

Rapport
2020:1

Bedre søknadsbehandling 2020

*Forskningen, analysene og argumentene som ligger til grunn
for hvordan Stiftelsen Dam organiserer sine tildelingsprosesser*

Jan-Ole Hesselberg

Rapport
2020:1

Bedre søknadsbehandling 2020

*Forskningen, analysene og argumentene som ligger til grunn
for hvordan Stiftelsen Dam organiserer sine tildelingsprosesser*

Jan-Ole Hesselberg

Rapport 2020:1

Utgitt av Stiftelsen Dam

Adresse: Akersgata 28, 0158 Oslo, Norge

Kontaktperson: Programsjef Jan-Ole Hesselberg, joh@dam.no

Bilde [Max Rovensky](#) på [Unsplash](#)

ISBN 978-82-691947-0-8

Copyright Stiftelsen Dam: [CC BY-NC 4.0](#)

www.dam.no

Forord

Stiftelsen Dam deler ut penger til prosjekter i regi av frivillige helseorganisasjoner i Norge. Vår jobb er å sikre at vi velger de beste prosjektene, slik at de rundt 300 millionene vi deler ut hvert år, kommer til best mulig nytte.

Derfor jobber vi systematisk med å kvalitetssikre de vurderingene som gjøres. Hos oss handler det om å lete grundig etter svakheter i våre tildelingsprosesser, holde oss oppdatert på hva forskningen sier om beslutningspsykologi og våre kjerneoppgaver, utvikle nye metoder og teste dem gjennom egen forskning.

I tråd med Stiftelsen Dams retningslinjer for åpenhet (se dam.no/retningslinjer-for-apenhet), publiserer vi resultatene i offentlig tilgjengelige regneark og rapporter. Denne rapporten er et ledd i dette.

Vi håper rapporten kan inspirere alle som jobber med søknadsbehandling til å reflektere over egne vurderingsprosesser og til å dele mer informasjon om dem.

Vi ønsker kritiske blikk på våre prosesser velkomne og ber leseren av rapporten melde alle innspill, store som små, via dam.no/bedre-soknadsbehandling.

Oslo, 18. juni 2020

Hans Christian Lillehagen
Generalsekretær
Stiftelsen Dam

Jan-Ole Hesselberg
Programsjef
Stiftelsen Dam

Innhold

1 Leseveiledning	10
1.1 Sentrale begreper	10
2 Hva vi bekymrer oss for	11
2.1 Ressursbruk	11
2.2 Uønsket påvirkning av søknadsbehandlingen	13
2.2.1 Interessekonflikter	13
2.2.2 Kognitive og sosiale bias	14
2.2.3 Støy	15
2.2.4 Hva vi har gjort for redusere uønsket påvirkning	17
3 Søknadsbehandlingen i våre programmer	20
3.1 Helse	21
3.2 Forskning	22
3.3 Utvikling	23
3.4 Ekspres	24
3.5 Kavlifondets program for helseforskning	24
4 Valgene vi har tatt og hvorfor	25
4.1 Fire felles vurderingskriterier i alle våre programmer	26
4.2 Sjudelt karakterskala	28
4.3 Beregnet total karakter på søknaden basert på kriterie karakterene	29
4.4 Faste ekspertgrupper, ingen eksterne eksperter	29
4.5 Bredt sammensatte fagutvalg	31
4.6 Flere uavhengige vurderinger på hver søknad	33
4.7 Prekvalifiseringsrunde	34
4.8 Siling av søknadene	38
4.9 Mindre grupper av fagpersoner, fremfor stort plenum	39
4.10 Ingen konsensus karakter og forbud mot karaktersnakk i gruppemøter	42
4.11 Administrativ fremfor faglig møteledelse	45
4.12 Brukerrepresentanter i søknadsvurderingen	45
4.13 Beregnet rangering etter gruppebehandling	47
4.14 Tilbakemelding til søker	51
5 Veien videre	53
6 Hva alle som driver søknadsbehandling bør gjøre	57
7 Referanser	60

1 Leseveiledning

Rapporten er delt i tre hoveddeler:

- “Hva vi bekymrer oss for” handler om de tingene vi mener truer god søknadsbehandling
- “Søknadsbehandlingen i våre programmer” beskriver de ulike programmene våre
- “Valgene vi har tatt og hvorfor” er hoveddelen. Den er organisert kronologisk, slik at den starter med de elementene som dukker opp tidlig i søknadsbehandlingen, men må ikke leses i den oppsatte rekkefølgen.

