

4.067.2:351.054:43/76:72:4.05:62:336.6:159.9

93-327

OORDEEL OVER KUNST

Kwaliteitsbeoordelingen in de beeldende kunst

Boekmanstichting - Bibliotheek
Herengracht 415
1017 EP Amsterdam
Tel. 6243739

Auteurs:

Paul Hekkert en Piet van Wieringen

**Dit onderzoek is uitgevoerd door de Vakgroep
Psychologie, Faculteit der Bewegingswetenschappen
van de Vrije Universiteit, Amsterdam, in opdracht
van het Directoraat-Generaal voor Culturele Zaken,
van het Ministerie van Welzijn, Volksgezondheid
en Cultuur.**

Rijswijk, mei 1993

When critics disagree the artist is in accord with himself.

Oscar Wilde

Everyone wants to understand art.

Why not try to understand the song of a bird?

Pablo Picasso

Boekmanstichting - Bibliotheek
Houttuinplein 415
1017 EP Amsterdam
Tel. 6243739

Inhoudsopgave

Beeldbeoordeling - Bibliotheek
nummer 415
1017 EP Amsterdam
Tel. 6243739

Voorwoord ... 7

I Inleiding ... 9

I.1	Het kunstbeleid in Nederland	9
I.1.1	Legitimering van het beleid	9
I.1.2	Uitvoering van het beleid	10
I.2	Uitvoering van het beleid bij het Fonds BKVB	10
I.2.1	Beoordelingsprocedure bij de commissie Individuele Subsidies	11
I.3	Het esthetisch oordeel: schoonheid of kwaliteit?	11

II Onderzoek I: oordelen van deskundigen ... 15

II.1	Doelstelling	15
II.2	Selectie van de experts	15
II.2.1	Achtergrond van de experts	15
II.3	Selectie van het beoordelingsmateriaal	16
II.3.1	Achtergrond van de aanvragers	17
II.4	Methode	18
II.4.1	Criteria	18
II.4.2	Opzet van het onderzoek	20
II.4.3	Individuele oordelen	21
II.4.4	Commissie-oordelen	21
II.4.5	Verantwoording van de opzet	22

III Resultaten en conclusies van onderzoek I ... 23

III.1	Test-hertest betrouwbaarheid	23
III.1.1	Betrouwbaarheid van de beoordelaars	23
III.1.2	Betrouwbaarheid van de schalen	23
III.2	Overeenstemming tussen de beoordelaars	24
III.3	Relatie tussen de criteria en het kwaliteitsoordeel	25
III.4	Het eindoordeel van de commissies m.b.t. honorering van de aanvragen	26
III.4.1	Samenhang tussen het individuele kwaliteitsoordeel en het gezamenlijke eindoordeel	26
III.4.2	Overeenstemming tussen de eindoordelen	26
III.4.3	Vervanging van het eindoordeel	27
III.4.4	Bepaling van het benodigde aantal beoordelaars	28
III.5	Vergelijking met commissies van het Fonds BKVB	28
III.6	Verschillen tussen oordelen van kunstbeschouwers en kunstenaars	28

IV	Onderzoek II: oordelen van geïnteresseerde leken ...	31
IV.1	Doelstelling	31
IV.2	Selectie van de beoordelaars	31
IV.3	Selectie van het beoordelingsmateriaal	32
IV.4	Methode	32
V	Resultaten en conclusies van onderzoek II ...	33
V.1	Overeenstemming tussen de beoordelaars	33
V.2	Relaties tussen het kwaliteitsoordeel en de overige criteria	34
V.3	Vergelijking met de oordelen van experts	35
V.3.1	Overeenstemming tussen geïnteresseerde leken en experts	35
V.3.2	Verschillen in scores op de criteria	36
V.3.3	Overeenstemming met beslissingen van het Fonds BKVB	36
VI	Discussie ...	39
VI.1	Verschillen tussen oordelen van experts en geïnteresseerde leken	39
VI.2	Vergelijking van twee door experts uitgevoerde beoordelingsprocedures	39
VI.3	Vervanging van het kwaliteitsoordeel door een oordeel gebaseerd op criteriumscores	41
VI.4	Generaliseerbaarheid van de resultaten	41
VII	Aanbevelingen ...	43
VII.1	Kanttekeningen bij de aanbevelingen	43
VII.2	Voordelen van de aanbevolen procedure	45
	Noten ...	47
	Referenties ...	51
	Bijlage 1 ...	53

Voorwoord

In opdracht van het ministerie van WVC hebben wij in 1991/1992 twee onderzoeken uitgevoerd naar kwaliteitsbeoordelingen in de beeldende kunst.

In het eerste onderzoek ging het om oordelen van experts. Hierin kwamen vragen aan de orde als: Hoe consistent zijn zij in hun oordelen? Welke criteria spelen daarbij een rol? In hoeverre zijn zij het met elkaar eens? Door de keuze van het te beoordelen materiaal en de wijze waarop het onderzoek was opgezet was het mogelijk relaties te leggen met de praktijk binnen beoordelingscommissies van het Fonds voor beeldende kunsten, vormgeving en bouwkunst. Over dit onderzoek is in een rapport getiteld "Kwaliteit als selectie criterium: Een empirisch onderzoek naar kwaliteitsbeoordelingen in de beeldende kunst" verslag gedaan¹.

In het tweede onderzoek werd vervolgens nagegaan in hoeverre personen die over minder expertise op het terrein van de beeldende kunst beschikken, maar daarin wel geïnteresseerd zijn, ten aanzien van hetzelfde materiaal tot overeenkomstige, dan wel afwijkende oordelen komen dan de experts uit het eerste onderzoek. De bevindingen van dit tweede onderzoek zijn weergegeven in een rapport met als titel "Geïnteresseerde leken versus experts: Een vergelijkend empirisch onderzoek naar kwaliteitsbeoordelingen in de beeldende kunst"².

In het thans voorliggende rapport worden beide onderzoeken in meer beknopte vorm verslagen en op elkaar betrokken. Na een inleidend eerste hoofdstuk wordt in hoofdstuk II en III verslag gedaan van het onderzoek bij deskundigen op het gebied van de beeldende kunst. In de hoofdstukken IV en V komt vervolgens het onderzoek aan de orde bij geïnteresseerde leken. Hoofdstuk VI bevat een discussie van de resultaten uit beide onderzoeken waarbij de bevindingen bij de experts centraal staan. Naar aanleiding van die resultaten wordt in het slothoofdstuk VII een aantal aanbevelingen gedaan die kunnen bijdragen tot een optimalisering van de procedure in beoordelingscommissies.

Voor een uitgebreide behandeling van de theoretische achtergrond, alsmede voor een volledige bespreking van alle uitgevoerde analyses en gevonden resultaten verwijzen wij de lezer naar de oorspronkelijke rapporten.

Tot dusverre is een door experts uitgesproken oordeel over kwaliteit nog maar vrij zelden tot onderwerp van empirische studie gemaakt. Hofstee³, die ten aanzien van beoordelingen van de kwaliteit van subsidie-aanvragen voor wetenschappelijk onderzoek eenzelfde probleem signaleert, ziet als oorzaak daarvoor onder meer de arbeidsintensiviteit van een empirische aanpak. Daarnaast "hebben beoordelingscommissies in het algemeen niet de neiging, gegevens met betrekking tot het beoordelingsproces aan de grote klok te hangen"⁴. Het lijkt ons toe dat Hofstee's opmerkingen ook van toepassing zijn op de schaarste aan onderzoek naar het werk van beoordelingscommissies in de kunst.

Met dit rapport hopen wij een eerste stap te zetten om in deze lacune te voorzien.

Het onderzoek is mede tot stand gekomen dankzij de bereidwillige medewerking van de volgende instanties en instellingen:

- Stichting Fonds voor beeldende kunsten, vormgeving en bouwkunst
- Praktijkburo Beeldende Kunstopdrachten
- Stichting Kunstzinnige Vorming Rotterdam
- Federatie van Amateur Beeldende Kunstenaarsverenigingen
- Stichting Beeldende Amateurkunst
- Stichting Beeldende Kunst

Daarnaast zij wij Vladimir Konečni (Universiteit van Californië, San Diego), Fred Wagemans (Museum Fodor, Amsterdam) en Geurt Imanse (Stedelijk Museum, Amsterdam) erkentelijk voor hun deskundige adviezen bij de opzet van het onderzoek.

Bij de uitvoering van het onderzoek was de praktische hulp van de studenten Margit Biemans en Marcel Dekker van groot belang.

Verder willen wij uiteraard de kunstenaars bedanken voor de toestemming om hun dia's te gebruiken in dit onderzoek.

Tenslotte zijn we de beoordelaars dankbaar voor hun inzet en enthousiasme bij het beoordelen van de honderden dia's.

Paul Hekkert &
Piet van Wieringen

Amsterdam, 24 maart 1993

I Inleiding

I.1 Het kunstbeleid in Nederland

Ondanks de heftige discussies die voortdurend plaatsvinden over de wijze waarop het kunstbeleid van de overheid gestalte krijgt, wordt het uitgangspunt dat een dergelijk beleid noodzakelijk is slechts zelden aangevochten. Ook al acht driekwart van de Nederlandse bevolking, volgens een enquête van het *Sociaal en Cultureel Planbureau*, bezuinigingen op de kunsten wenselijk⁵, toch zijn slechts weinig politici van mening dat de overheid het voortbestaan van de kunsten over moet laten aan principes van het marktmechanisme.

Vele kunstvormen en individuele kunstenaars zijn voor een groot deel afhankelijk geworden van subsidies en andere vormen van overheidssteun en zouden zich op een vrije markt niet voldoende kunnen handhaven. Wat maakt de kunst nu zo belangrijk dat de overheid daar jaarlijks miljoenen aan spendeert?

I.1.1 *Legitimering van het beleid*

Bij de beweegredenen die worden aangedragen ter legitimering van het kunstbeleid kan een onderscheid worden gemaakt tussen twee soorten van motieven of belangen⁶. Het eerste is vooral een ideologisch motief: de overheid veronderstelt dat kunst een opvoedkundige waarde heeft voor het individu. Omdat de consument dit belang zelf niet voldoende inziet dient de overheid als opvoeder te hulp te schieten. Economen noemen dit het 'merit'-argument.

Daarnaast dient de kunst ook een algemeen belang: zij bevordert de kwaliteit van de samenleving, houdt culturele waarden en tradities in stand en levert internationaal prestige op. Dit laatste wordt ook wel het 'externe effecten'-argument genoemd. Beide belangen hebben een grote rol gespeeld bij de legitimering van het kunstbeleid door de jaren heen.

Oosterbaan Martinius geeft in zijn onlangs verschenen proefschrift⁷ een uitvoerig overzicht van de motieven en beweegredenen die door politici zijn aangevoerd ter verdediging van het kunstbeleid. Sinds 1945 is het beeldende kunstbeleid van de overheid lange tijd gericht geweest op cultuurspreiding. De overheid beschouwde kunst aanvankelijk als belangrijk voor de esthetische vorming en later als welzijnsbevorderend en stelde zichzelf de taak "het deelnemen van zo groot mogelijke groepen der bevolking aan het kunstleven te bevorderen"⁸. Deze 'merit'-argumentatie leidde tot een beleid dat vooral gericht was op sociale en geografische spreiding van kunst. Voor mensen uit alle delen van het land en onder alle sociale lagen van de bevolking moest de kunst toegankelijk worden gemaakt. Daarnaast werd er naar gestreefd relatief veel kunstenaars sociale zekerheid te bieden waartoe in 1956 de Beeldende Kunstenaars Regeling werd opgericht (zie verder par. I.1.2).

Aan het begin van de jaren tachtig maakte bovengeschetst beleid plaats voor een beleid dat zich meer richt op het bevorderen van de artistieke kwaliteit van kunstwerken⁹. De hantering van kwaliteitscriteria kan echter leiden tot een a-functionele 'kunstkunst'¹⁰ waardoor het op gespannen voet komt te staan met de belangen van het publiek, de afnamecomponent¹¹. Desondanks staat de artistieke kwaliteit vooralsnog centraal in de uitvoering van het kunstbeleid.

I.1.2 *Uitvoering van het beleid*

Het beleid op het gebied van de beeldende kunsten wordt uitgevoerd door de *hoofdafdeling beeldende kunsten, bouwkunst en vormgeving* van het ministerie van WVC. Uit de informatiebrochure van deze hoofdafdeling blijkt dat bovengenoemde belangen in de vier aandachtsgebieden van het beleid tot uitdrukking komen¹²:

- presentatie van en informatie over kunst;
- marktverruiming;
- spreiding van kunst;
- faciliteiten en subsidiemogelijkheden voor individuele beeldende kunstenaars.

De praktische uitvoering van het beleid bestaat uit regelingen, opdrachten, aankopen en subsidies. Bij een groot aantal van die beleidsinstrumenten fungeert de Raad voor de Kunst als adviesorgaan. Ook de provincies en de vier grote gemeenten kennen dergelijke raden die eveneens vooral adviserende taken hebben. Daarnaast bestaat op het terrein van de scheppende kunsten een aantal door de overheid gefinancierde fondsen¹³. Deze fondsenstructuur is de meest vergaande vorm waarin de overheid de uitvoering van beleid heeft uitbesteed.

Het fonds dat verantwoordelijk is voor een belangrijk deel van het beleid op het gebied van de beeldende kunsten is het recentelijk opgerichte *Stichting Fonds voor de beeldende kunsten, vormgeving en bouwkunst* (in het vervolg aangeduid met Fonds BKVB) dat ook direct het grootste budget te besteden kreeg. Dit Fonds BKVB voert namelijk sinds 1 januari 1988 de individuele subsidieregeling en de regeling Beroepskostenvergoedingen van het ministerie van WVC uit. Deze financiële ondersteuning van individuele kunstenaars was tot 1986 nog ondergebracht bij het ministerie van Sociale Zaken. In die toenmalige Beeldende Kunstenaars Regeling (de BKR, opgeheven in 1987) golden naast artistieke ook sociale criteria. Het beleid was er met name op gericht noodlijdende kunstenaars van een bescheiden inkomen te voorzien. Nu de regeling is overgeheveld naar het ministerie van WVC is van die laatste, sociale, criteria afscheid genomen. Overeenkomstig de hierboven geschetste veranderingen in de legitimering van het kunstbeleid, worden de individuele subsidies aan beeldende kunstenaars nu verstrekt op basis van een kwalitatieve beoordeling.

I.2 **Uitvoering van het beleid bij het Fonds BKVB**

Het Fonds BKVB is verantwoordelijk voor twee subsidievormen, namelijk de regeling *Individuele Subsidies* (IS) en de regeling *Beroepskostenvergoedingen* (BKV). Bij de aanvang van de regeling BKV werd in eerste instantie de mate van professionaliteit beoordeeld waarna, in verband met het beschikbare budget, een op kwaliteit gebaseerde prioriteitstelling volgde. Het gevolg van deze professionaliteitstoets was dat afgewezen kunstenaars niet alleen een subsidie werd onthouden, "maar dat zij naast deze financiële klap ook nog een morele trap na kregen"¹⁴. Het Fonds BKVB heeft hierop de procedure bijgesteld en nu is ook bij deze regeling het criterium van de artistieke kwaliteit doorslaggevend geworden. Alhoewel deze procedure daarmee haaks is komen te staan op de doelstelling van de regeling, namelijk kunstenaars tijdelijk in staat stellen hun beroepsuitoefening te continueren (overeenkomstig de voormalige BKR), verwoordde al snel na de instelling van de regeling ook een commissielid de mening van het Fonds hierover aldus: "De beroepskostenvergoedingen mogen geen sociale regeling worden, het gaat immers om de kunst die gehonoreerd wordt"¹⁵.

Bij de commissie IS is dit kwaliteitscriterium van aanvang af al de alles bepalende factor in de beoordeling. In verband met de opzet van het onderzoek gaan we in de volgende paragraaf

specifieker in op de gehanteerde procedure bij deze commissie.

