarin groepsleden geaccepteerd willen worden. Daarentegen zullen informatiele beïnvloedingsprocessen veeleer een rol spelen in situaties waarin groepsleden voor het bereiken van hun doel afhankelijk zijn van de informatie die door andere groepsleden wordt verstrekt. Verklaringen op grond van deze beïnvloedingsprocessen worden eveneens teruggevonden in de groeps polarisatie literatuur, waarbij normatieve beïnvloedingsprocessen een plaats hebben gekregen in de sociale vergelijkingstheorie en informatiele beïnvloedingsprocessen in de argumentatietheorie. Isenberg (1986) concludeert echter dat het onderscheid tussen beide theorieën minder groot is dan vaak wordt aangenomen. Normatieve en informatiele

beïnvloedingsprocessen zouden in wezen niet los van elkaar kunnen worden gezien, aangezien de groep als belangrijke bron van informatie geaccepteerd moet worden om informatiele beïnvloeding mogelijk te maken. Normatieve beïnvloedingsprocessen zouden dus een voorwaarde zijn voor informatiele beïnvloedingsprocessen. Ook Burnstein & Vinokur (1977) vechten de veronderstelling aan dat beide beïnvloedingsprocessen naast elkaar zouden bestaan. Zij stellen dat elke beïnvloeding in wezen informatieel is.

In lijn met deze ideeën wordt in het huidige onderzoek verwacht dat proefpersonen op basis van summere informatie in staat zullen zijn zelf hun argumenten te genereren, maar dat feedback in de vorm van argumenten tot de beschikking over meer en betere argumenten zal leiden. In dit onderzoek wordt dan ook gekeken naar de invloed van het soort informatie op de *kwaliteit van groepsbeoordelingen* en wordt gebruik gemaakt van een taak waarin de kwaliteit van de beoordeling is vast te stellen. De informatie wordt zowel in een summere statistische vorm (rangorde informatie) als in een uitgebreidere vorm (beargumenteerde rangorde informatie) verstrekt.

Door de keuze van een *argumentrijke* taak zal een, in informatieel opzicht, zo groot mogelijk verschil tussen de beide feedback condities worden gecreëerd. In tegenstelling tot het onderzoek van Rowe e.a. (1991b) wordt het aantal ronden, mede gebaseerd op Brockhoff's (1975) onderzoeksresultaten waarbij stabilisatie in het beoordelingsproces pas na drie ronden bleek op te treden, uitgebreid tot drie.

De belangrijkste voorspelling is dan ook dat de kwaliteit van de oplossing door het verstrekken van feedback zal toenemen, waarbij het verstrekken van een uitgebreidere vorm van feedback tot meer toename in de kwaliteit van de oplossing zal leiden dan het verstrekken van een summere vorm van feedback. Daarbij wordt verwacht dat de invloed van feedback op de kwaliteit van de oplossing toeneemt met het aantal ronden. Teneinde uit te sluiten dat de kwaliteit van de oplossing toeneemt omdat nogmaals over het probleem kan worden nagedacht (Dalkey, 1972), wordt een controle conditie opgenomen. Exploratief zal de relatie tussen de kwaliteit van de oplossing en het vertrouwen in de kwaliteit van de oplossing onderzocht worden, waarbij verwacht wordt dat proefpersonen initieel meer vertrouwen in hun oplossing zullen stellen dan eigenlijk gerechtvaardigd is (Lichtenstein, Fischhoff & Phillips, 1982).

II. Methode

Proefpersonen

Aan het onderzoek deden 60 studenten (34 vrouwen en 26 mannen) van de *Bristol Polytechnic* mee. Zij kregen 6 Engelse ponden betaald voor hun

deelname. De studenten werden willekeurig verdeeld over vijftien groepen van vier personen. Deze vijftien groepen werden vervolgens aselekt toegewezen aan één van de drie condities. Per conditie deden dus twintig proefpersonen verdeeld over vijf groepen mee.