Målet er at rapporten kan fungere som et oppslagsverk for alle som interesserer seg for søknadsbehandling generelt og Stiftelsen Dams programmer spesielt.

1.1 Sentrale begreper

Fagutvalgsmedlem

Tilsvarende begreper: Fagfelle, reviewer, ekspert, saksordfører

De som vurderer våre søknader. Vi bruker fagutvalgsmedlem (FUM) og ikke “fagfelle”, “reviewer” eller “ekspert”, fordi disse begrepene er tett knyttet til vurderingen av forskningssøknader. De fleste av våre midler er ikke forskningsmidler og vurderes av fagutvalg bestående av personer med svært forskjellig bakgrunn og kompetanse. Tidligere kalte vi fagutvalgsmedlemmene for “saksordførere”.

Program

Tilsvarende begreper: Tilskuddsordning, ordning

Vi deler ut midlene gjennom “programmer”. Tidligere omtalte vi dette som “tilskuddsordning” eller bare “ordning”. Innenfor hvert program lyser vi ut midler.

Utlysning

Beskriver hvem som kan søke, hva de kan søke på og når de kan søke. Vi har fire programmer (Helse, Forskning, Utvikling og Ekspress) som beskrives grundigere under Søknadsbehandlingen i våre programmer.

2 Hva vi bekymrer oss for

Ved første øyekast fremstår antakelig ikke søknadsbehandling og pengeutdeling som en spesielt komplisert øvelse. Vurder søknader, gi penger til de beste, ferdig. Dessverre er det flere faktorer som kompliserer.

For det første må ressursene som brukes på utarbeidelsen av søknadene og vurderingen av dem veies opp mot nytten. For det andre må utdeleren (i dette tilfellet Stiftelsen Dam) sikre seg at det ikke uønskede elementer blir utslagsgivende i søknadsbehandlingen, enten det er interessekonflikter, bias eller støy.

2.1 Ressursbruk

Det finnes nesten ikke grenser for hvor omfattende man kan gjøre søknadsbehandling. Det er nesten alltid hensiktsmessig å bruke flere eksperter i vurderingen og det finnes heller nesten ingen grenser for hvor mye tid de potensielt kan legge i granskningen av detaljene i søknaden. Men kostnadene må stå i stil med nytten. Derfor jobber vi systematisk med å begrense unødig ressursbruk. Ikke bare med tanke på søknadsvurderingen, men også i søkernes arbeid med søknaden.

2.1.1 Kostnader forbundet med vurderingen

Som denne rapporten vil vise, er ønsket om å begrense unødvendig ressursbruk i forbindelse med vurderingene noe som ligger til grunn for mange av endringene vi har gjort de siste årene. Vi har for eksempel valgt å avvike store plenumsmøter der ekspertene diskuterer søknader, til fordel for mindre og kortere gruppemøter som ikke krever at vi leier lokaler og som gjør det enklere å gjennomføre digitale møter.

Vi har også valgt å tilpasse omfanget av vurderingsprosessene til søknadenes omfang. I stedet for å følge med på hva det koster å vurdere hver enkelt søknad, måler vi nå hva det koster å vurdere hver krone det søkes om.

I Tabell 1 ser du kostnadene forbundet med behandlingen av en enkelt søknad innenfor våre ulike programmer. Som du kan se er kostnadene forbundet med vårt enkleste program, Ekspress, svært lave per søknad (246 kroner) sammenliknet med Forskning som er vårt mest omfattende program (4.391 kroner per søknad).