I.2.1 Beoordelingsprocedure bij de commissie Individuele Subsidies

De regeling IS kent zes subsidievormen: startstipendia, werkbeurzen, reisbeurzen, studiebeurzen, projectsubsidies en presentatiesubsidies. De eerste twee bieden degenen die daarvoor in aanmerking komen de mogelijkheid zich gedurende een bepaalde periode te concentreren op de verdere ontwikkeling van hun scheppend werk. Het startstipendium is alleen bestemd voor kunstenaars die nog niet langer dan twee jaar geleden een opleiding aan een HBO-academie of een Technische Universiteit hebben afgerond; op een werkbeurs kunnen alle overige beeldende kunstenaars, vormgevers, architecten en fotografen een beroep doen. De overige vier subsidievormen hebben betrekking op de ontplooiing van bijzondere activiteiten die voor het kunstenaarschap van belang zijn en kunnen tegenwoordig door iedere kunstenaar die professioneel werkzaam is worden aangevraagd.

Om in aanmerking te komen voor een van deze subsidies dient een kunstenaar een uitgebreid aanvraagformulier in te vullen. In dit formulier vermeldt de kunstenaar, naast persoonsgegevens, onder meer informatie over zijn opleiding, marktgerichte activiteiten, scheppende werkzaamheden en een motivatie voor de aanvraag van de subsidie. Daarnaast dient de kunstenaar documentatie-materiaal in te zenden van recent werk, in de vorm van dia's, foto's, catalogi, videobanden e.d.. Een aanvraag wordt eerst formeel getoetst op onder meer volledigheid, alvorens in behandeling te worden genomen.

De behandeling van de aanvraag geschiedt in een van de zes subcommissies (werkgroepen), vier voor de beeldende kunst, één voor de vormgeving en één voor de bouwkunst, elk bestaande uit zes leden (bij vormgeving acht). De commissieleden ontvangen voorafgaand aan de vergadering een kopie van het aanvraagformulier die zij thuis moeten doornemen. Tijdens de vergadering wordt aan de hand van het beeldmateriaal de aanvraag behandeld. Nadat ieder lid een beargumenteerde, eerste indruk van het werk heeft gegeven, ontstaat er in de meeste gevallen een discussie. Komt de commissie er niet goed uit, bijvoorbeeld omdat men meent het verhaal van de kunstenaar nog eens te moeten horen, dan wordt besloten dat enkele leden een atelierbezoek afleggen. Deze aanvraag komt dan terug in een volgende vergadering. Is de commissie uiteindelijk tot een beslissing gekomen dan brengt zij advies uit aan het fondsbestuur dat op basis hiervan overgaat tot toekenning of afwijzing van de aanvraag.

I.3 Het esthetisch oordeel: schoonheid of kwaliteit?

In de voorgaande paragrafen is regelmatig gesproken over kwaliteit, het criterium dat een centrale plek heeft gekregen in het kunstbeleid van de laatste jaren. Waarom gaat het nu bij kwaliteit, en hoe komt een oordeel daarover tot stand?

In tegenstelling tot het begrip schoonheid is kwaliteit een criterium dat verwijst naar een functie, naar deugdelijkheid, juistheid of bekwaamheid. Bij een kwalitatief goede stoel doelen we vooral op comfort en stevigheid en zelden op de schoonheid van de vorm. Ten aanzien van kunstwerken wordt een dergelijk onderscheid ook toegepast. Naast kenmerken van schoonheid, een natuurgetrouwe afbeelding of een kleurrijk tafereel, die een 'streling van de zintuigen' veroorzaken, kan een kunstwerk ook eigenschappen bezitten die voor een 'juiste' beoordeling verlangen dat de subjectiviteit van dit zintuiglijk oordeel wordt overstegen. Hierbij gaat het dan met name om eigenschappen die van wezenlijk belang worden geacht voor de kunst als zodanig. De oordelen worden dan contemplatief van aard: heeft de kunst-

naar iets te vertellen; weet hij aan deze ideeën op een adequate wijze vorm te geven; draagt het werk bij aan de ontwikkeling van de kunstenaar of voegt het zelfs een nieuw element toe aan de kunsten als geheel? Een bevestigend antwoord op dergelijke vragen geeft een indicatie over de kwaliteit van een kunstwerk. Het gaat dan dus niet meer om decoratieve schoonheid van kleur of afbeelding; dergelijke oordelen betreffen slechts de persoonlijke smaak. Een kwalitatief hoogstaand kunstwerk mag vies en grauw van kleur zijn en kan een zowel zeer figuratieve als volkomen abstracte voorstelling bevatten; dit soort elementen is ondergeschikt. Waar het om gaat is dat alle elementen functioneel zijn voor datgene wat de kunstenaar wenst uit te drukken.

Voor een 'zuiver' esthetisch oordeel dient de persoonlijke smaak dan ook onderdrukt te worden; het is een kennisoordeel¹⁶ of een reflexief oordeel¹⁷ dat ontkoppeld moet worden van de vertroebelende prikkeling van de zintuigen. Volgens Kant is zo'n oordeel alleen mogelijk indien er geen object-gerichte belangen in het spel zijn. "Iedereen moet erkennen dat een oordeel over schoonheid, waarin zich ook maar het geringste belang mengt, zeer partijdig en geen zuiver smaakoordeel is. Men mag volstrekt niet met het bestaan van de zaak ingenomen zijn, maar men moet in dit opzicht volledig onverschillig staan, om in smaakaangelegenheden de rol van rechter te spelen"¹⁸. De beschouwer mag daarentegen wel op zichzelf gerichte belangen, als interesse, hebben of belang hebben bij het smaakoordeel als zodanig. Het esthetisch oordeel is een contemplatieve aangelegenheid: de beschouwer heeft geenszins een praktisch doel voor ogen en is niet uit op het verwerven van kennis; zijn oordeel dient in volledige vrijheid tot stand te komen.

Deze voorwaarden voor een zuiver esthetisch oordeel zien we veelvuldig terug in de hedendaagse beoordelingspraktijk. "Observers of art are often warned about separating their feelings (of liking) for a work from their judgment of its merit"¹⁹. De 'juiste' beoordeling kan alleen uitgevoerd worden door "De Ideale Beschouwer", en deze heeft "geen smaak"²⁰. Bij de selectie van commissieleden voor het Fonds BKVB wordt er dan ook op gelet dat deze mensen over de grenzen van hun eigen persoonlijke voorkeur heen kunnen kijken²¹. Adviseurs van de Amsterdamse Kunstraad zeggen hier desgevraagd ook naar te streven al erkennen zij inderdaad dat de persoonlijke smaak een aanzienlijke rol speelt bij de beslissingen²². Men beseft ook wel dat "... it's hard to attach much sense to such an extreme demand"²³. Een beperkte bijdrage van de intuïtie is daarom toegestaan²⁴. Toch dient het bij de oordeelsvorming te gaan "over heel feitelijke, benoembare zaken, waarover argumentatie en een gefundeerd eindoordeel mogelijk is"²⁵. Het instrument hiervoor zijn de criteria waarop het oordeel gebaseerd wordt.

Volgens Oosterbaan Martinius gaat het bij dergelijke criteria niet om toereikende, voor eens en altijd geldende gronden, maar om "abstracties, uit de kunst afgeleide begrippen, pogingen om de essentie van elkaars voorkeuren onder woorden te brengen"²⁶. "Ze zijn onder druk van de verantwoordingsdrang tot stand gekomen en zijn er vooral voor de buitenwereld"²⁷. Linnemann suggereert zelfs dat ze slechts zouden kunnen dienen "als legitimatie van de eigen smaak"²⁸. Hoe dit ook moge zijn, de commissies van het Fonds BKVB worden geacht te werken met criteria die, hoewel niet altijd expliciet genoemd, toch "het fundament vormen van het advies". Wanneer zij worden toegepast op één specifiek oeuvre "worden deze criteria heel concrete werktuigen in de handen van de commissie"²⁹.

In de volgende hoofdstukken zal onderzoek gerapporteerd worden waarin werd nagegaan in hoeverre beoordelaars in staat zijn op betrouwbare, consistente wijze kunstwerken te beoordelen op kwaliteit en een aantal andere criteria. Daarbij werd tevens de samenhang tussen de criteria bestudeerd. Het onderzoek betrof zowel experts (onderzoek 1) als in kunst

geïnteresseerde leken (onderzoek 2). Het onderzoek bij experts was zo opgezet dat enerzijds de bestaande beoordelingspraktijk bij het Fonds BKVB tot op zekere hoogte werd gesimuleerd, terwijl anderzijds mogelijke voordelen van een alternatieve beoordelingsprocedure konden worden onderzocht.

Het onderzoek bij de geïnteresseerde leken diende primair om verschillen en overeenkomsten tussen hun oordelen en die van de experts in kaart te brengen.

II Onderzoek I: oordelen van deskundigen

II.1 Doelstelling

Het onderzoek had tot doel inzicht te krijgen in een aantal aspecten samenhangende met de wijze waarop experts de kwaliteit van werken op het gebied van de beeldende kunst beoordelen. Deze aspecten betroffen met name:

1. De betrouwbaarheid van de desbetreffende oordelen in termen van
 - a. stabiliteit van de oordelen over de tijd, en
 - b. overeenstemming tussen de oordelen die door verschillende experts of door uit experts samengestelde commissies worden gegeven.
2. De mate waarin de kwaliteitsoordelen samenhangen met oordelen gegeven ten aanzien van andere criteria via welke de werken geëvalueerd worden.

Alhoewel dat niet tot de oorspronkelijke doelstelling van het onderzoek behoorde zijn op grond van de resultaten daarvan ook enige praktische aanbevelingen voor een specifieke beoordelingsprocedure geformuleerd.

II.2 Selectie van de experts

De selectie van de experts is uitgevoerd door medewerkers van het ministerie van WVC. In overeenstemming met de gebruikelijke samenstelling van commissies die belast zijn met kunstbeoordelingen zijn deskundigen afkomstig uit twee onderscheidbare groepen benaderd voor medewerking aan het onderzoek.

De eerste groep bestond uit 33 kunstbeschouwers. Deze personen waren alle professioneel actief op diverse terreinen binnen de beeldende kunstsector. Daarnaast is een tweede groep benaderd die bestond uit 30 professioneel actieve kunstenaars. In het verdere verloop van dit rapport worden de twee groepen aangeduid met 'beschouwers' en 'kunstenaars'.

Van de 63 benaderde deskundigen waren er uiteindelijk 34 bereid mee te werken aan het onderzoek: 18 beschouwers en 16 kunstenaars. Van de niet-medewerkers was het grootste deel onbereikbaar, deels, vermoedelijk, door een onjuiste adressering en deels door een verblijf in het buitenland. Een aantal aangeschreven experts weigerde medewerking omdat zij zich niet konden verenigen met de aard van het onderzoek; een klein aantal ontbrak het eenvoudigweg aan de noodzakelijke tijd.

II.2.1 *Achtergrond van de experts*

De gemiddelde leeftijd van de 34 deskundigen was 44.7 jaar (range van 32 tot 71 jaar). De gehele groep bestond uit 11 vrouwen en 23 mannen, afkomstig uit bijna alle regio's van het land.

De 16 participerende kunstenaars behoorden zonder uitzondering tot de toplaag van Nederlandse kunstenaars wat betreft verkoop van hun werk en de exposities waaraan zij deelnemen. Onder de 18 beschouwers bevonden zich drie galeriehouders, negen conservatoren van toonaangevende musea voor moderne kunst, één kunstcriticus en vijf medewerkers van diverse kunstinstellingen.

II.3 Selectie van het beoordelingsmateriaal

Met het oog op onbevooroordeelde beoordelingen werd het wenselijk geacht dat de kunstwerken onbekend zouden zijn aan de beoordelaars. Om die reden zijn 30 clusters dia's geselecteerd die beginnende kunstenaars hadden ingediend bij hun aanvraag voor een Startstipendium bij het Fonds BKVB. Dit type aanvraag staat alleen open voor die kunstenaars die niet langer dan twee jaar voor de indiening van hun aanvraag zijn afgestudeerd aan een Nederlandse academie voor Beeldende Kunst. De kunstenaars van wie de aanvragen en het werk is gebruikt voor het onderzoek worden in het vervolg aangeduid met 'aanvragers', dit om verwarring met de kunstenaars die beoordelen te voorkomen.

Met het gebruik van aanvragen voor een Startstipendium werd dus verwacht dat het werk (en de aanvrager) nauwelijks bekend zou zijn bij de deskundigen. Voorafgaand aan de beoordelingen werd nagegaan of deze vooronderstelling juist was. Op de vraag "bent u bekend met het werk van deze kunstenaar", werd in één geval door 7, in één geval door 6, en in twee gevallen door 5 beoordelaars bevestigend geantwoord. Het werk van de overige 26 aanvragers werd door twee of minder beoordelaars herkend. Gezien de grote kennis en ervaring van de experts is de onbekendheid met de geselecteerde werken bevredigend; geen enkel cluster werd door meer dan 20% van de beoordelaars herkend.

Ondanks de relatieve onbekendheid van de aanvragers en hun werk kon aan de andere kant wel aangenomen worden dat het werk aan minimaal te stellen eisen zou voldoen. Alle aanvragers hadden immers een kunstopleiding van enkele jaren doorlopen en vele onder hen hadden reeds enige marktgerichte activiteiten als kunstenaar ontplooid in de vorm van verkopen, exposities, verwerven van opdrachten of publicaties.

De selectie van de aanvragen werd uitgevoerd aan de hand van vier criteria:

1. Een gelijk aantal van door het Fonds BKVB afgewezen en gehonoreerde aanvragen.

Met dit criterium werd een flinke spreiding in kwaliteit van het werk beoogd. Bovendien is een gelijke verdeling wenselijk voor een statistische toetsing van de verschillen in beoordeling tussen de aanvragen.

2. Alleen aanvragen op het terrein van de beeldende kunst.

Aanvragen op het gebied van bouwkunst en vormgeving werden buiten beschouwing gelaten, alsmede aanvragen beeldende kunst met als specialisatie keramiek en audio-visuele en textiele vormgeving. Voor al deze typen aanvragen wordt een specifieke expertise van de beoordelaar noodzakelijk geacht.

3. Een grote spreiding in aard van het werk.

Met uitsluiting van de bij 2 genoemde specialisaties werd een spreiding in de aard van het werk nagestreefd die overeenkomt met het gebruikelijke aanbod bij het Fonds. Naast aanvragen tekenen/grafiek/schilderen werden daarom ook aanvragen beeldhouwkunst en ruimtelijk/monumentaal werk geselecteerd. Deze typen aanvragen worden bij het Fonds ook allemaal door de subcommissie beeldende kunst behandeld.

4. De visuele documentatie bij de aanvraag bestaat uit dia's.

In verband met de opzet van het onderzoek (zie § II.5) kon alleen werk op dia worden gepresenteerd. Daarom zijn in het onderzoek geen aanvragen betrokken waarbij andersoortige documentatie (b.v. videobanden, foto's of tekeningen) is geleverd.

II.3.1 *Achtergrond van de aanvragers*

De aanvragers waren 19 vrouwen en 11 mannen met een gemiddelde leeftijd van 28.7 jaar (range van 24 tot 46 jaar).

In tabel 1 staan de laatstgenoten, en afgeronde, opleidingen van de aanvragers weergegeven. Twee van de genoemde instituten zijn post-academische werkplaatsen: de Rijksacademie in Amsterdam en Ateliers '63 in Haarlem. De vijf aanvragers die bij deze opleidingen staan vermeld hebben tevens een opleiding aan een academie voor beeldende kunst gevolgd.

Aangezien de aanvragers nog maar kort geleden zijn afgestudeerd is hun huidige woonplaats in bijna alle gevallen gelijk aan de plaats waar zij de opleiding volgden.

Tabel 2 geeft een overzicht van de huidige specialisatie van de aanvragers en geeft daarmee tevens een indicatie over de aard van het bijgeleverde werk.

Gemiddeld werden bij de aanvragen 17.7 dia's (range van 12 to 21) ingediend.

De helft van de aanvragen is door het Fonds afgewezen. De 15 andere aanvragen zijn gehonoreerd met een Startstipendium, in 3 gevallen echter pas nadat een atelierbezoek had plaatsgevonden.