Taak

De taak die gebruikt werd, heet de *NASA Moon Survival Task*. Deze taak vereist het aangeven van de volgorde van belangrijkheid van vijftien items (waaronder zuurstof, water en een magnetisch kompas), die nodig zijn om te kunnen overleven op de maan. De rangorde zoals door NASA experts vastgesteld (met een concordantiemaat van Kendall van .82), dient hierbij als *objectief* criterium om de kwaliteit van de oplossing te bepalen (Hall & Watson, 1970). Het gaat om een taak die vaker in onderzoek naar groepsdynamische processen is gebruikt en door Erffmeyer & Lane (1984) wordt omschreven als een taak die uitnodigt tot het beargumenteren van de oplossing.

Onderzoeksopzet

Gekozen werd voor drie condities, waarbij het soort informatie dat als feedback werd verstrekt, varieerde. In alle drie de condities kregen de proefpersonen hun eigen rangorde en argumenten uit de ronde ervoor als basis-informatie uitgereikt. In conditie 1, de summere feedback conditie, werd daarnaast informatie over de rangorde van drie andere groepsleden als feedback verstrekt. In conditie 2, de uitgebreidere feedback conditie, werd daarnaast zowel informatie over de rangorde van drie andere groepsleden als informatie over hun argumenten als feedback verstrekt. In conditie 3, de controle conditie, werd geen extra informatie als feedback verstrekt; de proefpersonen kregen alleen de basis-informatie uitgereikt. In totaal vonden drie ronden plaats, waarbij in de tweede ronde feedback werd gegeven in de vorm van informatie uit de eerste ronde en in de derde ronde feedback in de vorm van informatie uit de tweede ronde.

Procedure

De drie ronden of iteraties vonden op een maandag, woensdag en donderdag van dezelfde week plaats. De proefpersonen werden op het moment van inschrijven geïnformeerd over het verloop van het onderzoek. Zij konden op alle drie de dagen tussen half tien 's morgens en half drie 's middags terecht voor deelname aan het onderzoek. Het onderzoek duurde in totaal 1.5 uur, zodat per keer op een onderzoeksduur van 30 minuten moest worden gerekend. Het experiment werd geïntroduceerd als een onderzoek naar oplossingsstrategieën. De proefpersonen wisten niet dat de door hen gegeven oplossingen als feedback zouden worden gebruikt in de volgende ronde. Ze werden verzocht niet met elkaar over het onderzoek te praten tot na de laatste ronde. In de tweede en de derde

ronde werd de proefpersonen gevraagd hun beoordeling desgewenst bij te stellen. Zij konden hierbij gebruik maken van de basis-informatie. Veertig van de zestig proefpersonen konden eveneens gebruik maken van informatie van anderen, die ofwel in de vorm van rangorde informatie ofwel in de vorm van beargumenteerde rangorde informatie werd verstrekt. Deze informatie werd met behulp van een samenvattingsformulier gegeven. Tabel 1 toont ter illustratie een (onvertaald) voorbeeld van beargumenteerde rangorde informatie die in de uitgebreidere feedback conditie als feedback werd verstrekt. In de summieere feedback conditie werd uitsluitend rangorde informatie als feedback verstrekt.

Tabel 1. Voorbeeld van het soort informatie dat in de uitgebreidere feedback conditie werd verstrekt

Box of matches

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	rank
				1						1		1		1	frequency

reasons:

- rank 5 : To ignite the heater? Provide light if in darkness.
 rank 11 : I doubt these would even light, because there is very little oxygen; how a heater works I know not either.
 rank 13 : It was difficult to rank this, as we are not told if it is required for running the heater and joining the rope together (if not, then its only practical use would be in cooking).
 rank 15 : Oxygen free environment, matches will not ignite.

Afhankelijke variabelen

De belangrijkste afhankelijke variabele was de accuratesse of kwaliteit van de oplossing. Deze werd berekend door de absolute afwijkingsscores van de individuele rangorde tot de door NASA vastgestelde rangorde op te tellen. Zo leverde een individuele rangorde van zuurstof (1), water (2), sterrenkaart (4), eten (3), FM radio (8), touw (10), verbandtrommel (12), parachute (14), reddingsboot (7), lichtpijlen (13), pistool (6), melkpoeder (9), verwarmingselement (11), magnetisch kompas (5) en lucifersdoosje (15), na vergelijking met de door NASA vastgestelde rangorde van respectievelijk 1 t/m 15 een totale afwijkingsscore van 44 op. Een afname in de score impliceert een toename in de accuratesse van de oplossing. Het vertrouwen in de oplossing werd gemeten door een percentage tussen 0% en 100% aan te geven. Dit percentage betrof het vertrouwen dat gesteld

werd in de keuze van de vijf belangrijkste items nodig voor overleving op de maan (de zogenaamde *Top Vijf*).