Problemet med denne måten å måle kostnadene på, er at den ikke tar hensyn til søknadenes omfang. I Ekspress søkes det om ca 26.000 kroner i gjennomsnitt og maksimalt 30.000 kroner, mens det i Forskning søkes om ca 2,2 millioner i snitt. I tillegg er det store variasjoner i hvor stor andel av søknadene som innvilges. Fra 13 prosent i Forskning til omtrent 40 prosent i Ekspress. Måler vi kostnadene per 100.000 kroner som bevilges, justeres det både for søknadenes omfang og for innvilgelsesandel. Som tabellen viser, snus kostnadsbildet på hodet og vurderingene i Forskning blir faktisk rimeligere enn i Ekspress (1.375 mot 3.041 per 100.000 bevilgede kroner).

Tabell 1: Kostnader forbundet med fagtuvalgene som vurderer søknader

	Helse 2019	Forskning 2019	Ekspress 2019	Utvikling 2020*
Per søknad	2.228	4.391	246	1.413
Per 100.000 bevilgede kroner	1.575	1.375	3.041	886

*basert på antall innsendte skissesøknader (i prekvalifiseringsfasen)

2.1.2 Ressursbruk hos søker

Kostnadene forbundet med vurderingene av søknadene er enkle å måle, men utgjør kun en svært liten andel av de totale utgiftene. Den største andelen av ressursbruken knyttet til søknadsprosesser bæres av søkerne. De bruker en betydelig mengde tid på å skrive søknader som i mange tilfeller ikke blir innvilget.

Både våre egne undersøkelser og undersøkelser fra Norges forskningsråd viser at det går med svært mye tid til å søke midler til forskning. Som Tabell 2 viser, brukte våre søkere i gjennomsnitt 8,1 ukeverk per søknad. En undersøkelse gjennomført av NIFU for Norges forskningsråd viste at gjennomsnittet hos dem lå på 9,0 ukeverk (Ramberg 2017). Tidsbruken er vesentlig lavere i programmet Helse, men også her bruker søkerne en betydelig mengde tid.

Tabell 2: Søkeres selvrapporterte tidsbruk i ukeverk

Gjennomsnittlig tidsbruk (median i parentes)	Norges forskningsråd	Forskning 2019	Helse 2019
Ukeverk ved egen institusjon	6,8 (4,0)	6,6 (4,0)	2,5 (1,1)
Ukeverk ved eksterne partnerinstitusjoner	2,2 (1,0)	1,6 (0,6)	0,9 (0,3)
Totalt	9,0 (5,0)	8,1 (4,6)	3,4 (1,4)

Dersom vi multipliserer tallene fra tabellen over med antallet avslag, får vi et estimat på den totale tidsbruken forbundet med søknadene som ikke blir innvilget. For søknadsrundene i 2019 tilsvarer det 41,5 årsverk for programmet Forskning og 55,8 årsverk for programmet Helse. Det blir også mulig å beregne hvor mye tid som brukes per bevilgede krone. Det gir et sammenlignbart tall på tvers av våre programmer. Tabell 3 gir en oversikt over disse tallene.

Tabell 3: Antall avslag, total tidsbruk for avslag og dagsverk per 100.000 bevilgede kroner

	Helse 2019	Forskning 2019	Ekspress 2019	Utvikling 2020*
Antall avslag	755	234	1.498	119*
Årsverk totalt for avslag	55,8	41,5	2,6	7,8*
Dagsverk per 100.000 bevilgede kroner	11,9	12,7	7,9	12,1

*) Utvikling 2020 benyttet en totrinnsprosess med en prekvalifiseringsfase. Tallene er for begge trinnene totalt sett.

En undersøkelse fra National Health and Medical Research Council i Australia forsøkte å estimere den totale ressursbruken forbundet med søknadsprosesser, inkludert administrative kostnader, utgifter til søknadsvurdering, ressurs- og tidsbruk i forskningsinstitusjonene som søker og så videre.

Undersøkelsen konkluderte med at forskningsorganisasjonene stod for 85 prosent av ressursbruken, nettopp fordi søknadsskrivingen tar så mye tid (Graves et al. 2011).

Det er viktig å understreke at denne ressursbruken ikke kan beskrives som “bortkastet”. I arbeidet med søknaden bygger søkerne egen kunnskap, både om temaet prosjektet omhandler og om prosjektplanlegging og søknadsskriving. I tillegg bygges det nettverk med samarbeidspartnere. Videre kan hele eller deler av søknaden gjenbrukes i andre søknadsprosesser.