Tabel 1. Laatstgenoten opleiding van de 30 aanvragers.

Opleiding	aantal
Amsterdam, Gerrit Rietveld Academie	4
Amsterdam, Rijksacademie	1
Arnhem, Hogeschool voor de Kunsten	3
Den Bosch, Kon. Academie voor Kunst en Vormgeving	4
Den Haag, Kon. Academie van Beeldende Kunsten	1
Enschede, Academie voor Beeldende Kunsten AKI	3
Haarlem, Ateliers '63	4
Kampen, Chr. Hogeschool voor de Kunsten C. Huygens	1
Rotterdam, Academie van Beeldende Kunsten	3
Tilburg, Academie voor Beeldende Vorming	2
Utrecht, Hogeschool voor de Kunsten	4

Tabel 2. Huidige specialisatie van de 30 aanvragers.

Specialisatie	aantal
Tekenen/schilderen/grafiek	20
Ruimtelijke vormgeving	3
Beeldhouwen	5
Gemengde media	2

II.4 Methode

Het onderzoek richtte zich voor een deel op individuele, en voor een deel op commissie-beoordelingen. De procedure die werd gevolgd bij deze twee vormen van beoordelen wordt uitvoerig besproken in par. II.4.3 en II.4.4.

In het deel van het onderzoek waarin de individuele oordelen centraal stonden moesten de 30 clusters dia's worden beoordeeld op een aantal criteria. Deze beoordelingen werden uitgevoerd aan de hand van zogenaamde beoordelingsschalen. De selectie van de criteria en de achtergrond van de schaaltechniek worden besproken in de volgende twee paragrafen.

II.4.1 Criteria

De criteria waarop in het onderzoek de individuele beoordelingen werden uitgevoerd zijn ontleend aan de beoordelings-praktijk, aan psychologische onderzoeken en aan filosofische beschouwingen over het beoordelen van kunst.

De eerste zes criteria (van vakmanschap tot ontwikkeling) zijn ontleend aan het artikel van Cramer³⁰. Deze criteria "vormen het fundament van het advies" over een subsidie-aanvraag bij een beoordelingscommissie van het Fonds BKVB.

- **Vakmanschap**
- **Inhoud**
- **Zeggingskracht**
- **Oorspronkelijkheid**

Deze vier criteria zijn ook al eens in psychologisch onderzoek gehanteerd bij het beoordelen van beeldende kunst. Ze kwamen tevens als uiterst relevant voor het beoordelen van kunst naar voren in een onderzoek van Lindauer & Long³¹. Zij vroegen een groep van kunstenaar-docenten en galeriehouders het belang aan te geven van een groot aantal geselecteerde criteria. De drie hoogstgenoteerde criteria waren: kracht, vakmanschap en concept (idee of betekenis). Aan het criterium 'vernieuwing' werd een iets kleiner belang toegekend.

De laatste twee aan Cramer ontleende criteria,

- **Coherentie van het cluster** en
- **De ontwikkeling die de kunstenaar doormaakt**, zijn niet eerder opgenomen geweest in psychologisch onderzoek. Deze criteria zijn alleen hanteerbaar wanneer meerdere werken van één kunstenaar worden beoordeeld. Een dergelijke presentatie van kunstwerken is niet eerder (in onderzoek) toegepast en dat maakt het onderhavige onderzoek in dit opzicht zowel vernieuwend als meer relevant voor de huidige beoordelingspraktijk.

Het volgende criterium,

- **Zinvolheid**, is in eerste instantie ontleend aan een artikel van van der Tas³². Hij doelt met dit begrip op de 'zin' of betekenis van een kunstwerk; in hoeverre refereert het kunstwerk aan het 'zijn', de werkelijkheid die buiten het kunstwerk of de kunst in het algemeen is gelegen? De volgende twee criteria zijn vaak toegepast in experimenteel psychologisch onderzoek naar kunstbeoordelingen.

- **Complexiteit**
- **Dynamiek**

Het is algemeen aanvaard dat de intentie van de maker in het werk tot uitdrukking komt, waarbij dan soms het belang van

- **Integriteit** wordt benadrukt³³.

Van de bovengenoemde 10 criteria zal worden nagegaan in hoeverre zij bijdragen aan de totstandkoming van het kwaliteitsoordeel. Dat uiteindelijke criterium van:

- **Kwaliteit**, is ook in vele onderzoeken door middel van een schaalbeoordeling of een rapportcijfer gestructureerd beoordeeld.

Na selectie is de lijst met 11 criteria³⁴ doorgesproken met twee conservatoren van musea voor moderne kunst. Alhoewel zij bij een aantal criteria twijfels uitten over de relevantie voor de hedendaagse kunst en sommige criteria moeilijk te beoordelen vonden, werd geen enkel criterium als onzinnig beschouwd. Er was dus geen gegronde reden de lijst te wijzigen. Bovendien zou met het onderzoek de irrelevantie of onbruikbaarheid van een criterium vanzelf worden aangetoond.

De 11 criteria moesten wel op een gestructureerde wijze worden aangeboden om de wijze van beoordelen voor alle beoordelaars gelijk te maken. Daarom is besloten de criteria aan de hand van bipolaire schalen te laten beoordelen. Er is hiervoor gebruik gemaakt van een 9-punts schaal. Naast het grote aantal antwoordmogelijkheden, wat alleen zinvol is bij een grote expertise van de beoordelaars³⁵, biedt het oneven aantal antwoordmogelijkheden de beoordelaar de mogelijkheid te kiezen voor een neutrale middenpositie. De wijze waarop de criteria op de schalen zijn ondergebracht is weergegeven in tabel 3. In de meeste gevallen is het criterium zelf als adjectief gebruikt. Het criterium 'inhoud' is geïnterpreteerd in termen van 'rijk aan inhoud', 'zinvolheid' is vertaald in de mate waarop de thematiek van het werk betrekking heeft op de kunst zelf ('kunst-intrinsiek') of iets daarbuiten ('kunst-overschrijdend').

De volgorde van de schalen is voorgesteld door één van de geraadpleegde conservatoren. Voor hem was deze volgorde de meest logische en best hanteerbare. Om die reden hebben we deze volgorde tijdens alle fasen van het onderzoek constant gehouden.

De criteria werden verder niet omschreven; de interpretatie van de schaaladjectieven werd overgelaten aan de beoordelaar.

Tabel 3. De 11 criteria en de bijbehorende beoordelingsschalen.

Criterium		Schaal
Complexiteit:	Eenvoudig	Complex
Dynamiek:	Statisch	Dynamisch
Coherentie:	Onsamenhangend	Samenhangend
Vakmanschap:	Afwezigheid van vakmanschap	Aanwezigheid van vakmanschap
Inhoud:	Arm	Rijk
Zinvolheid:	Kunst-intrinsieke thematiek	Kunst-overschrijdende thematiek
Zeggingskracht:	Gebrek aan zeggingskracht	Zeggingskracht
Ontwikkeling:	Negatief	Positief
Oorspronkelijkheid:	Niet oorspronkelijk	Oorspronkelijk
Integriteit van kunstenaar:	Niet integer	Integer
Kwaliteit:	Slecht	Goed

II.4.2 Opzet van het onderzoek

Het onderzoek was verdeeld over drie zittingen met tussenpozen van één maand. Iedere zitting bestond uit zes gelijke sessies waarin 4 tot 6 beoordelaars hun beoordelingen uitvoerden. De beoordelaars werden op basis van beschikbaarheid bij één van de zes sessies ingedeeld waarbij zoveel mogelijk per sessie een gelijke verdeling van beschouwers en kunstenaars is aangehouden. Een sessie nam ongeveer drie uur in beslag³⁶. In tabel 4 is een schematische opzet van de opbouw van de drie onderzoekzittingen weergegeven. Helaas konden niet alle 34 beoordelaars op alle drie de zittingen aanwezig zijn, deels door ziekte, deels door andere verplichtingen. Het aantal beoordelaars per zitting staat in tabel 4 tussen haakjes vermeld.

Tijdens zitting I werden 20 clusters dia's individueel beoordeeld op de 11 schalen. In zitting II en III vonden er naast de individuele ook groepsbeoordelingen plaats.

In zitting II werden de 10 overige clusters dia's eerst individueel en vervolgens door de commissie als geheel beoordeeld. Tijdens de derde zitting werden de aanvragen uit zitting I nogmaals beoordeeld, 10 opnieuw individueel op alle criteria, de 10 andere door de commissies als geheel. Die laatste 10 kregen wel nogmaals een individueel kwaliteitsoordeel. In iedere sessie werden de clusters dia's in een nieuwe, aselechte volgorde gepresenteerd om mogelijke ongewenste volgorde-effecten te vermijden.

In tabel 4 staat naast het aantal behandelde aanvragen ook vermeld hoe die aantallen telkens waren opgebouwd uit bij het Fonds gehonoreerde (pos.) en afgewezen (neg.) aanvragen.

Tabel 4: Schematisch overzicht van de onderzoeksopzet (voor uitleg, zie tekst).

Zitting I (n = 33)	Zitting II (n = 30)	Zitting III (n = 25)
20 aanvragen: 13 pos. 7 neg.	10 aanvragen: 2 pos. 8 neg.	10 aanvragen uit zitting I: 7 pos. 3 neg.
INDIVIDUEEL	INDIVIDUEEL	INDIVIDUEEL: HERTEST
	Zelfde 10 aanvragen DISCUSSIE IN 6 COMMISSIES (2*6, 2*5 en 2*4 beoordelaars per commissie)	Andere 10 aanvragen uit zitting I: 6 pos. 4 neg. Individueel alleen kwaliteit: hertest DISCUSSIE IN 6 COMMISSIES (1*5 en 5*4 beoordelaars per commissie)

II.4.3 *Individuele oordelen*

De beoordelaars kregen voorafgaand aan de individuele beoordelingen een boekje bestaande uit een instructie en 10 (zitting II en III) of 20 (zitting I) beoordelingsformulieren met daarop de 11 schalen (voor iedere aanvraag één formulier). In de instructie werd vermeld dat het bij de beoordelingen ging om een persoonlijk oordeel en werd tevens het gebruik van de schalen uitgelegd. Over het werk werd verteld dat de dia's afkomstig waren van kunstenaars die een aanvraag hadden ingediend voor een Startstipendium en dat zij dus alle onlangs een kunstacademie hadden afgerond. Daarnaast werd nog vermeld dat de dia's in een chronologische volgorde vertoond zouden worden.

Nadat de beoordelaars de instructie hadden doorgenomen werd van ieder cluster één dia vertoond om een indruk te geven van het soort werk dat zij te zien zouden krijgen. Dit overzicht moest tevens een indicatie geven van de spreiding in kwaliteit waardoor het kon dienen als een referentiekader ten opzichte waarvan de eerste clusters beoordeeld konden worden. Na het overzicht volgde de vertoning van het cluster dia's. Bij iedere dia werd de titel, het formaat en het gebruikte materiaal van het origineel gegeven. Een dia bleef ongeveer 10 sec. in beeld. Na ieder cluster kregen de beoordelaars de tijd om de 13 schalen in te vullen. Zij konden voor de beoordelingen zelf het tempo bepalen. Als de laatste beoordelaar klaar was met het invullen werd begonnen met de vertoning van het volgende cluster totdat alle clusters beoordeeld waren.

Tijdens de drie zittingen zijn op deze wijze alle 30 aanvragen individueel beoordeeld, waarvan 20 zowel tijdens de eerste als derde zitting (10 op alle criteria en 10 alleen op kwaliteit).

II.4.4 *Commissie-oordelen*

Nadat de beoordelaars in zitting II en III hun individuele beoordelingen hadden uitgevoerd namen zij aan één tafel plaats en kregen zij een nieuwe instructie. De beoordelaars werd meegedeeld dat zij een fictieve beoordelingscommissie vormden die moest beslissen over 10 aanvragen voor een Startstipendium. De clusters dia's werden vervolgens opnieuw vertoond. Voorafgaand aan de discussie werd nu ook informatie verstrekt over de aanvragers. Deze informatie was afkomstig uit het aanvraagformulier dat zij hadden ingevuld bij hun aanvraag voor een Startstipendium en bestond onder meer uit persoonsgegevens (geslacht, leeftijd en woonplaats), de gevolgde opleiding, verkopen en exposities van het werk en een korte beschrijving van de intenties en inspiratiebron van de aanvrager. Tijdens de tweede zitting werd deze informatie voorgelezen gedurende de dia-vertoning. Bij de derde zitting werd de informatie direct na de vertoning gegeven omdat hier opnieuw de titel, het formaat en het gebruikte materiaal tijdens de vertoning moesten worden vermeld (het was immers twee maanden geleden dat de beoordelaars deze clusters gezien hadden). Bovendien moesten de beoordelaars, voordat de achtergrondinformatie werd verstrekt, eerst opnieuw een individueel kwaliteitsoordeel geven over het werk.

Na de vertoning van een cluster kregen de 'commissieleden' de gelegenheid ongeveer vijf minuten te discussiëren over het werk³⁷. De bedoeling van de discussie was te komen tot een, al dan niet eensgezind, eindoordeel over de aanvraag. Dit oordeel diende te bestaan uit een 'toekenning', 'twijfel/atelierbezoek' of 'afwijzing'³⁸. Tijdens de tweede en derde zitting zijn op deze wijze 20 aanvragen door 6 verschillende commissies beoordeeld.

II.4.5 *Verantwoording van de opzet*

Het onderzoek is uiteraard zo opgezet, dat de vragen die voortvloeien uit de globale doelstelling kunnen worden beantwoord. Aan de hand van die, in § II.1 genoemde, vragen zal nu de opzet worden verklaard.

1. Is een oordeel over de kunstwerken betrouwbaar?

a. Stabiliteit van het oordeel over de tijd.

De stabiliteit van een oordeel over de tijd, die een indicatie geeft van de mate waarin experts consistent oordelen, werd geoperationaliseerd als de **test-hertest betrouwbaarheid** ervan. Om deze te kunnen berekenen zijn 20 clusters dia's zowel tijdens de eerste als derde zitting individueel beoordeeld, 10 op alle criteria en 10 alleen op kwaliteit.

b. Overeenstemming tussen oordelen van verschillende beoordelaars.

Bij de mate waarin beoordelaars het met elkaar eens zijn is de **inter-beoordelaar betrouwbaarheid** in het geding. Voor de berekening daarvan zijn de individuele beoordelingen uit zitting I en II over alle 30 aanvragen beschikbaar.

2. Welke criteria dragen bij aan de totstandkoming van een kwaliteitsoordeel?

Ook voor de beantwoording van deze vraag zijn de beoordelingen over alle 30 aanvragen bruikbaar. Op basis van individuele en gemiddelde scores kan hiermee de relatie van ieder criterium met het kwaliteitsoordeel bepaald worden. Tevens kan worden nagegaan hoe de criteria onderling met elkaar samenhangen.

3. Hoe klinkt het individuele oordeel door in het eindoordeel van de commissie?

Over 20 aanvragen zijn in zitting II en III commissie-oordelen gevormd. Deze bieden de mogelijkheid na te gaan in hoeverre het individuele kwaliteitsoordeel in het groepsoordeel tot uitdrukking komt. Voor deze groepsoordelen werd tevens de inter-commissie betrouwbaarheid bepaald, de mate waarin de commissies het met elkaar eens zijn.

III Resultaten en conclusies van onderzoek I

In de volgende zes paragrafen worden de belangrijkste resultaten van het onderzoek besproken. De eerste twee paragrafen staan in het teken van de betrouwbaarheid: in hoeverre zijn experts consistent in hun beoordelingen (III.1) en in hoeverre zijn ze het met elkaar eens (III.2)? De volgende twee paragrafen hebben betrekking op de totstandkoming van het individuele kwaliteitsoordeel (III.3) en het eindoordeel van de commissie (III.4). Uit de resultaten wordt in paragraaf III.4 tevens een aantal conclusies getrokken ten aanzien van een mogelijke optimalisering van de onderzochte beoordelingsprocedure. In paragraaf III.5 worden de beoordelingen uit het onderzoek vergeleken met de door het Fonds BKVB genomen beslissingen. In de laatste paragraaf III.6 wordt tenslotte nagegaan in hoeverre er verschillen bestaan tussen de beoordelingen van beschouwers en kunstenaars.