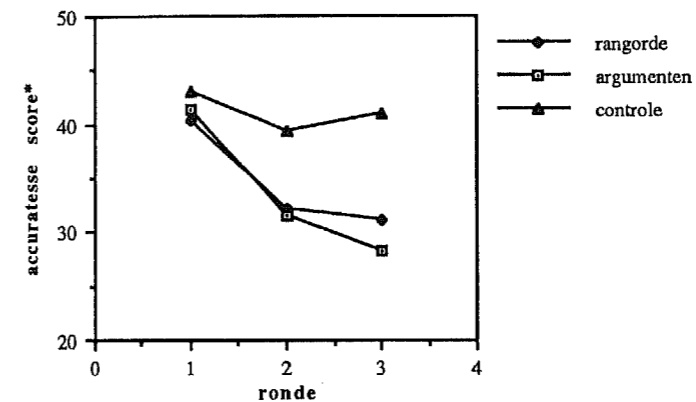
III. Resultaten

Controle op de randomisatie

Variantie-analyse uitgevoerd op de accuratesse scores van de eerste ronde laat geen significante verschillen zien tussen de condities.

Accuratesse

Om de hypothese te toetsen dat een toename in de kwaliteit van de oplossing kan worden toegeschreven aan de invloed van het soort informatie, zijn de gemiddelde scores berekend. In de summieere feedback conditie wordt een gemiddelde score van 40.4 in de eerste ronde, 32.2 in de tweede ronde en 31.1 in de derde ronde gevonden. In de uitgebreidere feedback conditie zijn deze scores respectievelijk 41.3, 31.6 en 28.3 en in de controle conditie respectievelijk 43.1, 39.4 en 40.9. In Figuur 1 worden deze resultaten grafisch weergegeven.



* in de grafiek staat een verlaging van de accuratesse score voor een verhoging van de accuratesse.

Figuur 1. Gemiddelde accuratesse per ronde

De belangrijkste hypothese wordt getoetst in een 3 (aard van de feedback) x 3 (aantal ronden) variantie-analyse met herhaalde metingen.

Allereerst wordt een hoofdeffect voor feedback ($F(2,114)=24.8, p<.001$) gevonden. Figuur 1 laat zien dat in de feedback condities de accuratesse hoger ligt dan in de controle conditie. Tukey's HSD procedure toont dat in de tweede ronde geen significante verschillen tussen de condities worden gevonden. In de derde ronde verschilt het gemiddelde van de controle conditie significant ($p<.05$) van de gemiddelden in de feedback condities. Daarnaast wordt een hoofdeffect voor het aantal ronden ($F(2,57)=3.7, p<.05$) vastgesteld, waaruit blijkt dat de accuratesse toeneemt met het aantal ronden. Deze beide hoofdeffecten worden echter gekwalificeerd door het verwachte interactie-effect ($F(4,114)=3.2, p<.05$), waaruit blijkt dat de verschillen tussen de condities terug te voeren zijn op de informatie die wel in de feedback condities is verstrekt en niet in de controle conditie.

Paarsgewijze vergelijkingen laten zien dat in de controle conditie geen significante verschillen aantoonbaar zijn tussen de accuratesse scores in de verschillende ronden. In de feedback condities daarentegen wel. In beide feedback condities wordt een significant verschil in accuratesse gevonden tussen de eerste en de tweede ronde: $t(1,19)=3.79, p<.001$ in de summier feedback conditie en $t(1,19)=4.32, p<.001$ in de uitgebreidere feedback conditie. Alleen in de uitgebreidere feedback conditie wordt echter een significant verschil tussen de tweede en derde ronde gevonden: $t(1,19)=3.72, p<.001$. Deze extra toename in accuratesse blijkt echter niet groot genoeg te zijn om te zorgen voor een significant verschil in accuratesse tussen de beide feedback condities in de derde ronde ($t(1,38)<1, n.s.$).