Da vi spurte våre søkere i Forskning 2020 “Hvilken betydning/nytte har arbeidet med den aktuelle søknaden for deg hvis søknaden ikke innvilges?”, svarte hele 91 prosent “middels”, “stor” eller “svært stor”. I den nevnte undersøkelsen hos Norges forskningsråd svarte over to tredjedeler det samme.

Vår holdning er likevel at det er en del av vårt ansvar å sørge for at søknadsprosessen belaster søkerne og dermed det norske samfunnet minst mulig. Flere av tiltakene som beskrives i denne rapporten har hatt som formål å begrense denne belastningen. Innføringen av skissesøknader (se [Prekvalifiseringsrunde](#)) og nye vurderingskriterier (se [Fire felles vurderingskriterier](#)) er blant noen av disse tiltakene.

2.2 Uønsket påvirkning av søknadsbehandlingen

Stiftelsen Dam har, som de fleste andre finansører, et sett med styringsdokumenter som beskriver hvordan våre midler skal deles ut. Utlysningene for de ulike programmene vedtas av vårt styre og der beskrives det hvem som kan søke på hva, og hvilke områder som skal prioriteres. Å sikre at våre tildelinger gjenspeiler kriteriene og prioriteringene i utlysningene, er sekretariatets viktigste oppgave.

Det er flere potensielle kilder til uønsket påvirkning av søknadsbehandlingen. Grensene mellom dem er ikke alltid tydelige, men det kan likevel være nyttig å dele dem inn i tre hovedkategorier:

- Interessekonflikter
- Kognitive og sosiale bias
- Støy

Selv om rapporten i det følgende fokuserer på vurderingene til fagutvalgsmedlemmene, utgjør disse tre kilder til uønsket påvirkning hos alle som er involverte i søknadsbehandlingen, inkludert sekretariatet og styret.

2.2.1 Interessekonflikter

Den norske forskeren Liv Langfeldt har kartlagt hva som påvirker faglige vurderinger av forskningssøknader (Langfeldt 2002) og skriver at *Både juridisk habilitet og andre interesser og forhold som kan medføre partiskhet eller skjevheter i tildelingsprosessene, er et sentralt spørsmål i behandling av søknader om forskningsmidler.*

En samtale vår programsjef Jan-Ole Hesselberg hadde med en merittert professor kan tjene som eksempel på slike interessekonflikter. Professoren hadde jobbet flere år med å vurdere søknader for Norges forskningsråd og fortalte om sin strategi i møtene med de andre ekspertene:

Professoren: Min hemmelige misjon var å få inn flere kvinner (...) ved å få de andre med på at mine vurderinger var til å stole på. I de søknadene jeg hadde sett meg ut som jeg skulle prøve å få gjennom, så måtte jeg forte meg å ta ordet. Jeg måtte komme inn med oppsnakkingen før de andre. Jeg kunne ikke tåle mer enn én nedsnakking. Da var det for sent.

Jan-Ole: Problemet for finansjøren, sånn som oss, er at vi har klare retningslinjer og vedtekter for hva vi ønsker å bruke penger på, så hvis man får inn eksperter som har en annen agenda, er det et problem.

Professoren: Men jeg var opptatt av å motvirke en skjevhet.

Paradokset bør være åpenbart. Hun ønsket å rette opp en *antatt* skjevfordeling ved aktivt å skjevfordele selv. Dersom alle gjør det, mister både søkerne og vi som finansjør fullstendig oversikt og utvelgelsesprosessen blir uforutsigbar. Hva om en i gruppen mener forskning på barn og unge bør prioriteres, mens en annen mener fagmiljøer i Nord-Norge bør løftes fram og en tredje mener kvalitativ forskning kommer dårlig ut? Dersom slike områder skal prioriteres må det gjøres på et overordnet plan og inkluderes i utlysningen av midlene, ikke innføres på bakrommet av eksperter med en egen agenda.

Det vil ikke være mulig å unngå slike interessekonflikter fullt og helt, men som rapporten vil vise, har vi gjennomført flere tiltak som delvis er myntet på å begrense utslaget av dem. Både innføringen av flere uavhengige vurderinger, forbud mot karaktersnakk i gruppemøtene og administrativ ledelse av gruppemøtene er blant disse tiltakene.