III.1 Test-hertest betrouwbaarheid

Zoals reeds besproken bij de uiteenzetting van de methode (§ II.4.2) werden de twintig aanvragen uit de eerste zitting tijdens de derde zitting opnieuw individueel beoordeeld. Voor tien aanvragen werden opnieuw alle schalen ingevuld, over de tien andere aanvragen werd nu alleen een kwaliteitsoordeel gegeven. Via deze hertest kon worden nagegaan in hoeverre de deskundigen de beoordelingsschalen op betrouwbare wijze hanteren in die zin dat beoordeling van hetzelfde werk door een zelfde beoordelaar op verschillende tijdstippen tot dezelfde oordelen leidt. Uiteraard kan er sprake zijn van verschillen tussen beoordelaars (de ene beoordelaar is consistentere in zijn oordelen dan een andere) en tussen schalen (de ene schaal leidt tot consistentere oordelen dan een andere).

III.1.1 *Betrouwbaarheid van de beoordelaars*

Helaas hadden niet alle deskundigen uit de eerste zitting ook de mogelijkheid op de derde zitting te verschijnen. Hierdoor kon van slechts 24 beoordelaars de consistentie van de beoordelingen worden bepaald.

Allereerst is voor iedere beoordelaar de test-hertest correlatie^a per schaal over de 10, respectievelijk 20 aanvragen berekend. Van iedere beoordelaar werd vervolgens de mediane waarde van de 11 test-hertest correlaties vastgesteld. Deze mediaan is te beschouwen als een indicatie van de overall betrouwbaarheid van een beoordelaar. Bij één beoordelaar bleek de betrouwbaarheid zeer gering ($r = .06$), bij de overige varieerde deze van matig ($r = .37$) tot goed ($r = .87$) met een gemiddelde van $.64$. Over het algemeen blijkt er dus van een acceptabele betrouwbaarheid sprake te zijn.

III.1.2 *Betrouwbaarheid van de schalen*

Per schaal werd nagegaan in hoeverre deze door de individuele beoordelaars op betrouwbare wijze werd gehanteerd. Daartoe werd van de 24 test-hertest correlaties per schaal de mediane waarde vastgesteld die te beschouwen is als een indicatie voor die betrouwbaarheid. De mediane waarden varieerden van $.47$ tot $.68$, wat wijst op een redelijke betrouwbaarheid.

^a Tenzij anders vermeld, zijn alle genoemde correlaties Pearson product-moment correlaties.

De schalen thematiek ($r = .50$) en integriteit ($r = .47$) bleken het minst betrouwbaar, hetgeen in overeenstemming is met het feit dat meerdere beoordelaars aangaven er veel moeite mee te hebben om het werk op deze schalen te beoordelen.

Conclusie 1

Over het algemeen is er sprake van een behoorlijke consistentie, zowel ten aanzien van de beoordelaars als ten aanzien van de gehanteerde schalen. Dit resultaat bevestigt de bruikbaarheid van de toegepaste schaalmethode en schept vertrouwen in de wijze waarop de deskundigen zich van hun beoordelingstaak hebben gekweten.

III.2 Overeenstemming tussen de beoordelaars

Nu we hebben gezien dat de beoordelingen redelijk betrouwbaar zijn uitgevoerd, wordt het interessant na te gaan in hoeverre de beoordelaars het onderling eens zijn. Hiertoe zijn voor iedere schaal afzonderlijk de correlaties berekend tussen de scores van alle mogelijke paren beoordelaars, die vervolgens werden gemiddeld. Deze gemiddelde correlaties geven per schaal aan in hoeverre de beoordelaars het onderling eens zijn; ze zijn weergegeven in tabel 5. De overeenstemming is zeer gering op de schalen thematiek en integriteit. Dit resultaat is in overeenstemming met de relatief geringe test-hertest betrouwbaarheid van deze schalen (zie § III.1.2).

Tabel 5. Gemiddelde correlaties tussen de scores van de 34 beoordelaars op de 11 schalen.

n=34 beoordelaars	gemiddelde correlatie
complexiteit	.15
dynamiek	.15
samenhang	.34
vakmanschap	.21
inhoud	.24
thematiek	.09
zeggingskracht	.24
ontwikkeling	.21
oorspronkelijkheid	.21
integriteit	.15
kwaliteit	.27

Met betrekking tot het kwaliteitsoordeel stemt de gevonden waarde van .27 redelijk overeen met resultaten uit eerder onderzoek naar kwaliteitsoordelen van kunstexperts, waarbij

gemiddelde correlaties van .29 en .24 werden gevonden³⁹. Ook de overeenstemming op de criteria vakmanschap en oorspronkelijkheid is nagenoeg gelijk aan die uit het onderzoek van Getzels en Csikszentmihalyi, waarin voor deze criteria gemiddelde correlaties van respectievelijk .22 en .31 werden gevonden.

Hoewel genoemde waarden op een geringe overeenstemming tussen beoordelaars wijzen, is deze toch niet lager dan werd vastgesteld bij onafhankelijk van elkaar werkende beoordelaars die wetenschappelijke onderzoeksvoorstellen of artikelen dienden te evalueren⁴⁰. Er lijkt dan ook geen reden de betrouwbaarheid van beoordelingen in de kunst lager aan te slaan dan die van beoordelingen in de wetenschap.

Conclusie 2

De gemiddelde overeenstemming tussen de oordelen over eenzelfde kunstwerk door twee beoordelaars is gering. Desondanks is de overeenstemming met betrekking tot het kwaliteitsoordeel niet slechter dan bij beoordelingen van wetenschappelijke manuscripten het geval bleek.

III.3. Relatie tussen de criteria en het kwaliteitsoordeel

Een van de centrale vragen uit dit onderzoek betreft de samenhang tussen het kwaliteitsoordeel en de overige criteria. Op de criteria dynamiek en samenhang na bleken deze alle significant gerelateerd te zijn aan dat kwaliteitsoordeel. Met behulp van een stapsgewijze regressie-analyse werd nagegaan op grond van welke combinatie van criteria het kwaliteitsoordeel het best kan worden voorspeld.

Uit deze analyse bleek dat de vier criteria **inhoud, vakmanschap, oorspronkelijkheid en ontwikkeling** onafhankelijk van elkaar een bijdrage leveren aan het kwaliteitsoordeel en gezamenlijk 97% van de variantie daarvan verklaren. Dit betekent dat het kwaliteitsoordeel het best geconstrueerd kan worden via de criteria inhoud, vakmanschap, oorspronkelijkheid en ontwikkeling. Zijn de oordelen op deze vier criteria bekend, dan is het mogelijk het kwaliteitsoordeel met zeer grote nauwkeurigheid te voorspellen.

Gegeven dit resultaat is het de moeite waard na te gaan of de somscore van de vier bijdragende criteria het kwaliteitsoordeel zou kunnen vervangen. Dit zou het voordeel hebben dat bekendmaking van de scores op de vier bijdragende criteria de beoordelingsprocedure voor alle betrokkenen inzichtelijker zou maken. De op basis van die somscores berekende gemiddelde overeenstemming tussen beoordelaars blijkt nagenoeg gelijk aan de overeenstemming tussen hun kwaliteitsoordelen, $r_{\text{gem}} = .26$.

Tevens bleek dat deze somscore niet slechts gemiddeld over de observatoren maar ook voor elk van hen individueel een vrij goede voorspelling van hun kwaliteitsoordeel opleverde (het desbetreffende percentage verklaarde variantie varieerde van 50 tot 95%, met een gemiddelde van 85%).

Conclusie 3

Het over observatoren gemiddelde kwaliteitsoordeel over een cluster dia's kan nauwkeurig worden gereconstrueerd op grond van de gemiddelde oordelen over inhoud, vakmanschap, oorspronkelijkheid en ontwikkeling van het werk. Deze vier criteria voorspellen tevens in behoorlijke mate de variantie in het kwaliteitsoordeel voor iedere observator afzonderlijk. Op grond hiervan is het verdedigbaar het kwaliteitsoordeel te vervangen door de som van de scores op deze vier criteria.

III.4 Het eindoordeel van de commissies m.b.t. honorering van de aanvragen

Tot nu toe hebben wij ons bij de bespreking van de resultaten beperkt tot de individuele beoordelingen van de 30 geselecteerde aanvragen. Tijdens de tweede en derde zitting is er echter over in totaal 20 van deze 30 aanvragen eveneens een groepsoordeel gevormd betreffende het al dan niet honoreren van de aanvragen (zie voor de gevolgde procedure bij de discussie § II.4.4). Zoals vermeld was de samenstelling van de commissies tijdens de tweede zitting anders dan tijdens de derde zitting. Wel is in elk van de zittingen door zes commissies over tien aanvragen een eindoordeel gegeven.

III.4.1 *Samenhang tussen het individuele kwaliteitsoordeel en het gezamenlijke eindoordeel*

Als de eindoordelen van de commissies beschouwd worden als gerangordend (een 3 = toekenning is beter dan een 2 = twijfel en dat is beter dan een 1 = afwijzing van de aanvraag), is het mogelijk de overeenstemming tussen het individuele kwaliteitsoordeel en het eindoordeel van de betreffende commissie te berekenen met de Spearman rangorde correlatie coëfficiënt (r_s). Opvallend was dat bij 6 van de 12 commissies slechts het kwaliteitsoordeel van één beoordelaar significant correleerde met het eindoordeel van de commissie. Dit resultaat wekt de suggestie dat het eindoordeel van een commissie vaak in vrij sterke mate de reflectie is van het oordeel van slechts één, dominant dan wel overtuigend, commissielid. In het geval van een gelijkwaardige bijdrage van ieder lid aan het eindoordeel zou hun kwaliteitsoordeel in min of meer overeenkomstige mate gerelateerd moeten zijn aan dat eindoordeel. Op basis van de correlaties blijken slechts bij vier van de twaalf commissies de leden evenredig bij te dragen aan het eindoordeel.

Conclusie 4

Waar binnen sommige commissies het eindoordeel betreffende het al dan niet honoreren van een aanvraag een redelijke weerspiegeling is van de individuele kwaliteitsoordelen, reflecteert binnen andere commissies het eindoordeel voornamelijk het oordeel van één van haar leden, waaraan de overige leden zich tijdens de discussie hebben aangepast.

III.4.2 *Overeenstemming tussen de eindoordelen*

Voor zowel zitting II als zitting III werden correlaties berekend tussen de eindoordelen van de zes commissies. In zitting II heeft de gemiddelde correlatie over alle mogelijke paren van deze zes commissies een waarde van .50. Deze overeenstemming is deels te verklaren uit het feit dat 5 van de 10 tijdens deze zitting behandelde aanvragen unaniem van zeer lage kwaliteit werden bevonden⁴¹. In zitting III, waar die 10 aanvragen aan de commissies werden voorgelegd die in zitting I bij de individuele beoordelingen de grootste variantie vertoonden in de kwaliteits-beoordelingen, bleek de overeenstemming aanmerkelijk geringer ($r_{s,gem} = .22$).

De matige overeenstemming tussen de eindoordelen van de commissies kan meerdere oorzaken hebben. Het is mogelijk dat het een direkt gevolg is van de geringe overeenstemming tussen de afzonderlijke beoordelaars (zie § III.2). Het gebrek aan duidelijke overeenstemming kan echter ook veroorzaakt zijn door sociale processen die een rol spelen bij de consensus-vorming. Deze laatste mogelijkheid wordt ondersteund door de in de vorige

paragraaf vermelde bevinding dat het eindoordeel van een commissie vaak het oordeel van slechts één commissielid reflecteert. Voor zover het gebrek aan overeenstemming tussen de eindoordelen inderdaad hieruit voortkomt, zal een vervanging van het eindoordeel door een somscore van de individuele kwaliteitsoordelen een grotere overeenstemming tussen de commissies moeten opleveren. Deze mogelijkheid zal in de volgende paragraaf worden onderzocht.

III.4.3 Vervanging van het eindoordeel

Voor elk van de twaalf commissies is het gemiddelde berekend van de kwaliteitsoordelen van de individuele beoordelaars voor iedere aanvraag. De gemiddelde correlatie tussen deze gemiddelde kwaliteitscores is .61 voor zitting II en .34 voor zitting III. Om deze gemiddelde kwaliteitscores te kunnen vergelijken met de eindoordelen zijn de eerstgenoemde gecodeerd tot een één (afwijzing), twee (twijfel) of drie (toekenning)⁴², overeenkomstig de rangorde bij de eindoordelen. De gemiddelde correlaties blijken voor beide zittingen slechts lichtelijk te dalen als gevolg van deze reductie, $r_{s, \text{gem}} = .56$ voor zitting II en $r_{s, \text{gem}} = .30$ voor zitting III. De overeenstemming op basis van de, al dan niet gecodeerde, kwaliteitscores is dus hoger dan die op basis van de eindoordelen, waar respectievelijk waarden van .50 en .22 werden gevonden (zie § III.4.2).

Aangezien de eindoordelen tot stand zijn gekomen nadat de commissieleden de informatie uit het aanvraagformulier hadden vernomen en over het getoonde werk hadden gediscussieerd, is het zinvol na te gaan of de gemiddelde kwaliteitscores wel hetzelfde meten als de eindoordelen van de commissies. In dit verband kan het volgende worden opgemerkt.

De gecodeerde gemiddelde kwaliteitscore van een commissie uit zitting II correleert hoog met het eindoordeel van diezelfde commissie, de desbetreffende correlatiecoëfficiënten variëren van $r_s = .65$ tot $r_s = 1.00$, met een gemiddelde r_s van .96. Voor de commissies uit zitting III zijn deze correlaties echter beduidend lager en variëren van $r_s = .16$ tot $r_s = .69$, met een gemiddelde r_s van .39. Bij deze laatste commissies meet de gemiddelde kwaliteitscore dus (deels) iets anders dan het eindoordeel. Dit valt toe te schrijven aan het feit dat het eindoordeel bij deze commissies vaak het kwaliteitsoordeel van één enkel lid bleek te weerspiegelen. In plaats van de gemiddelde kwaliteitscore als onbruikbaar te beschouwen is het o.i. veeleer aan te raden deze toch als alternatief voor het eindoordeel te gebruiken. Bij gebrek aan een extern criterium om de validiteit^a van beide scores te bepalen, kan de adequaatheid van de scores alleen uitgedrukt worden in termen van betrouwbaarheid. Op basis van die betrouwbaarheid zijn de gemiddelde kwaliteitscores te prefereren boven de eindscores.

Conclusie 5

Geconcludeerd kan worden dat de gemiddelde kwaliteitscores van de leden van een commissie een grotere overeenstemming opleveren tussen de commissies dan de via discussie tot stand gekomen eindoordelen. De beoordelingsprocedure kan dus betrouwbaarder gemaakt worden door de eindoordelen te vervangen door die gemiddelde kwaliteitscores.

^a Het gaat bij de validiteit om de vraag: in hoeverre meet je met de scores het verschil in kwaliteit dat je wenst te meten. Om dat te kunnen nagaan dien je te beschikken over een objectief en extern criterium. Het zal eenieder duidelijk zijn dat een dergelijk criterium voor kwaliteit niet bestaat. In de discussie (§ VI.2) zal hierop nader worden ingegaan.

III.4.4 *Bepaling van het benodigde aantal beoordelaars*

Op grond van de tot dusverre gepresenteerde resultaten blijkt de beoordelingsprocedure aan betrouwbaarheid en objectiviteit⁴³ te winnen wanneer het eindoordeel van een commissie als geheel vervangen wordt door het gemiddelde kwaliteitsoordeel van een aantal individueel oordelende deskundigen.

De betrouwbaarheid van dit gemiddelde kwaliteitsoordeel zal toenemen met het aantal beoordelaars⁴⁴. Indien overeenkomstig de praktijk binnen het Fonds BKVB wordt uitgegaan van zes beoordelaars kan, gegeven onze resultaten, een betrouwbaarheid van .69 verwacht worden. Toename van het aantal beoordelaars zal nog slechts een geringe verbetering van deze betrouwbaarheid tot gevolg hebben.