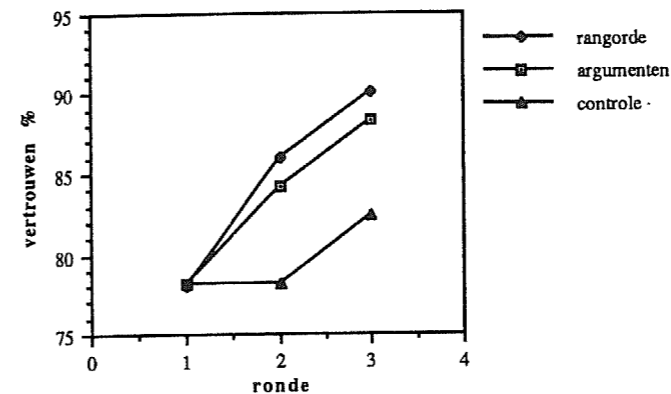
Kijken we naar de standaardafwijkingen in de opeenvolgende ronden, dan valt op dat deze zoals verwacht in de feedback condities meer zijn afgenomen dan in de controle conditie. In de controle conditie is de standaardafwijking 12.8 in de eerste ronde, 12.5 in de tweede ronde en 11.4 in de derde ronde. In de summier feedback conditie respectievelijk 11.9, 10.4 en 9.7 en in de uitgebreidere feedback conditie 10.8, 10.1 en 8.5. Deze afname illustreert een grotere mate van consensus in de uitgebreidere feedback conditie.

Top Vijf

In de summier feedback conditie wordt een gemiddeld vertrouwen in het aantal juiste items in de Top Vijf gevonden van 78.0% in de eerste ronde, 85.9% in de tweede ronde en 90.0% in de derde ronde. In de uitgebreidere feedback conditie worden gemiddelden van respectievelijk 78.2%, 84.2% en 88.3% gevonden en in de controle conditie van respectievelijk 78.2%, 78.2% en 82.4%. Deze resultaten worden in Figuur 2 grafisch weergegeven.

Variante-analyse met herhaalde metingen levert uitsluitend een hoofdeffect voor het aantal ronden op ($F(2,114)=12.3, p<.001$). Uit de gemiddelden komt naar voren dat in alle drie de condities het vertrouwen toeneemt met het aantal ronden.

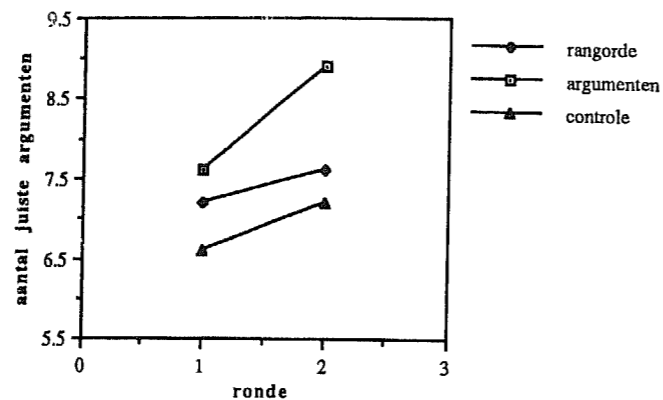
Het aantal (objectief) juiste items in de Top Vijf blijkt in de eerste ronde significant te correleren met het vertrouwen ($r=.50, p<.001$). In de tweede en derde ronde is deze correlatie niet significant (respectievelijk $r=.29$ en $r=.15$).