2.2.2 Kognitive og sosiale bias

Det er viktig å understreke at uønskede skjevheter i vurderingene ikke trenger å være et resultat av bevisste handlinger eller åpenbare interessekonflikter. I samtalen over, har eksperten en tydelig agenda og har bevisst utviklet en strategi for å nå fram med den, men det vanligste er antakelig at utfallet påvirkes på mer subtile måter, gjennom kognitive og sosiale bias¹.

En slik subtil kilde til påvirkning er eksperten inne på selv: **Hvem som snakker først.**

Forskning viser også at dette kan være en utfordring i slike søknadsprosesser. I en gjennomgang av forskningen på sosiale prosesser i grupper som vurderer forskningssøknader, konkluderer forskerne med at “*What the first speaker starts the discussion with is found to be decisive [...]. The strong or weak points of the applicant mentioned first are strongly supported by other panelists*” (van Arensbergen et al. 2014).

¹ Bias oversettes gjerne med “skjevhet”. I dette tilfellet sikter eksperten til at hun mener at menn har fått en for stor andel av innvilgelsene – en systematisk skjevhet i favør av mannlige søkere. [Store norske leksikon](#) beskriver bias som “utsagn, valg og vurderinger som på systematisk måte avviker fra det som stemmer overens med virkeligheten”

Både sosiale og kognitive prosesser bidrar til “første taler”-effekten og en av disse er den såkalte **forankringseffekten** (“anchoring” på engelsk). Når vi gjør vurderinger der svaret ikke er åpenbart, bruker vi ofte et referansepunkt. Tenk for eksempel på hvordan du besvarer dette spørsmålet: *Hvor mange mennesker bor det i Hamburg?*

Dersom du ikke allerede vet svaret, er sannsynligheten stor for at du bruker en by du kjenner innbyggertallet på som referansepunkt og deretter justerer ut fra det. Kanskje tar du utgangspunkt i Oslos rundt 670.000 innbyggere? Og kanskje tenker du at Hamburg er 2-3 ganger større? Og det fører kanskje til at du tipper 1,5 millioner?

Strategien fungerer ofte godt når vi skal gjøre overslag (det bor 1,8 millioner i Hamburg), men problemet er at forankringspunktet ofte velges for oss og i tillegg kan være ganske tilfeldig valgt. Som for eksempel når man diskuterer karactersetting i en gruppe med eksperter som vurderer søknader. Mye tyder på at den første karakteren som tilfeldigvis nevnes vil fungere som et forankringspunkt for de andre deltakerne og påvirke deres karactersetting. Forankringseffekten er godt dokumentert i mange sammenhenger (Gilovich et al. 2002) og har blant annet vist seg å påvirke straffen dommere gir (Englich et al. 2006) og budgivning under auksjoner (Ariely et al. 2003).

Både “første taler”-effekten og forankringseffekten er to helt naturlige, menneskelige tendenser som gjør oss i stand til å ta raske beslutninger som påfallende ofte er relativt nøyaktige. Men i situasjoner hvor man tar viktige beslutninger der man har klare ønsker for hva som skal og ikke skal vektlegges, kan disse tendensene bidra til systematiske skjevheter. Derfor forsøker vi å lage prosesser som demmer opp for dem. Vi har klare regler for hvilken rekkefølge søknadene skal vurderes i, og i gruppemøtene i programmet Forskning har vi nedlagt forbud mot å snakke om karakterer (se Forbud mot karactersnakk i gruppemøter).

2.2.3 Støy

For ikke lenge siden skjedde det en liten feil i datasystemet vårt som gjorde noe med fordelingen av søknadene til en av våre eksperter. Han visste det ikke selv, men i rekken av de rundt 70 søknadene han fikk til vurdering dukket en søknad opp to ganger.

Første gang han fikk søknaden pløyde han gjennom de ti sidene med prosjektbeskrivelse og landet på at det var en god søknad. Karakter 4 på en skala fra en til syv. To uker senere ga han seg i kast med nøyaktig samme søknad på nytt og merket det ikke. Hva skjedde? Søknaden var den samme, men det var ikke ekspertens tolkning av den. Han ga den bunnkarakteren 1.