Conclusie 6

Een beslissing over de kwaliteit van een aanvraag kan met een bevredigende betrouwbaarheid ($r = .69$) gebaseerd worden op het gemiddelde kwaliteitsoordeel van zes onafhankelijke beoordelaars.

III.5 **Vergelijking met commissies van het Fonds BKVB**

De mate van overeenstemming tussen de beoordelingen van de commissies uit ons onderzoek en die van het Fonds kan als volgt inzichtelijk worden gemaakt. Van de 20 beoordeelde aanvragen bleek de gemiddelde kans om door een van de commissies niet in eerste instantie te worden afgewezen 31%. Voor de 8 aanvragen die door het Fonds niet in eerste instantie waren afgewezen bleek die kans 54% te zijn. Voor de 12 aanvragen die door het Fonds wel in eerste instantie waren afgewezen was die kans veel kleiner, te weten 11%. Met andere woorden: Een door het Fonds afgewezen aanvraag heeft een vijf maal zo kleine kans om door een commissie uit ons onderzoek niet te worden afgewezen als een door het Fonds niet afgewezen aanvraag.

Conclusie 7

Gezien de mate van overeenstemming tussen de beoordelingen uit het onderzoek en de beslissingen van het Fonds mag geconcludeerd worden dat de experts uit het onderzoek qua deskundigheid en beoordelingscapaciteit overeenkomen met de commissieleden van het Fonds. Dit rechtvaardigt een generalisering van de resultaten van het onderzoek naar de beoordelingspraktijk.

III.6 **Verschillen tussen oordelen van kunstbeschouwers en kunstenaars**

Tot nu toe hebben we de beoordelaars uit het onderzoek als één homogene groep deskundigen behandeld. In de beoordelingspraktijk wordt echter onderscheid gemaakt tussen twee typen experts: kunstbeschouwers en kunstenaars. Commissies worden in de regel samengesteld uit vertegenwoordigers van beide groepen. Gezien hun achtergrond en de positie die zij innemen in de kunstwereld, lijkt een tweedeling tussen beschouwer-expert en kunstenaar-expert een verantwoord onderscheid. Het is echter maar de vraag of deze verschillende achtergronden ook werkelijk tot verschillen in beoordeling leiden.

Een directe toetsing tussen de gemiddelde scores van beschouwers en kunstenaars leverde voor de criteria complexiteit (3.86 vs. 4.44), samenhang (6.33 vs. 5.95), thematiek (3.78 vs. 3.46), vakmanschap (5.70 vs. 5.37), ontwikkeling (4.97 vs. 4.56) en oorspronkelijkheid (3.99 vs. 3.62) een significant ($p < .05$) resultaat op. Behalve bij complexiteit ligt de gemiddelde score van de kunstenaars ten aanzien van deze criteria dus lager dan die van de beschouwers. Dit werkt echter niet door in de gegeven kwaliteitsoordelen, deze liggen bij de kunstenaars gemiddeld zelfs iets hoger dan bij de beschouwers (4.23 vs. 4.08), een overigens niet significant verschil. Ten aanzien van de correlaties tussen de gemiddelde scores van beide groepen over de 30 aanvragen kan opgemerkt worden dat deze bevredigend is. Met betrekking tot kwaliteit is de desbetreffende correlatiecoëfficiënt $r = .79$, voor de criteria inhoud, vakmanschap, oorspronkelijkheid en ontwikkeling, die tezamen het kwaliteitsoordeel in hoge mate voorspellen, zijn de correlaties respectievelijk .70, .80, .78 en .61.

Conclusie 8

Er bestaat een grote overeenstemming tussen de beoordelingen van kunstbeschouwers en kunstenaars. Gegeven echter dat de kunstenaars op enkele relevante criteria gemiddeld iets negatiever oordelen dan de beschouwers, verdient het aanbeveling het percentage kunstenaars (c.q. beschouwers) binnen commissies constant te houden indien de eindoordeelen mede op deze criteria worden gebaseerd. Indien slechts via gegeven kwaliteitsoordelen het eindoordeel bepaald wordt lijkt dat niet noodzakelijk.

Alvorens op een meer algemene discussie naar aanleiding van de resultaten van het eerste onderzoek in te gaan (zie hoofdstuk VI), wordt nu eerst het tweede onderzoek, waarbij geïnteresseerde leken als beoordelaars fungeerden, aan de orde gesteld.

IV Onderzoek II: oordelen van geïnteresseerde leken

IV.1 Doelstelling

Dit onderzoek had primair tot doel na te gaan in hoeverre oordelen over beeldende kunst door geïnteresseerde leken, die zich als amateur op actieve of passieve wijze met kunst bezighouden, overeenstemmen met die van de experts uit onderzoek I. Die overeenstemming had betrekking op

1. het kwaliteitsoordeel en de oordelen op de andere criteria waarop het werk werd geëvalueerd;
2. de criteria waarmee dat kwaliteitsoordeel het best voorspeld kan worden.

Daarnaast werd ook voor deze groep observatoren de betrouwbaarheid van de desbetreffende oordelen vastgesteld in termen van hun onderlinge overeenstemming.

IV.2 Selectie van de beoordelaars

Gezien de doelstelling van het onderzoek werd gezocht naar personen die zich actief dan wel passief als amateur met kunst bezighielden en daarmee van hun interesse voor kunst blijk gaven. Via medewerkers van de afdeling amateuristische kunstbeoefening van het ministerie van WVC zijn de Stichting Kunstzinnige Vorming Rotterdam, de Federatie van Amateur Beeldende Kunstenaarsverenigingen en de Stichting Beeldende Amateurkunst benaderd met het verzoek een lijst op te stellen van amateur beeldende kunstenaars. Een soortgelijk verzoek ging uit naar de Stichting Beeldende Kunst in verband met een lijst van regelmatige deelnemers aan de kunstuitleen.

Op basis van de desbetreffende lijsten zijn 20 actieve kunstbeoefenaars en 27 passieve kunstliefhebbers aangeschreven met het verzoek deel te nemen aan het onderzoek. Van deze 47 personen hebben er uiteindelijk 26 aan het onderzoek geparticipeerd: 12 beoefenaars en 14 liefhebbers. Van de overigen hadden er twee vergeten hun adres te vermelden, konden er twee op basis van de opgegeven voorkeuren niet worden ingeroosterd, was er één "helaas niet beschikbaar" en hadden er zestien niet gereageerd op het verzoek.

De gemiddelde leeftijd van de 26 geïnteresseerde leken was 43 jaar (range van 23 t/m 68 jaar). De groep bestond uit 14 vrouwen en 12 mannen.

Hoewel van deze beoordelaars, gezien de criteria van selectie, verwacht mocht worden dat zij enige ervaring met en belangstelling voor beeldende kunst zouden hebben, bestond over de mate van expertise toch onduidelijkheid. Om die reden werd achteraf aan alle beoordelaars een vragenlijst gestuurd waarmee specifiek inzicht kon worden verkregen in de achtergrond en interesse van de participanten op het terrein der beeldende kunst. Van 25 van de 26 deelnemers is uiteindelijk een ingevulde vragenlijst retour ontvangen.

Hieruit kon worden afgeleid dat de meeste participerende beoordelaars een redelijke ervaring hebben met het scheppen van kunst; meer dan de helft volgde een cursus en beoefende zelf één of meerdere activiteiten op het terrein van de beeldende kunst. Een meerderheid van de beoordelaars getuigde ook van de nodige belangstelling voor beeldende kunst via het geregeld bezoeken van musea of galleries en het lezen van recensies of tijdschriften op dit gebied.

De totale groep is echter weinig homogeen wat betreft de specifieke aard van de achtergrond

en de mate van belangstelling; tussen de beoordelaars zijn hierin flinke verschillen zichtbaar. Mede in verband met het feit dat een groot deel van de beoordelaars zich enigszins expert acht, lijkt de benaming 'geïnteresseerde leken' toch het meest toepasselijk.

IV.3 Selectie van het beoordelingsmateriaal

Het beoordelingsmateriaal was identiek aan dat gebruikt tijdens onderzoek I, echter met dien verstande dat thans slechts 26 van de oorspronkelijk 30 clusters dia's beschikbaar waren.

IV.4 Methode

De beoordelaars gaven hun beoordelingen tijdens één sessie van ruim drie uur. In totaal waren er zes van zulke sessies georganiseerd, waarbij minimaal twee, en maximaal zes beoordelaars aanwezig waren. Voor hun deelname ontvingen de beoordelaars een vergoeding van f 25,=/uur.

De instructie aan de beoordelaars en de presentatie van de dia's waren conform die bij het eerste onderzoek (zie § II.4.3).

Van de daar gehanteerde beoordelingsschalen werden er in dit onderzoek 9 gehandhaafd. De criteria, thematiek en integriteit, die bij de experts niet bleken te voldoen werden thans achterwege gelaten (zie §§ III.1.2 en III.2).

Uiteindelijk waren zo van 26 leken de beoordelingen op 9 schalen van 26 clusters dia's beschikbaar voor analyse. De resultaten daarvan worden in het volgende hoofdstuk besproken.

V Resultaten en conclusies van onderzoek II

In dit hoofdstuk worden de belangrijkste resultaten van onderzoek II besproken. Paragraaf V.1 is gewijd aan de inter-beoordelaar betrouwbaarheid, dat wil zeggen aan de vraag in hoeverre de beoordelaars onderling overeenstemmen in hun beoordeling van de aanvragen. In paragraaf V.2 wordt vervolgens nagegaan hoe de oordelen op de verschillende schalen onderling en met het kwaliteitsoordeel samenhangen. Waar mogelijk zullen de resultaten uit deze twee paragrafen worden vergeleken met die ten aanzien van de experts uit het eerste onderzoek. De overeenstemming tussen de beoordelingen van de leken en de experts komt in paragraaf V.3 aan de orde. Daarin wordt tevens nagegaan in hoeverre uit de kwaliteitsbeoordelingen van de leken afgeleide beslissingen gerelateerd zijn aan de beslissingen van het Fonds BKVB.

V.1 Overeenstemming tussen de beoordelaars

Voor iedere schaal afzonderlijk zijn correlaties berekend tussen de scores van telkens twee beoordelaars over alle 26 aanvragen. In het totaal werden zo per schaal 325 correlatie coëfficiënten berekend die vervolgens gemiddeld werden. Deze gemiddelde correlaties zijn opgenomen in tabel 6 en geven voor iedere schaal de mate van overeenstemming tussen de beoordelaars aan. Net als bij de experts is deze inter-beoordelaar betrouwbaarheid over het algemeen gering. Ten opzichte van die experts is er echter een aantal verschillen waarneembaar voor wat betreft de mate van overeenstemming op de criteria.

Tabel 6. De gemiddelde correlatie per schaal tussen de scores van de 26 beoordelaars voor de 26 aanvragen.

criterium	Gemiddelde correlatie
Complexiteit	.31
Dynamiek	.26
Samenhang	.21
Vakmanschap	.25
Inhoud	.15
Zeggingskracht	.21
Ontwikkeling	.15
Oorspronkelijkheid	.10
Kwaliteit	.21

De overeenstemming ten aanzien van de criteria complexiteit en dynamiek is bij de leken groter dan bij de experts, de desbetreffende gemiddelde correlatie coëfficiënten zijn respectievelijk .31 versus .18 voor complexiteit en .26 versus .15 voor dynamiek. De overeenstemming

ten aanzien van het criterium oorspronkelijkheid is daarentegen bij de leken kleiner, de gemiddelde correlatie coëfficiënten zijn respectievelijk .10 en .27. Ook over de kwaliteit van de aanvragen stemden de leken onderling ($r_{\text{gem}} = .21$) iets minder overeen dan de experts onderling ($r_{\text{gem}} = .27$) deden.

De gevonden waarden voor de overeenstemming zijn voor een aantal criteria lager dan vergelijkbare correlaties uit eerder onderzoek. Getzels en Csikszentmihalyi vonden voor beoordelingen van leken op de criteria kwaliteit, vakmanschap en oorspronkelijkheid gemiddelde correlaties van respectievelijk .35, .40 en .48⁴⁵. In een onderzoek van Hendriks was de overeenstemming tussen leken .30 voor wat betreft hun kwaliteitsbeoordelingen⁴⁶. Een mogelijke verklaring van de grotere inter-beoordelaar overeenstemming in die onderzoeken is dat de samenstelling van de betrokken groep beoordelaars veel homogener was dan in dit onderzoek. In beide studies werd de groep leken gevormd door studenten die qua opleiding en leeftijd nauwelijks uiteenliepen. Bovendien hebben we in § IV.2 al opgemerkt dat de leken in dit onderzoek onderling vrij sterk verschillen voor wat betreft hun ervaring en interesse op het gebied van de kunst. De genoemde studies presenteerden geen gegevens over die achtergrond van de participerende studenten, maar aangenomen mag worden dat, gezien de eenzijdige selectie, die beoordelaars ook wat betreft hun (gebrek aan) kunst-specifieke expertise een grotere homogeniteit vertoonden.

Conclusie 1

De gemiddelde overeenstemming tussen de beoordelingen van 26 aanvragen door twee beoordelaars is op alle criteria gering. Met betrekking tot het kwaliteitsoordeel is de gevonden waarde van .21 iets kleiner dan de gemiddelde overeenstemming van de experts uit het eerste onderzoek en eveneens kleiner dan de waarden die voor groepen leken in ander onderzoek werden gevonden.

V.2 Relaties tussen het kwaliteitsoordeel en de overige criteria

Per aanvraag werd voor iedere schaal het gemiddelde van de scores van de 26 beoordelaars berekend. Vervolgens werden over alle 26 aanvragen deze gemiddelde schaalscores met elkaar gecorreleerd.

Met uitzondering van complexiteit en dynamiek, bleken alle criteria significant gerelateerd aan het kwaliteitsoordeel. De correlaties zijn in grote mate overeenkomstig aan de waarden die gevonden werden voor de experts. Het verband tussen oorspronkelijkheid en kwaliteit was echter significant zwakker dan bij de experts ($r = .40$ vs. $r = .88$; $p < .05$) terwijl de relatie tussen vakmanschap en kwaliteit significant sterker bleek ($r = .97$ vs. $r = .84$; $p < .05$).

Dit resultaat is in overeenstemming met de eerder genoemde resultaten van Getzels en Csikszentmihalyi waaruit bleek dat bij experts originaliteit een belangrijker indicator is van kwaliteit dan vakmanschap, terwijl bij leken het omgekeerde het geval is.

Om te kunnen bepalen welke criteria van invloed zijn op de totstandkoming van het kwaliteitsoordeel werd dezelfde procedure (een stapsgewijze regressie-analyse, zie § III.3) uitgevoerd als bij de experts.

Hieruit bleek dat de criteria **vakmanschap** en **ontwikkeling** gezamenlijk 97% van de variantie in het kwaliteitsoordeel verklaarden, evenveel als een combinatie van vier criteria dat deed bij de experts. Tot laatstgenoemde vier criteria behoorden naast vakmanschap en ontwikkeling

ook de criteria inhoud en oorspronkelijkheid. Deze laatste twee leverden bij de geïnteresseerde leken echter geen onafhankelijke bijdrage aan de totstandkoming van het kwaliteitsoordeel. Voor wat betreft de individuele beoordelaars gold dat de som van de scores op beide criteria bij twee van hen slechts 6 respectievelijk 16% van de variantie in het kwaliteitsoordeel verklaarden. Bij de overige beoordelaars echter varieerde dit percentage van 49 tot 91% met een gemiddelde van 74%.