Figuur 2. Gemiddeld vertrouwen in de Top Vijf per ronde

Kwaliteit van de argumenten

Gekeken is naar verschillen in de kwaliteit van de gebruikte argumenten tussen de drie condities. De argumenten die sterk overeenkwamen met de argumenten van NASA werden gescoord als 1 en de argumenten die enige overeenkomst vertoonden met de argumenten van NASA als een 1/2. Hiertoe werden alle argumenten door twee onafhankelijke beoordelaars gescoord. Cohen's kappa berekend over de scores van 24 proefpersonen leverde een significante overeenstemmingsmaat op van .86 ($p<.001$) in de eerste ronde en van .83 ($p<.001$) in de tweede ronde. Besloten werd het gemiddelde van de scores van beide beoordelaars te gebruiken voor het berekenen van het totale aantal juiste argumenten per persoon. Omdat de proefpersonen in de derde ronde geen argumenten meer hoefden te geven, hebben de resultaten betrekking op de eerste en de tweede ronde. Dit levert in de summier feedback conditie een gemiddeld aantal juiste argumenten van 7.2 in de eerste ronde en 7.6 in de tweede ronde op. In de uitgebreidere feedback conditie wordt een gemiddeld aantal juiste argumenten gevonden van respectievelijk 7.6 en 8.9 en in de controle conditie van respectievelijk 6.6 en 7.2. In Figuur 3 worden deze resultaten grafisch weergegeven.



Figuur 3. Gemiddeld aantal juiste argumenten per ronde

Variantie-analyse laat in de eerste ronde geen verschil zien tussen de drie condities. Een significant hoofdeffect voor condities ($F(2,57)=6.63$, $p<.01$) en een significant hoofdeffect voor rondes ($F(1,57)=29.14$, $p<.001$) wordt met behulp van variantie-analyse met herhaalde metingen vastgesteld, waarbij de hoofdeffecten gekwalificeerd worden door een significant interactie-effect ($F(2,57)=3.58$, $p<.05$). Tukey's HSD procedure laat zien dat in de tweede ronde het gemiddelde in de uitgebreidere feedback conditie significant ($p<.05$) verschilt van zowel het gemiddelde in de summere feedback conditie als van het gemiddelde in de controle conditie.

IV. Discussie

De resultaten van het onderzoek ondersteunen de hypothese dat het verstrekken van feedback leidt tot een toename in de kwaliteit van de oplossing. Er wordt echter geen significant verschil tussen de beide feedback condities gevonden. Wel laten de resultaten zien dat de kwaliteit van de oplossing in de latere fase van het beoordelingsproces meer is toegenomen indien uitgebreidere feedback wordt verstrekt dan indien summere feedback wordt verstrekt.

De resultaten lijken niet verklaard te kunnen worden door een verschil in informatiele beïnvloeding. Zowel informatie in de vorm van rangorde

informatie als informatie in de vorm van beargumenteerde rangorde informatie draagt bij aan het verhogen van de kwaliteit van de oplossing. Aangezien alleen in de uitgebreidere feedback conditie een duidelijke toename van het aantal juiste argumenten is geconstateerd, wordt de veronderstelling dat de kwaliteit van de oplossing af zou hangen van het aantal juiste argumenten onvoldoende ondersteund.

De resultaten stemmen in grote lijnen overeen met de onderzoeksresultaten van Rowe e.a. (1991b). De resultaten van het huidige onderzoek laten echter zien dat indien meer tijd en aandacht wordt besteed aan het iteratieve proces in de vorm van meer rondes en een uitgebreidere vorm van feedback, dit de kwaliteit van groepsbeoordelingen ten goede kan komen.

De resultaten met betrekking tot de relatie tussen het aantal juiste items en het vertrouwen in de oplossing zijn niet in overeenstemming met Lichtenstein e.a. (1982), aangezien het vertrouwen in de kwaliteit van de oplossing in de eerste ronde wel op overtuigende wijze blijkt te correleren met het aantal juiste items in de Top Vijf. Deze resultaten lijken veeleer in overeenstemming met de onderzoeksresultaten van Koriat, Lichtenstein & Fischhoff (1980) en Sniezek, Paese & Switzer (1990), die aantoonde dat afwegingsprocessen en inzicht in vóór- en tegenargumenten kunnen leiden tot een afname in de overschatting van het vertrouwen in de kwaliteit van de eigen oplossing. In dat opzicht kan de *NASA Moon Survival Task* als afwegingsproces in zichzelf worden gezien.