Det er dessverre ikke en unik hendelse og vi opplever både at ekspertene er uenige med seg selv og med andre eksperter som vurderer samme søknad. Den israelsk-amerikanske psykologen og nobelprisvinneren Daniel Kahneman kaller dette “støy” (eng. “noise”) og skiller det fra bias (Kahneman et al. 2016). Begge deler er kilder uønsket påvirkning av menneskelige vurderingsprosesser, men bias er *systematiske og forutsigbare* skjevheter. Det er ikke støy. Kahneman bruker baderomsvekten for å beskrive forskjellen. Den skjev måler (er “biased”) dersom den systematisk viser et par kilo over eller under din egentlige vekt og den er støyete (“noisy”) hvis den viser ulik vekt avhengig av hvor du tilfeldigvis plasserer føttene dine på vekten. I det første

tilfellet får du omtrent samme vekt hver gang du veier deg, i det andre tilfellet får du ulike målinger hver gang.

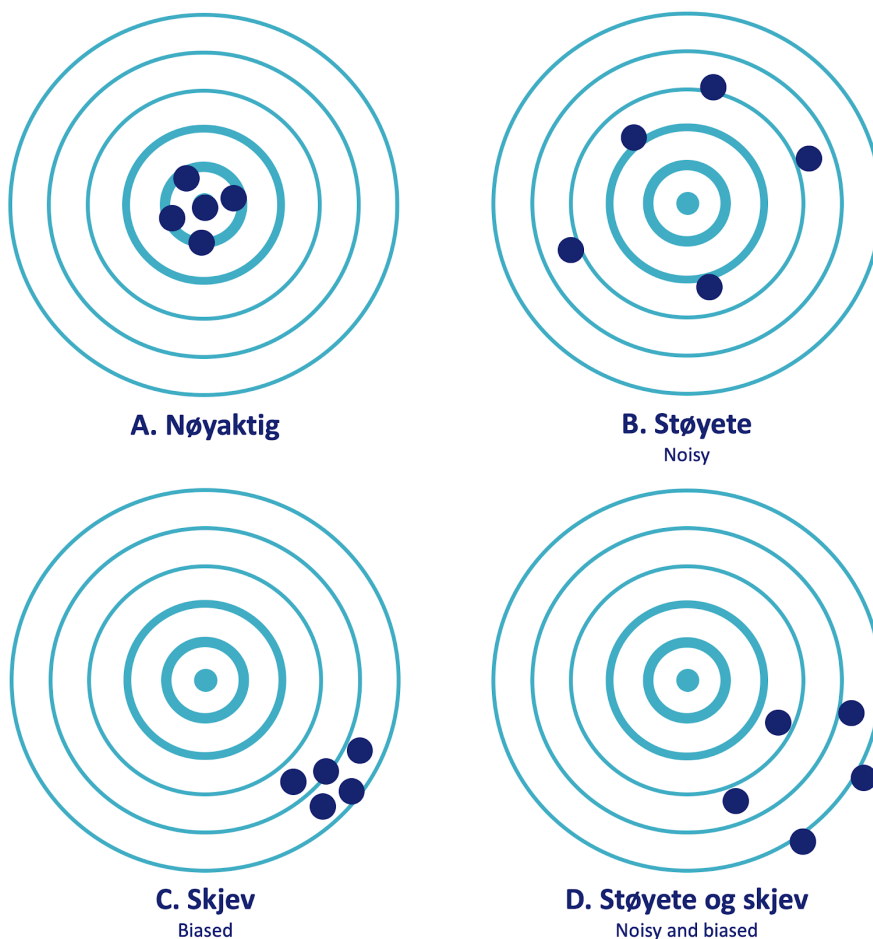
Forskjellen mellom støy og bias kan også presenteres visuelt. I Figur 1 ser du hvordan støy og bias påvirker nøyaktigheten til fire ulike grupper av eksperter som alle vurderer samme søknad. Midten av blinken representerer søknadens "faktiske kvalitet". Det er mange måter å definere "faktisk kvalitet" på, men det er ikke vesentlig i dette tilfellet.

I gruppe A er vurderingene nøyaktige. De ulike ekspertene er enige i sine vurderinger og vurderingene sammenfaller med den faktiske kvaliteten.