Conclusie 2

*Waar bij de experts het gemiddelde kwaliteitsoordeel het best voorspeld kon worden door de som van de scores op de criteria inhoud, vakmanschap, oorspronkelijkheid en ontwikkeling, is bij de geïnteresseerde leken uit dit onderzoek de som van de scores op twee van deze vier, namelijk **vakmanschap** en **ontwikkeling**, voldoende om dezelfde hoeveelheid van 97% van de variantie in het kwaliteitsoordeel te voorspellen.*

V.3 Vergelijking met de oordelen van experts

De centrale vraag in dit onderzoek is in hoeverre de oordelen van de leken overeenstemmen met die van de experts. In par. III.3.1 zal deze vraag worden uitgewerkt waarna in par. III.3.2 wordt nagegaan in welke mate de leken in absolute zin anders oordelen dan de experts. In de laatste paragraaf wordt gekeken of hypothetische beslissingen over toekenning of afwijzing van de aanvragen door de leken, afgeleid uit hun kwaliteitsoordelen, overeenstemmen met de, volgens dezelfde methode, afgeleide beslissingen van de experts en de werkelijke beslissingen van het Fonds BKVB.

V.3.1 Overeenstemming tussen geïnteresseerde leken en experts

In tabel 7 zijn voor iedere beoordelingsschaal de correlaties tussen de gemiddelde scores van de experts (uit het vorige onderzoek) en van de geïnteresseerde leken (uit het huidige onderzoek) vermeld. Met betrekking tot kwaliteit blijkt er van geen enkele overeenstemming tussen de experts en de geïnteresseerde leken sprake te zijn. Wel zijn de leken het betrekkelijk eens met de experts als het gaat om de oorspronkelijkheid van het werk. Een significante overeenstemming tussen leken en experts werd verder gevonden voor de criteria complexiteit, dynamiek en samenhang. In de eerste paragraaf van hoofdstuk VI zal nader worden ingegaan op deze resultaten⁴⁷.

Het ontbreken van overeenstemming tussen leken en experts aangaande de kwaliteit van de werken is conform de resultaten uit eerdere onderzoeken van Getzels en Csikszentmihalyi en Hendriks die respectievelijk correlaties van .22 en -.08 vonden tussen de kwaliteitsoordelen van beide groepen⁴⁸. De kwaliteitsoordelen van de 'relatieve' experts uit het onderzoek van Hendriks correleerden hoger met de expert-oordelen ($r = .42$) wat erop wijst dat de beoordelaars uit ons onderzoek beschouwd moeten worden als leken.

Ook voor de criteria oorspronkelijkheid en, zij het in mindere mate, vakmanschap zijn de gevonden relaties in redelijke mate overeenkomstig die uit het onderzoek van Getzels en Csikszentmihalyi, waarbij tussen leken en experts op deze criteria correlaties van respectievelijk .60 (versus .65 in dit onderzoek) en .32 (versus .14) gevonden werden.

Tabel 7. Overeenstemming per criterium tussen de gemiddelde scores van de leken en de experts over alle 26 aanvragen.

Leken vs. experts	
Criterion	Alle aanvragen (n=26)
Complexiteit	.42*
Dynamiek	.82**
Samenhang	.82**
Vakmanschap	.14
Inhoud	-.21
Zeggingskracht	-.09
Ontwikkeling	-.08
Oorspronkelijkheid	.65**
Affiniteit	-.35
Interessantheid	-.08
Kwaliteit	-.12

* $p < .05$ ** $p < .01$

V.3.2 *Verschillen in scores op de criteria*

Een toetsing van de verschillen tussen de gemiddelde scores van de geïnteresseerde leken en de experts leverde t -waarden op die varieerden van $t = 2.06$ ($p = .05$) voor samenhang tot $t = 9.10$ ($p < .01$) voor oorspronkelijkheid. Uit de richting van de verschillen bleek dat de leken de beoordeelde werken positiever waardeerden dan de experts.

V.3.3 *Overeenstemming met beslissingen van het Fonds BKVB*

In de vorige twee paragrafen hebben we gezien dat de geïnteresseerde leken in vele opzichten duidelijk anders hebben geoordeeld over de 26 aanvragen dan de experts uit het eerste onderzoek. Dat verschil in oordeel zou uiteraard ook tot uitdrukking komen in toe- of afwijzingen van aanvragen voor stipendia. Om een indicatie hiervan te krijgen zijn voor zowel de 34 experts als voor de 26 geïnteresseerde leken de gemiddelde kwaliteitscores per aanvraag berekend. Deze scores kunnen als betrouwbare schatters van het gemiddelde expert- en lekenoordeel worden beschouwd.

Van de 26 beoordeelde aanvragen waren er door het Fonds 12 direct gehonoreerd, 12 afgewezen en werd er over 2 pas een (positieve) beslissing genomen nadat er een atelierbezoek was afgelegd. Op basis van die verdeling werd zowel bij de leken als de experts, aan de 12 aanvragen met de hoogste gemiddelde kwaliteitscore een 3 (= toewijzing), aan de

daarop volgende twee een 2 (= atelierbezoek) en aan de laagste 12 een 1 (= afwijzing) toegerekend⁴⁹.

Zou de totale groep van experts uit het eerste onderzoek in 5 van de 26 gevallen (ongeveer 19% van de aanvragen) tot een andere beslissing zijn gekomen dan de commissie van het Fonds, de leken zouden over 19 van de 26 aanvragen (73%) een andere beslissing hebben genomen. De overeenstemming tussen de beslissingen kan ook met een Spearman rangorde correlatie coëfficiënt worden weergegeven. Tussen het Fonds en de experts is deze correlatie .71 ($p < .01$), tussen het Fonds en de leken -.33 en tussen de experts en de leken -.08 (zij kwamen over 10 van de 26 aanvragen tot eenzelfde beslissing).

Conclusie 3

Met betrekking tot de kwaliteit van de aanvragen bestaat er geen overeenstemming tussen de oordelen van de leken en die van de experts. Bovendien beoordelen de leken het werk op alle criteria gemiddeld positiever dan de experts. Zouden de beslissingen over toe- of afwijzing van de aanvragen worden overgelaten aan de leken uit dit onderzoek, dan zouden deze in sterke mate afwijken van de beslissingen van het Fonds en die van de experts uit het eerste onderzoek.

VI Discussie

In dit hoofdstuk zullen wij ons beperken tot een bespreking van de belangrijkste resultaten van de twee uitgevoerde onderzoeken. Uitgaande van deze resultaten zal dan in het slot-hoofdstuk een aantal aanbevelingen geformuleerd worden.

VI.1 Verschillen tussen oordelen van experts en geïnteresseerde leken

Op een aantal criteria, te weten complexiteit, dynamiek, samenhang en oorspronkelijkheid was er van een significante overeenstemming tussen de gemiddelde oordelen van experts en leken sprake. Dit werkte echter niet door in de kwaliteitsbeoordelingen van het getoonde werk. Duidelijk werd dat op zulke beoordelingen gebaseerde beslissingen inzake toekenning van subsidies door leken sterk zouden afwijken van die door experts. Al met al bevestigt dit resultaat de veel gehoorde mening dat de voorkeuren van kunstexperts niet gedeeld worden door personen die de benodigde expertise niet bezitten. Zelfs de verschillen tussen de oordelen van experts en die van geïnteresseerde leken, die toch enigszins bekend zijn met beeldende kunst, blijken al ver uiteen te lopen.

Over deze verschillen valt nog het volgende op te merken. Het feit dat oorspronkelijkheid bij de experts wel, en bij de geïnteresseerde leken niet gerelateerd was aan het kwaliteitsoordeel is in overeenstemming met theoretische overwegingen zoals die onder meer door Parsons zijn aangedragen⁵⁰. Hierbij wordt namelijk verondersteld dat experts hun esthetische oordeel voor een belangrijk deel baseren op de traditie waarin het werk geplaatst kan worden. Voegt een werk iets toe aan deze traditie, m.a.w. is het in een bepaald opzicht vernieuwend, dan getuigt het daarmee van kwaliteit. De sterke nadruk in negatieve zin die de experts in het eerste onderzoek aan de dag legden ten aanzien van epigonistische aspecten in het werk, geeft blijk van dit kunst-historische belang.

De geïnteresseerde leken zijn weliswaar in staat de oorspronkelijkheid van het werk te herkennen, maar hun kwaliteitsoordeel wordt daardoor niet of nauwelijks beïnvloed. Dat kwaliteitsoordeel hangt bij de leken, evenals bij de experts, wel in sterke mate samen met het oordeel over vakmanschap. Laatstgenoemd oordeel echter valt, gezien de geringe overeenstemming dienaangaande, bij de leken anders uit dan bij de experts.

Gegeven bovengenoemde verschillen tussen leken en experts zullen in de volgende paragrafen slechts de resultaten van het eerste onderzoek verder worden behandeld.

VI.2 Vergelijking van twee door experts uitgevoerde beoordelingsprocedures

In paragraaf III.4.3 werden de gemiddelde kwaliteitsoordelen vergeleken met de eindoordelen van de commissies die tot stand kwamen na discussie, waarbij in die discussie persoonlijke gegevens van de aanvragers werden meegewogen. Daarbij bleek de eerstgenoemde procedure tot betrouwbaarder resultaten te leiden dan de tweede. Mede daarom werd gesuggereerd de toewijzingsbeslissingen te baseren op de gemiddelde kwaliteitsoordelen en de discussie achterwege te laten.

Om enig inzicht te krijgen in het praktisch effect hiervan is het noodzakelijk de validiteit van beide selectieprocedures vast te stellen. Daartoe zouden we eigenlijk moeten beschikken over een extern criterium dat de 'ware' kwaliteit van de aanvragen aangeeft. Een mogelijk

criterium zou kunnen zijn de mate waarin een aanvrager over vijf of tien jaar succesvol is als kunstenaar. Het blijft echter zeer de vraag of die mate van succes, als die al objectief zou zijn vast te stellen, een werkelijk valide maat is voor de kwaliteit van een kunstenaar. Kortom, gezien het intrinsiek subjectieve karakter van het kwaliteitscriterium zal een voor ieder acceptabel objectief extern criterium wel nooit beschikbaar komen.

Om, gegeven deze situatie, toch iets over de validiteit te kunnen zeggen is een schatting gemaakt van het (fictieve) rendement van de beide beoordelingsprocedures (zie bijlage 1). Hieruit is gebleken dat de procedure zonder discussie en kennis van persoonlijke gegevens (procedure 2) een kleine verbetering in rendement oplevert ten opzichte van de procedure waarbij wel een discussie plaatsvond en informatie uit het aanvraagformulier werd betrokken (procedure 1). Deze verbetering is op zichzelf wellicht te gering om procedure 2 duidelijk te prefereren. Zoals reeds kort aangeduid in par. III.4.3 zien wij echter ook bijkomstige voordelen in procedure 2. Achterwege laten van zowel de discussie als de verstrekking van persoonlijke gegevens van de aanvragers leidt ertoe dat:

1. de beslissingen niet in onevenredige mate beïnvloed kunnen worden door dominerende commissieleden of tot stand komen op basis van andere, ongewenste, groeps-dynamische processen die bij de besluitvorming kunnen optreden;
2. de beslissingen louter op het werk betrekking hebben. Observaties tijdens de discussie brachten onder meer aan het licht dat bepaalde gegevens uit het informatie-formulier door de beoordelaars op zeer verschillende wijze in het oordeel werden betrokken. Deze subjectieve interpretatie van de verstrekte informatie wordt met procedure 2 vermeden. Bovendien weten de beoordeelde aanvragers in dat geval dat de beslissingen niet beïnvloed zijn door persoonlijke kenmerken als sexe, regionale herkomst, opleidingsinstituut etc..

Daarnaast kan nog het volgende aangetekend worden. Wanneer zowel de resultaten van procedure 1 als procedure 2 vergeleken worden met de door het Fonds genomen beslissingen, dan blijkt procedure 1 tot een wat lagere overeenstemming met de Fondsoordelen te leiden dan procedure 2 (de desbetreffende correlaties waren voor zitting II respectievelijk .85 en .89 en voor zitting III .40 en .53). Waar bij de beslissingen van het Fonds persoonlijke gegevens worden verstrekt en via discussie een eindoordeel gevormd wordt, leidt introductie van deze elementen in ons onderzoek dus tot een lagere overeenstemming met de Fondsbeslissingen dan wanneer ze achterwege worden gelaten. Dit sterkt ons in onze mening dat de persoonlijke gegevens en de discussie veeleer tot 'ruis' dan tot een adequater eindoordeel leiden.

In het uitgevoerde onderzoek zijn het al dan niet verstrekken van persoonlijke gegevens en het al dan niet plaats laten vinden van een discussie niet onafhankelijk van elkaar gemanipuleerd (d.w.z. de twee elementen kwamen geen van beide, of beide tegelijk voor), zodat geen definitieve uitspraken gedaan kunnen worden over de vraag naar het effect van het verstrekken van persoonlijke gegevens zonder dat een discussie plaats vindt, of het effect van een discussie waarbij persoonlijke gegevens onbekend zijn. De zojuist aangevoerde overwegingen achten wij echter dermate relevant dat het ons desondanks beter voorkomt zowel van verstrekking van persoonlijke gegevens als van discussie af te zien.

Het door ons gevoerde pleidooi om bij de kwaliteitsbeoordeling de discussie achterwege te laten is reeds eerder door Amabile gevoerd met betrekking tot oordelen over creativiteit van kunstwerken⁵¹. Een dergelijk advies is echter niet in overeenstemming met de opvatting van Cramer, waar zij opmerkt dat juist de discussie zou leiden tot een oordeel "...dat meer is dan de som van zes afzonderlijke, subjectieve stemmen: en in die meerwaarde ligt het objectieve element van het beoordelingsproces besloten"⁵². Het zal inmiddels duidelijk zijn dat wij zowel op theoretische als op empirische gronden deze opvatting menen te moeten afwijzen.

VI.3 Vervanging van het kwaliteitsoordeel door een oordeel gebaseerd op criteriumscores

In paragraaf III.3 werd aangetoond dat de individuele kwaliteitsoordelen waarop in procedure 2 uiteindelijk de beslissing wordt gebaseerd, slechts in zeer geringe mate veranderen indien deze worden vervangen door de gesommeerde individuele scores op de criteria inhoud, vakmanschap, oorspronkelijkheid en ontwikkeling. De over alle commissieleden gemiddelde scores op deze vier criteria bleken immers gezamenlijk het gemiddelde kwaliteitsoordeel voor 97% te kunnen verklaren. Gegeven dit zeer sterke verband zal er in psychometrisch opzicht zeer weinig veranderen indien in plaats van de gemiddelde, direct aangegeven kwaliteitscores, gebruik gemaakt wordt van een kwaliteitscore die tot stand komt via optelling, c.q. middeling van de individuele scores op de genoemde criteria. Een voordeel hiervan is dat de gemiddelde scores op de vier criteria desgewenst aan de aanvrager bekend kunnen worden gemaakt. Tot op heden luidt de motivatie van het Fonds met betrekking tot de afwijzing van een aanvraag bijvoorbeeld als volgt: "De werkgroep acht het werk van de aanvrager qua vorm en inhoud niet voldoende uitgesproken. Het werk bezit naar mening van de werkgroep niet voldoende autonome beeldende kwaliteit" (Uit brieven van het fondsbestuur aan drie kunstenaars die hun aanvragen voor een startstipendium afgewezen zagen). Via expliciete scores op de vier relevante criteria kan de kunstenaars een genuanceerder inzicht worden geboden in de beoordeling van hun aanvraag.

VI.4 Generaliseerbaarheid van de resultaten

In de nu volgende bespreking van de generaliseerbaarheid van de onderzoeksresultaten zullen we ons beperken tot mogelijke generaliseringen naar andere vormen van individuele subsidie bij het Fonds BKVB. Daarbij wordt uiteraard tevens alleen gesproken over aanvragen op het terrein van de beeldende kunst.