Samengevat heeft het onderzoek tot meer inzicht in de invloed van het soort informatie op de kwaliteit van groepsbeoordelingen geleid. Het verstrekken van feedback in een summere vorm blijkt een noodzakelijke en tevens voldoende voorwaarde te zijn om de kwaliteit van groepsbeoordelingen te verhogen. Vertaald naar de praktijk lijken de onderzoeksresultaten de keuze van een summere vorm van feedback bij de toepassing van de Delphi techniek zodoende te rechtvaardigen.

Hoewel in het huidige onderzoek de kwaliteit van de oplossing in gunstige zin wordt beïnvloed, is door Rowe e.a. (1991b) gesuggereerd dat nadelige beïnvloedingsprocessen kunnen optreden indien een meerderheid op voorhand een onjuist standpunt inneemt. Dergelijke beïnvloedingsprocessen zouden verder uitgezocht kunnen worden in onderzoek naar de invloed van het soort informatie op de *richting van de beïnvloeding*. Toekomstig Delphi onderzoek zou zich moeten richten op de invloed van *onjuiste* versus *juiste* informatie op de kwaliteit van groepsbeoordelingen.

Literatuur

Boje, D.M., & Murnighan, J.K. (1982). Group confidence pressures in iterative decisions. *Management Science*, 28, 1187-1196.

- Brockhoff, K. (1975). The performance of forecasting groups in computer dialogue and face to face discussion. In H.A. Linstone & M. Turoff (Eds.), *The Delphi Method: Techniques and Applications*. London: Addison-Wesley.
- Burnstein, E., & Vinokur, A. (1977). Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology*, 13, 315-332.
- Dalkey, N.C. (1972). Delphi. In J.P. Martino (Ed.), *An introduction to technological forecasting*. New York: Gordon and Breach, Science Publishers, Inc.
- Dalkey, N.C., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9, 458-467.
- Deutsch, M., & Gerard, H.B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, 51, 629-636.
- Erfmeyer, R.C., & Lane, I.M. (1984). Quality and acceptance of an evaluative task: The effects of four group decision-making formats. *Group & Organization Studies*, 9, 509-529.
- Hall, J., & Watson, W.H. (1970). The effects of a normative intervention on group decision-making performance. *Human Relations*, 23, 299-317.
- Hill, G.W. (1982). Group versus individual performance: Are N+1 heads better than one? *Psychological Bulletin*, 91, 517-539.
- Isenberg, D.J. (1986). Group polarization. A critical review and meta-analysis. *Journal of Personality and Social Psychology*, 50, 1141-1151.
- Janis, I.L. (1982). *Groupthink*. Boston: Houghton Mifflin Company.
- Koriat, A., Lichtenstein, S. & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107-118.
- Lichtenstein, S., Fischhoff, B., & Phillips, L.D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A.Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
- Linstone, H.A., & Turoff, M. (1975). *The Delphi Method: Techniques and Applications*. London: Addison-Wesley.
- Parenté, F.J., Anderson, J.K., Myers, P., & O'Brien, T. (1984). An examination of factors contributing to delphi accuracy. *Journal of Forecasting*, 3, 173-182.
- Riggs, W.E. (1983). The Delphi Method: An experimental evaluation. *Technological Forecasting and Social Change*, 23, 89-94.
- Rowe, G., Wright, G., & Bolger, F. (1991a). Delphi: A re-evaluation of research and theory. *Technological Forecasting and Social Change*, 39, 235-251.
- Rowe, G., Bolger, F., & Wright, G. (1991b; aangeboden ter publicatie). The Delphi method: An investigation of richness of feedback and change in majority/minority opinions.
- Sackman, H. (1974). *Delphi critique*. Lexington, MA: Lexington Books.
- Snizek, J.A., Paese, P.W., & Switzer III, F.S. (1990). The effect of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes*, 46, 264-282.
- Steiner, I.D. (1972). *Group process and productivity*. New York: Academic Press.
- Van de Ven, A., & Delbecq, A.L. (1971). Nominal versus interacting group processes for committee decision-making effectiveness. *Academic Management Journal*, 14, 203-212.
- Van den Beemt, F.C.H.D., & Le Pair, C. (1991). Grading the grain: Consistent evaluation of research proposals. *Research Evaluation*, 1, 3-10.