I gruppe B er vurderingene støyete. Ekspertene er ikke enige og vurderingene deres er heller ikke treffsikre.

I gruppe C er vurderingene skjeve ("biased"). Ekspertene er enige seg imellom, men treffer ikke med vurderingene sine.

I gruppe D er vurderingene både støyete og skjeve. Ekspertene er verken enige eller treffsikre og i tillegg tenderer vurderingene i en bestemt retning.



Figur 1: Hvordan støy og bias påvirker nøyaktighet

I vårt tilfelle er støy det som på en *uforutsigbar* eller *ukontrollerbar* måte gjør at 1) samme ekspert gir like søknader forskjellige karakterer eller 2) ulike eksperter gir samme søknad forskjellige karakterer.

Forskning innen beslutningspsykologi har vist at veldig mange faktorer kan bidra til støy i menneskelige vurderinger:

- *Hvilke søknader som tilfeldigvis nettopp er vurdert danner referanserammen* for vurderingen av de påfølgende søknadene. Dette har blant annet vist seg å påvirke utfallet i vurderingen av både lånesøknader og asylsøknader (Chen et al. 2016).
- *Hvor sliten du er.* Ikke bare om du er fysisk sliten, men også hvor mentalt sliten du er. Flere studier har antydnet at man blir sliten av å ta samme type beslutninger og at det påvirker konklusjonene man lander på. En kjent studie fra 2011 viste for eksempel at dommere som vurderte innsatte for prøveløslatelse, ble markant strengere utover dagen, helt til de tok pauser (Danziger et al. 2011). Fenomenet faller inn under det som omtales som “choice overload” (Chernev et al. 2015)
- *Humøret ditt.* Humøret vårt kan påvirke vurderingene våre. (Johnson and Tversky 1983; Slovic and Peters 2006)
- *Om du har dårlig tid.* Flere studier viser at tidspress påvirker måten vi tar beslutninger på (Dhar and Nowlis 1999; Rieskamp and Hoffrage 1999)
- *Været.* Studier har blant annet vist at investeringer i aksjemarkedet muligens påvirkes av været (Kramer and Weber 2012; Saunders 1993)
- *I hvilken grad du må på do.* Ja, til og med dette er det forsket på. (Tuk et al. 2010)

Det er ikke rimelig å anta at alle disse faktorene påvirker våre beslutningsprosesser i enhver vurderingsprosess. Poenget med listen er i stedet å vise hvor mange ting som *potensielt* kan påvirke våre beslutninger og som *kan* bidra til den store variasjonen vi ser i vurderingene til våre eksperter. Disse potensielle kildene til støy begrenser seg uheldigvis ikke til disse eksemplene og sammen utgjør de en betydelig utfordring for alle som tar viktige beslutninger basert på menneskelige vurderinger.

Det er grunnen til at vi systematisk kartlegger innslaget av støy og gjennomfører det Kahneman kaller “noise audits” eller “støyrevisjon”, ved å måle enigheten (inter-rater reliabiliteten) mellom våre eksperter. Er enigheten veldig lav, betyr det normalt høyt innslag av støy. Det er godt dokumentert at enigheten i søknadsprosesser som våre, er problematisk lav (Marsh et al. 2008; Marsh et al. 2007; Guthrie et al. 2017) og våre programmer er heller intet unntak. Selv om problemet gjelder mange, er det like fullt et problem og noe vi forsøker å gjøre noe med (se [Veien videre](#)).

2.2.4 Hva vi har gjort for redusere uønsket påvirkning

Vi har tatt flere grep for å redusere både støy, bias og interessekonflikter.

For det første har vi sørget for å sikre innsamling av informasjon om søknadene vi får og vurderingene til ekspertene våre. Det gjør oss i stand til å gjennomføre både støyrevisjon (se over) og å analysere tildelingene våre for biaser. Spiller søkerens kjønn en rolle for om man får innvilgelse eller ikke? Får store, mektige aktører lettere innvilgelse enn de små? Og påvirker søknadssummen sjansene for innvilgelse? Dette er blant noen av tingene vi holder øye med og som vi også deler med

omverdenen (se blant annet innlegget [Dette har vi lært av å behandle 3.667 søknader](#) i Khrono 24.09.19).