In dit onderzoek zijn uitsluitend aanvragen voor een startstipendium beoordeeld, een selectie waarvan de beoordelaars op de hoogte waren gesteld. De indieners behoorden dus alle tot een specifieke groep beginnende kunstenaars, die minder dan twee jaar geleden hun opleiding hadden afgerond (zie § II.3.1). Daarom is het niet mogelijk 'keiharde' uitspraken te doen betreffende generaliseerbaarheid van de resultaten naar andere subsidiesoorten. Er lijkt echter weinig redenen toe om aan te nemen dat de door ons bestudeerde procedures bij de beoordeling van **werkbeurzen** tot sterk afwijkende resultaten zouden leiden. Net als de startstipendia hebben de werkbeurzen namelijk tot doel "om kunstenaars de tijd te gunnen zich op hun scheppend werk te concentreren zonder de druk van 'brood-op-de-plank'"⁵³. Daar een werkbeurs juist kan worden aangevraagd door kunstenaars die langer dan twee jaar professioneel actief zijn, kan overwogen worden, voorafgaande aan de beoordeling van dit type aanvraag, aan te geven hoe lang de aanvrager reeds als kunstenaar werkzaam is. Deze variabele kan bij de startstipendia uiteraard nog nauwelijks een rol spelen.

Wat betreft de vier andere vormen van individuele subsidies, te weten studiebeurzen, reisbeurzen, projectsubsidies en presentatiesubsidies, lijkt generalisering van de resultaten niet zonder meer geoorloofd. Bij deze subsidiesoorten speelt de motivatie van de aanvrager om juist voor die specifieke opleiding, reisbestemming, project of presentatie een aanvraag in te dienen een belangrijke rol. Overigens nemen deze laatstgenoemde vier subsidievormen tezamen slechts 20.9% van het totaal aantal aanvragen individuele subsidies voor hun reke-

ning, terwijl de aanvragen voor startstipendia en werkbeurzen respectievelijk 40.6% en 38.5% van alle aanvragen uitmaken⁵⁴.

VII Aanbevelingen

De in hoofdstuk 3 gepresenteerde conclusies en de in het vorige hoofdstuk gemaakte overwegingen leiden tot de hieronder genoemde aanbevelingen ten aanzien van de bij de beoordeling van aanvragen voor een startstipendium te volgen procedure.

1. *Beoordelingscommissies bestaan uit zes deskundigen, waaronder kunstbeschouwers en kunstenaars. Het percentage kunstenaars (c.q. kunstbeschouwers) binnen de verschillende commissies wordt constant gehouden.*
2. *De commissieleden blijven onkundig van persoonlijke gegevens van de aanvragers.*
3. *Ieder commissielid beoordeelt alle, aan de commissie voorgelegde, aanvragen individueel en onafhankelijk van de andere commissieleden op de vier criteria Inhoud, Vakmanschap, Oorspronkelijkheid en Ontwikkeling. Deze beoordelingen worden gestructureerd uitgevoerd op een 9-punts schaal.*
4. *Per criterium worden de individueel gegeven scores gemiddeld over de zes beoordelaars. Deze vier gemiddelde scores kunnen desgewenst aan de aanvrager bekend worden gemaakt.*
5. *Per commissielid worden tevens de scores op de vier criteria opgeteld. De resulterende somscore wordt beschouwd als de operationalisering van het individuele kwaliteitsoordeel over een aanvraag. Het uiteindelijke kwaliteitsoordeel over een aanvraag komt tot stand door de individuele somscores over de zes beoordelaars te middelen.*
6. *De binnen een commissie behandelde aanvragen worden op grond van de aldus geconstrueerde kwaliteitsoordelen geprioriteerd. Vervolgens beslist de commissie in onderling overleg (1) welke aanvragen direct gehonoreerd worden, (2) welke aanvragen een atelierbezoek vereisen en (3) welke aanvragen direct afgewezen worden, dit echter met dien verstande dat de prioriteringen van aanvragen binnen categorie (1) stuk voor stuk hoger zijn dan binnen categorie (2), en dat die op hun beurt weer stuk voor stuk hoger zijn dan die binnen categorie (3).*

VII.1 Kanttekeningen bij de aanbevelingen

Hieronder volgen nog enige opmerkingen met betrekking tot de aanbevelingen. Daarbij worden deze waar mogelijk vergeleken met aanbevelingen betreffende een beoordelingssysteem zoals die in de al vaker aangehaalde discussienota van de BBK werden geformuleerd⁵⁵. ad 1:

In de discussienota van de BBK luidt de eerste aanbeveling dat *beoordelaars gezocht moeten worden onder kunstenaars, kunsthistorici of andere beschouwers*. Onze eerste aanbeveling is hiermee in overeenstemming. Uit het onderzoek is bovendien duidelijk naar voren gekomen dat commissies bestaande uit (geïnteresseerde) leden tot andere beslissingen zouden komen. Op grond daarvan zien we geen reden de huidige samenstelling van vier kunstenaars en twee kunstbeschouwers in een commissie te wijzigen.

ad 2:

De aanvrager zou nu kunnen volstaan met het verstrekken van enige persoonlijke gegevens ten behoeve van de statistiek. Naast de al genoemde voordelen van deze aanbeveling wordt de aanvrager dan tevens ontslagen van het invullen van het aanvraagformulier. Deze activiteit wordt door vele kunstenaars als een zware belasting ervaren.

ad 3:

Ook deze aanbeveling is grotendeels gelijk aan die in de nota van de BBK waar gesteld wordt dat *het beoordelen van kwaliteit dient te geschieden aan de hand van de aspecten vakmatigheid, zeggingskracht en oorspronkelijkheid*. Over het criterium zeggingskracht werd door Linnemann al eerder opgemerkt dat het daarbij eigenlijk hoort te gaan "om hoe, en in hoeverre, aan een *inhoud* adequaat vorm wordt gegeven"⁵⁶. De resultaten uit het onderzoek bevestigen dat zeggingskracht en inhoud zo goed als identieke criteria zijn⁵⁷.

Het vierde uit ons onderzoek als relevant naar voren gekomen criterium ontwikkeling is wellicht alleen van toepassing op aanvragen voor een startstipendium. Aangezien de aanbevelingen uit de nota betrekking hebben op een procedure voor kunstbeoordelingen in het algemeen is het niet verwonderlijk dat dit criterium daar buiten beschouwing is gelaten. De drie criteria inhoud, vakmanschap en oorspronkelijkheid verklaarden in ons onderzoek samen al 96% van de variantie in het kwaliteitsoordeel zodat het criterium ontwikkeling, bijvoorbeeld bij andere subsidiesoorten, zonder al te veel verlies aan validiteit van de somscore, achterwege kan worden gelaten.

ad 4:

Met deze aanbeveling wordt eveneens tegemoet gekomen aan een aanbeveling uit de nota: *In beoordelingssystematieken zal er mee rekening gehouden moeten worden dat een verantwoording achteraf op adequate wijze mogelijk is*.

ad 5:

In de nota werd gesteld dat *minderheidsstandpunten nooit een onevenredig zware invloed op het uiteindelijke commissieoordeel mogen hebben*. Door af te zien van discussie tussen de commissieleden wordt hierin al (deels) voorzien. Zo wordt immers voorkomen dat een afwijkend oordeel van een dominant lid tot groepsoordeel kan uitgroeien.

Een verdere correctie voor extreme oordelen zou kunnen worden bewerkstelligd door de laagste en hoogste scores niet mee te nemen in de berekening van het uiteindelijke kwaliteitsoordeel. Deze methode, waarbij de beide uiterste scores vervallen, komt overeen met jurybeoordelingen in verschillende takken van sport (bijv. kunstrijden en schoonspringen) en met de scoringsprocedure bij een aantal instanties waar wetenschappelijke aanvragen worden beoordeeld. Deze methode is in dit onderzoek niet toegepast waardoor er geen uitspraak gedaan kan worden over de grootte van het effect op de betrouwbaarheid en validiteit van de score die aldus ontstaat. Het lijkt evenwel zeer onwaarschijnlijk dat dat effect negatief zal zijn.

ad 6:

Het is niet op voorhand uit te sluiten dat verschillende commissies verschillen in de 'strengheid' waarmee de criteria gehanteerd worden. Het is daarom niet mogelijk absolute waarden van de eindscores vast te stellen waarboven aanvragen worden gehonoreerd, dan wel waaronder deze worden afgewezen. Daarom is de aanbeveling gedaan de eindscore alleen te gebruiken in relatieve zin, voor een prioriteitsbepaling van iedere aanvraag ten opzichte van de andere aanvragen.

Met deze procedure wordt het de commissies mogelijk gemaakt zelf een beslissing te koppelen aan de eindscores. Zo zal het kunnen voorkomen dat, over een totaal van bijvoorbeeld 20 behandelde aanvragen, de ene keer de eerste zeven aanvragen op de lijst

gehonoreerd worden (35%) terwijl een volgende keer slechts aan de twee hoogst geprioriteerde aanvragen een subsidie wordt toegekend (10%). Overigens kan met verloop van tijd nagegaan worden in hoeverre er inderdaad sprake is van systematische verschillen tussen de commissies met betrekking tot de bij toe- en afwijzing gehanteerde criteriumscores.

In geval van twijfel over een aanvraag kan besloten worden over te gaan tot een atelierbezoek. Deze aanvraag wordt dan in een later stadium opnieuw besproken, net zoals dat nu het geval is.

In de BBK-nota werd gesteld dat *discussie tussen deskundigen een belangrijke plaats moet innemen in de beoordelingssystematiek*. Gezien de bij aanbeveling 5 gemaakte opmerkingen en de alom, óók in de genoemde nota, uitgesproken bedenkingen over de sociale processen die de consensusvorming in zo'n discussie kunnen bepalen, is de zeer beperkte rol die onderling overleg in de aanbevolen procedure inneemt echter gerechtvaardigd.

Met uitzondering dus van het belang van de discussie, wordt met de aanbevolen procedure, die is opgesteld naar aanleiding van het onderzoek, gevolg gegeven aan de aanbevelingen uit de discussienota, die gebaseerd zijn op een inhoudelijke bespreking van de beoordelingsprocedure.

VII.2 Voordelen van de aanbevolen procedure

De aanbevolen beoordelingsprocedure heeft ten opzichte van de huidige procedure bij de commissie IS van het Fonds BKVB de volgende **vijf voordelen**:

1 Alle commissieleden zijn gedwongen een persoonlijk, onafhankelijk oordeel te geven. Het is niet langer mogelijk het eigen oordeel aan te passen aan dat van anderen of je te verschuilen achter de mening van de commissie. Ieder lid heeft zijn **individuele verantwoordelijkheid**.

2 Ieder commissielid moet het werk op **dezelfde subjectieve criteria** beoordelen.

3 De gemiddelde criteriumscores geven de betrokkenen **meer inzicht** in de wijze waarop het kwaliteitsoordeel tot stand is gekomen.

4 De voorgestelde eindscore, het gemiddelde van zes onafhankelijke somscores, is **betrouwbaarder** dan een eindoordeel dat tot stand is gekomen na een discussie en waarin persoonlijke gegevens van de aanvrager zijn betrokken.

5 Mogelijke, **ongewenste sociale processen** die tijdens een discussie kunnen optreden worden zo goed als **uitgesloten**. Weliswaar vindt er uiteindelijk onderling overleg plaats over toekenning van de aanvragen, maar de prioritering staat dan al vast. In vergelijking met de bestaande beoordelingsprocedure bij het Fonds BKVB is het belang van de discussie dus aanzienlijk teruggedrongen.

Noten

1. Hekkert & van Wieringen, 1992a.
2. Hekkert & van Wieringen, 1992b.
3. Hofstee, 1983.
4. Hofstee, 1983, p. 273.
5. Vermeld in P. van Klink. D'Ancona wordt te vroeg aan de schandpaal genageld. *NRC Handelsblad*, 6 juni 1991.
6. Abbing, 1989.
7. Oosterbaan Martinius, 1990.
8. Oosterbaan Martinius, 1990, p. 69.
9. Zie 'Evaluatie van het Beleid van de minister van WVC op het gebied van de Beeldende Kunsten, Bouwkunst en Vormgeving in het tijdvak 1984 t/m 1987', september 1988, p. 4.
10. Deze, aan Abraham de Swaan ontleende, term komt overeen met het aloude 'l'art pour l'art'. In zijn artikel 'Kunstkunst en gunstkunst' en het essay 'Kwaliteit is klasse' gaat de Swaan uitgebreid in op de vermeende onverenigbaarheid van een publieks- en een kwaliteitsgericht beleid (de Swaan, 1990a en 1990b).
11. Zie Van Grootheest, 1991.
12. Zie 'Informatiebrochure Hoofdafdeling Beeldende kunst, Bouwkunst en Vormgeving van het Ministerie van Welzijn, Volksgezondheid en Cultuur'. Rijswijk, februari 1989.
13. De hier gehanteerde tweedeling van adviesraden en fondsen is ontleend aan Oosterhuis (1990) en is naar zijn zeggen "een grove indeling die geen recht doet aan de juridische basis van deze instellingen" (p. 87).
14. Linnemann, 1991, p. 18.
15. Vermeld in L. Delfgaauw. Heftige discussies en moeilijke beslissingen: De beroepskostenvergadering. *Nieuwsbrief van het Fonds voor beeldende kunsten, vormgeving en bouwkunst*, 2, 1989.
16. Hegel, 1989.
17. Kant, 1966.
18. Kant, 1966, p. 3.
19. Lindauer & Long, 1986, p. 172.
20. Kraaijpoel, 1990, p. 110.
Een opvatting welke volledig indruist tegen deze voorwaarde van smaakonderdrukking is die van de museumdirecteur Mr. E. de Wilde : "Het is mijn ervaring dat elke andere overweging als die van de persoonlijke voorkeur tot een slechte keuze leidt" (vermeld in Kraaijpoel, 1990,

p. 162).

21. In de eerste *Nieuwsbrief* van het Fonds BKVB geeft Cato Cramer, voorzitter van de Adviescommissie Individuele Subsidies van dit fonds, inzicht in de criteria die het Fonds hanteert bij de selectie van haar commissieleden en bij de beoordelingen van de subsidieaanvragen. Aan dit artikel zal dan ook regelmatig worden gerefereerd. Zie: Cramer, 1988.

22. In Oosterhuis, 1990.

23. Wollheim, 1980, p. 70.

24. Cramer, 1988.

25. Cramer, 1988, p. 5.

26. Oosterbaan Martinius, 1990, p. 142.

27. Oosterbaan Martinius, 1990, p. 144.

28. Linnemann, 1991, p. 16.

29. Cramer, 1988, p. 4.

30. Cramer, 1988.

31. Lindauer & Long, 1986.

32. Van der Tas, 1990.

Dit criterium vinden we ook terug bij Poll (1972) die in dit verband spreekt van beeldspraak en daarmee wijst op "de spanning tussen kunstwerk en omringende realiteit". Een derde auteur, die reeds lang geleden benadrukte dat een kunstwerk meer moet zijn dan een natuurprodukt, d.w.z. een produkt dat louter en alleen op zichzelf staat en naar zichzelf verwijst (het 'l'art pour l'art'-idee), is Hegel (1989). Een kunstvoorwerp dient naast zintuiglijk object tevens wezenlijk voor de geest te bestaan, "opdat de geest erdoor beïnvloed zal worden en er een of andere bevrediging uit zal halen" (p. 87).

33. Zie o.a. Beardsley, 1970.

34. Naast deze 11 criteria waren in het onderzoek nog twee criteria opgenomen, namelijk **interessantheid** en **persoonlijke affiniteit**. Deze twee criteria hebben, net als kwaliteit, betrekking op een globaal of totaal-oordeel over het werk. Uit de zeer hoge correlaties van deze criteria met het kwaliteitsoordeel bleek dat de beoordelaars hiertussen nauwelijks onderscheid maakten. Met de in dit onderzoek gehanteerde methode bleek het dus niet mogelijk oordelen over de kwaliteit te scheiden van de persoonlijke smaak, de affiniteit met het werk. Zoals in de inleiding al naar voren gebracht (zie § I.3) lijkt het derhalve een onhaalbare eis de eigen voorkeur uit te schakelen in geval van kwaliteitsoordelen. Daarmee blijft de vraag staan of deze twee typen oordelen niet meer zijn dan een talig dan wel logisch onderscheid waar er in de praktijk geen verschil zichtbaar is.

35. Champney & Marshall, 1939.

36. Voor iedere zitting waaraan zij deelnamen ontvingen de beoordelaars een vergoeding conform het gebruikelijke honorarium voor leden van een beoordelingscommissie.

37. Na mondelinge toestemming van alle beoordelaars zijn de discussies op cassette opgenomen. Een kwalitatieve analyse van het argumentatie-proces tijdens deze discussies is uitgevoerd door Dekker de Vos (1992).

38. Naast dit commissie-oordeel is de beoordelaars ook gevraagd direkt na afloop van de discussie een individuele prioritering te geven van de aanvraag. In de praktijk bleek men dit persoonlijke oordeel vaak af te stemmen op het eindoordeel van de commissie. Hierop hebben we besloten deze prioriteitsoordelen bij de analyse van de resultaten buiten beschouwing te laten.

39. Respectievelijk Getzels & Csikszentmihalyi, 1976 en Hendriks, 1988.

40. Zie o.a. Hofstee, 1983; Scheerens & Beem, 1986; Cicchetti, 1991 en Martindale & Dailey, 1992.

41. Dit gegeven doet vermoeden dat de beoordelaars het meer eens zijn over welke aanvragen zeker afgewezen dienen te worden dan over te honoreren aanvragen, een tendens die in overeenstemming is met de praktijk bij de beoordelaars van wetenschappelijke manuscripten waar "reviewer agreement on decisions to reject manuscripts is appreciably higher than agreement on acceptance" (Cicchetti, 1991; p. 126).

42. Hierbij is uitgegaan van de wijze waarop de drie alternatieven waren verdeeld over de eindoordelen. Dit leidde tot de volgende regel voor de codering van de gemiddelde kwaliteitscores: gem. score $< 4.50 = 1$, $4.50 \leq \text{gem. score} \leq 5.00 = 2$ en gem. score $> 5.00 = 3$.

43. Door de discussie achterwege te laten wordt immers voorkomen dat ongewenste groepsdynamische processen een rol gaan spelen bij de besluitvorming.

44. Zie o.a. Cicchetti, 1991, p. 130.

Met de Spearman-Brown formule voor testverlenging is het mogelijk de betrouwbaarheid te schatten van gemiddelde kwaliteitscores gebaseerd op verschillende aantallen beoordelaars (Drenth, 1975, p. 226 e.v.).

45. Getzels & Csikszentmihalyi, 1976.

46. Hendriks, 1988.

47. Teneinde een meer gedetailleerd beeld te verkrijgen van de mate van overeenstemming tussen de geïnteresseerde leken en de experts zijn de 26 clusters dia's verdeeld in twee categorieën. De eerste categorie werd gevormd door 17 clusters bestaande uit (dia's van) schilderijen, tekeningen of grafiek, de tweede categorie van 8 clusters bevatte voornamelijk (dia's van) ruimtelijk werk. De kunstwerken uit deze laatste categorie waren overwegend conceptueel van aard.

Uit de over beide categorieën afzonderlijk uitgevoerde analyses bleek onder meer dat de overeenstemming over de oorspronkelijkheid van het werk alleen gold voor de eerste categorie van picturale werken. Over de beoordeling van de oorspronkelijkheid van de ruimtelijke werken waren de leken het oneens met de experts. Ook op de andere criteria bleek de overeenstemming tussen de leken en de experts duidelijk geringer ten aanzien van het ruimtelijke werk. Kennelijk is de aard van dit werk voor de geïnteresseerde leken dermate nieuw en onbekend dat zij moeite hebben met de beoordeling en waardering ervan. De leken

achtten de kwaliteit van dit type werk dan ook niet beter dan dat van de picturale werken, in tegenstelling tot de experts die het ruimtelijke werk significant hoger beoordeelden op kwaliteit (en op de andere criteria).

Voor een uitvoerige bespreking van de analyses over de afzonderlijke categorieën verwijzen wij verder naar het tweede onderzoeksrapport (Hekkert & van Wieringen, 1992b).

48. Zie Getzels & Csikszentmihalyi, 1976 en Hendriks, 1988.

49. De aldus geconstrueerde fictieve beslissingen van de experts wijken dus af van de in het eerste onderzoek gerapporteerde commissieoordelen en gemiddelde kwaliteitscores over 4 tot 6 experts. Hier wordt immers gebruik gemaakt van het gemiddelde kwaliteitsoordeel van alle 34 beoordelaars.

50. In zijn boek "How we understand art" onderscheidt Parsons (1987) op basis van een groot aantal interviews vijf ontwikkelingsstadia in het begrijpen van kunst die samenhangen met leeftijd, algemene en culturele vorming en geoefendheid in het kijken naar kunstwerken. Zijn theorie vertoont in vele opzichten verwantschap met die van Bourdieu (1989) en beide theorieën worden uitvoerig behandeld in de originele onderzoeksrapporten.

51. Zie Amabile, 1982.

52. Cramer, 1988, p.5.

53. Uit de toelichting bij het aanvraagformulier *Individuele subsidie beeldende kunsten*, Stichting Fonds voor beeldende kunsten, vormgeving en bouwkunst.

54. Deze percentages zijn gebaseerd op het totaal aantal aanvragen voor een individuele subsidie in 1988, 1989 en 1990, zie *Het Fonds in cijfers*, Amsterdam: Uitgave van Stichting Fonds voor beeldende kunsten, vormgeving en bouwkunst.

55. Linnemann, 1991, p. 20.

56. Linnemann, 1991, p. 14.

57. De correlatie tussen de scores op de criteria inhoud en zeggingskracht bedroeg .98. Vervanging van een inhoudscore door een score op zeggingskracht zal derhalve nauwelijks van invloed zijn op de somscore.

Referenties

- Abbing, H. (1989). *Een economie van de kunsten. Beschouwingen over kunst en kunstbeleid*. Groningen: Historische Uitgeverij.
- Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, **43**, 997-1013.
- Beardsley, M. C. (1970). *The possibility of criticism*. Detroit: Wayne State University Press.
- Bourdieu, P. (1989). *Distinction: A social critique of the judgement of taste*. London: Routledge. Oorspronkelijk verschenen als: *La distinction. Critique sociale du jugement*. Paris: Les Editions de Minuit, 1979.
- Champney, H. & Marshall, H. (1939). Optimal refinement of the rating scale. *Journal of Applied Psychology*, **23**, 323-331.
- Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, **14**, 119-186.
- Cramer, C. (1988). Het oordelen over kunst. *Nieuwsbrief van het Fonds voor beeldende kunsten, vormgeving en bouwkunst*, **1**, 4-5.
- Dekker de Vos, M. (1992). *De ontwikkeling van smaak: De individuele ontwikkeling van het esthetische oordeelsvermogen*. Doctoraalscriptie vakgroep Persoonlijkheidspsychologie, Rijks Universiteit Leiden.
- Drenth, P. J. D. (1975). *Inleiding in de testtheorie*. Deventer: Van Loghum Slaterus.
- Getzels, J. W. & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art*. New York: Wiley-Interscience.
- Grootheest, T. van (1991). Lezing opgenomen in symposium verslag: *Voorbeeldige opdrachten*. Groningen: Centrum Beeldende Kunst.
- Hegel, G. W. F. (1989). *Over de esthetiek*. Amsterdam: Boom Meppel. Oorspronkelijk verschenen als 'Einleitung' tot: *Vorlesungen über die Ästhetik*, 1835.
- Hekkert, P. & Wieringen, P. C. W., van (1992a). *Kwaliteit als selectie criterium: Een empirisch onderzoek naar kwaliteitsbeoordelingen in de beeldende kunst*. Onderzoeksrapport in opdracht van het ministerie van W.V.C..
- Hekkert, P. & Wieringen, P. C. W., van (1992b). *Geïnteresseerde leken versus experts: Een vergelijkend empirisch onderzoek naar kwaliteitsbeoordelingen in de beeldende kunst*. Onderzoeksrapport in opdracht van het ministerie van W.V.C..
- Hendriks, H. (1988). *Kwaliteitsbeoordeling in de beeldende kunst: Een empirisch onderzoek naar overeenstemming en sekse-bias in de oordelen*. Heymans Bulletins, HB-88-941-SW, Rijksuniversiteit Groningen.
- Hofstee, W. K. B. (1983). Beoordelingen van subsidie-aanvragen voor onderwijsresearch: Een psychometrische evaluatie. *Tijdschrift voor Onderwijsresearch*, **8**, 273-284.
- Kant, I. (1966). *Kritik der Urteilskraft und Schriften der Naturphilosophie*. Darmstadt: Weischedel (oorspronkelijke publicatie 1790).
- Kraaijpoel, D. (1990). *De nieuwe salon*. Groningen: Academie Minerva Pers.
- Lindauer, M. S. & Long, D. A. (1986). The criteria used to judge art: Marketplace and academic comparisons. *Empirical Studies of the Arts*, **4**, 163-174.
- Linnemann, F. (1991). *Kwaliteit, professionaliteit en beroepserkenning*. Amsterdam: Uitgave BBK.

- Martindale, C. & Dailey, A. (1992). The reliability of literary judgments. *Abstracts of the XIIIth International Congress on Empirical Aesthetics*, p. 39.
- Oosterbaan Martinius, W. (1990). *Schoonheid, welzijn, kwaliteit: Kunstbeleid en verantwoording na 1945*. Meppel: Ten Brink.
- Oosterhuis, R. (1990). Advisering over kunst en cultuur in Nederland. In: *Kunst en overheid: Beleid en praktijk*, pp. 83-105. Amsterdam: Boekmanstichting/Universiteit van Amsterdam.
- Parsons, M. J. (1987). *How we understand art: A cognitive developmental account of aesthetic experience*. Cambridge: Cambridge University Press.
- Poll, K. L. (1972). *De bekleemde elite*. Amsterdam: Meulenhoff.
- Scheerens, J. & Beem, A.L. (1986). Ratings of requests for grants in educational research. *Tijdschrift voor Onderwijsresearch*, **11**, 67-82.
- Swaan, A. de (1990a). Kunstkunst en gunstkunst. In: *Kunst en staat*, pp. 33-35. Amsterdam: Boekmanstichting/Vakgroep Culturele Studies Universiteit van Amsterdam (Oorspronkelijke publicatie 1982).
- Swaan, A. de (1990b). Kwaliteit is klasse. In: *Kunst en staat*, pp. 49-70. Amsterdam: Boekmanstichting/Vakgroep Culturele Studies Universiteit van Amsterdam (oorspronkelijke publicatie 1985).
- Tas, J. M. van der (1990). De objectivering van de subjectiviteit. *Boekmancahier*, **5**, 283-289.
- Wollheim, R. (1980). *Art and its objects. Second edition, with six supplementary essays*. Cambridge: Cambridge University Press.

Bijlage 1

Het meest acceptabele alternatief om iets over de validiteit van de in het onderzoek gehanteerde beoordelingsprocedures te kunnen zeggen wordt bereikt door (1) het gemiddelde oordeel van alle denkbare deskundigen als criterium voor kwaliteit te nemen en (2) de beoordelaars uit het onderzoek als aselekt gekozen personen uit die populatie van alle denkbare deskundigen te beschouwen. Uitgaande van de juistheid van deze veronderstellingen wordt voor procedure 1 (eendoordelen gevormd na discussie en met kennisneming van persoonlijke gegevens) en 2 (eendoordelen gebaseerd op gemiddelde kwaliteitscores) een validiteit van respectievelijk .61 en .66^a vastgesteld. Op basis van deze waarden kan in principe worden nagegaan in hoeverre de rendementen van bovengenoemde selectieprocedures van elkaar verschillen. Daartoe is het echter wel noodzakelijk dat een tweetal grootheden bekend zijn, c.q. geschat worden. Allereerst gaat het daarbij om de het percentage aanvragen dat gehonoreerd wordt (het **selectiepercentage**). Aangezien van de aanvragers van een startstipendium bij het Fonds in 1988, 1989 en 1990 ongeveer 20% gehonoreerd werd^b, lijkt het verantwoord dit percentage als selectiepercentage aan te houden. Vervolgens dient het percentage van alle aanvragers vastgesteld te worden, dat gehonoreerd zou *moeten* worden. Dit percentage is gelijk aan de **toevalskans**, d.w.z. de kans dat bij een aselekte (op toeval gebaseerde) toewijzing van een willekeurige aanvraag, die aanvraag het ook inderdaad verdiende toegewezen te worden. Er van uitgaande dat het selectiepercentage van 20% impliceert dat in feite ook ongeveer een dergelijk percentage van de aanvragen het verdient gehonoreerd te worden, zal de toevalskans daarvan niet al te sterk afwijken en stellen we deze eveneens op 20%.

Op grond van deze assumpties kunnen nu de **succespercentages** van de beide selectieprocedures worden vastgesteld, d.w.z. de kans dat, gegeven de desbetreffende selectieprocedure, een aanvraag die het verdient gehonoreerd te worden dat ook inderdaad wordt. Uitgaande van deze selectiepercentages is in tabel B1 weergegeven wat het resultaat zou zijn van procedure 1 en 2, alsmede van een aselekte toewijzingsprocedure, als er 500 aanvragen beoordeeld zouden worden.

Tabel B1. Fictieve resultaten van op procedure 1 (validiteit = .61), procedure 2 (validiteit = .66) en toeval gebaseerde beslissingen over 500 aanvragen bij een selectieratio van .20 en een toevalskans van .20.

Beslissing op basis van	Toegewezen		Afgewezen	
	<i>terecht</i>	<i>onterecht</i>	<i>terecht</i>	<i>onterecht</i>
Procedure 1	50	50	350	50
Procedure 2	53	47	353	47
Toeval	20	80	320	80

^a Deze waarde is gebaseerd op de betrouwbaarheid van de gecodeerde gemiddelde kwaliteitscores.

^b Deze gegevens zijn ontleend aan: *Het Fonds in cijfers*. Amsterdam: Uitgave van Stichting Fonds voor beeldende kunsten, vormgeving en bouwkunst.

De resultaten van procedure 1 en 2 hadden slechts betrekking op 20 van de 30 in het onderzoek betrokken aanvragen (zie § III.4). Deze 20 aanvragen waren niet aselekt uit de 30 beschikbare aanvragen getrokken, maar de 10 aanvragen die tijdens zitting I tot de meeste divergentie in de kwaliteitsbeoordeling hadden geleid maakten er deel van uit. Bovendien kwam het nogal eens voor dat de commissies niet uit zes, maar slechts uit vier of vijf leden bestonden. Daarom is het beter het rendement van de procedures te schatten op grond van de resultaten ten aanzien van alle dertig aanvragen en er daarbij van uit te gaan dat de commissies uit zes leden bestaan. Omdat eindoordeelen van de gezamenlijke groepen slechts ten aanzien van 20 aanvragen bekend zijn, maar de individuele oordelen over 30 aanvragen, is dit slechts mogelijk voor procedure 2. Uitgaande van dezelfde premissen als gehanteerd in de vorige paragraaf wordt nu een validiteit van .83 bereikt^c. Bij hetzelfde selectiepercentage van 20% en een toevalskans van 20% leidt dat tot een succespercentage van 67%. Gaan we opnieuw uit van 500 te behandelen aanvragen, dan zou deze procedure tot de in tabel B2 weergegeven beslissingen leiden. De in deze tabel gepresenteerde getallen geven inzicht in wat de door ons voorgestelde procedure aan winst oplevert relatief ten opzichte van een volledig willekeurige toe- of afwijzing van de aanvragen.

Tabel B2. Fictieve resultaten van op procedure 2 (validiteit = .83) en toeval gebaseerde beslissingen over 500 aanvragen bij een selectieratio van .20 en een toevalskans van .20.

Beslissing op basis van	Toegewezen		Afgewezen	
	<i>terecht</i>	<i>onterecht</i>	<i>terecht</i>	<i>onterecht</i>
Procedure 2	67	33	367	33
Toeval	20	80	320	80

Gezien de hierboven uitgevoerde vergelijking ligt het in de lijn der verwachting dat bekendmaking van persoonlijke gegevens, in combinatie met een discussie tussen de commissieleden tot een, zij het kleine, vermindering van het rendement van de selectieprocedure zal leiden.

^c Psychometrische redenen nopen ertoe deze waarde als een maximum te beschouwen, het is aannemelijk dat de validiteit de facto wat kleiner zal zijn.