For det andre standardiserer vi elementer i søknadsprosessen. Innføringen av maler for prosjektbeskrivelse og CV gjør for eksempel at ekspertenes preferanser for fonter, formatering og disposisjon ikke lenger kan få betydning. Vi har også standardisert rekkefølgen søknadene skal behandles i, slik at vi i størst mulig grad unngår uheldige rekkefølgeeffekter.

For det tredje er vi i gang med å utvikle opplæringsprogrammer for fagutvalgsmedlemmene våre, slik at de får en felles forståelse av hvordan kriteriene skal tolkes og karakterskalaen skal brukes. Målet er at det skal redusere både støy og bias.

For det fjerde unngår vi å bruke elementer i vurderingsprosessen som er unødvendige eller som reduserer kvaliteten på beslutningsgrunnlaget. Siden vurderingene er sårbare for støy, bias og interessekonflikter, bør det ikke innføres elementer der det ikke er behov for dem. Et eksempel på dette er at vi nå beregner en totalkarakter for hver søknad basert på karakterene på de ulike vurderingskriteriene fremfor å la fagutvalgsmedlemmene sette en totalkarakter (se [Beregnet totalkarakter](#)). Flere studier har for øvrig vist at det å eliminere menneskelige vurderinger i beslutningsprosesser i langt mer komplekse prosesser kan føre til bedre beslutninger (Kahneman et al. 2016)

Sist, men ikke minst, bruker vi flere eksperter for å vurdere hver enkelt søknad. Flere vurderinger betyr at vurderingen til den enkelte ekspert får mindre å si. Det betyr også at støy begrenses og at interessekonflikter som vi ikke klarer å plukke opp ikke får like stor betydning (se [Flere uavhengige vurderinger](#) for mer om dette).

Selv om vi har gjort mye for å sikre kvaliteten på våre vurderingsprosesser er vi åpenbart ikke i mål. Vi vet at vi opplæringen av ekspertene våre kan bli bedre, at vi i større grad kan anonymisere informasjon i søknadene som et ledd i å fjerne mulige kilder til uønsket påvirkning (som for eksempel søkerens navn, kjønn og alder) og at vi kan bli bedre på å kartlegge kilder til støy i situasjonene der ekspertene våre vurderer søknader individuelt. Både analysene og endringene tar imidlertid tid og krever noen ganger store investeringer (for eksempel i utviklingen av datasystemene vi bruker). Vi håper imidlertid at denne rapporten vil vise at vi er på god vei og at vi ikke mener vi er i mål.

Tabell 4: Utvalgte, potensielle kilder til uønsket påvirkning av søknadsbehandlingen

Område*	Potensielt bekymringsverdige tendenser
Faglige og personlige interesser og holdninger	<p>Kjønn. Bevisst og ubevisst vektning av kjønn i behandlingen av søknadene, i tilfeller der det ikke skal vektlegges (Bornmann et al. 2007; Forscher et al. 2019; Witteman et al. 2019; Helmer et al. 2017).</p> <p>Nettverk. Når bekjentskaper (også svært fjerne bekjentskaper) får bedre vurderinger (Teplitskiy et al. 2018).</p> <p>Politisk ideologi. Når egne politiske overbevisninger bevisst eller ubevisst påvirker vurderingen av søknadene (Rom and Musgrave 2014).</p> <p>Livssyn. At eget livssyn påvirker vurderingen av søknader som omhandler livssynssensitive tema, som for eksempel forskning på abort eller bruk av nikab (Smith and Blazeby 2018; Mrdjénovich 2020).</p> <p>Etnisitet. Bevisst eller ubevisst vektlegging av etnisitet uten at dette er begrunnet i vurderingskriteriene (Busetta et al. 2013; Midtbøen and Rogstad 2012).</p> <p>Metode. Når spesielle metodiske tilnærminger løftes bevisst eller ubevisst, uten at dette er begrunnet i kvaliteten på prosjektet.</p> <p>Geografi. Når prosjektets geografiske tilknytning vektlegges bevisst eller ubevisst, i tilfeller der det ikke er en del av kriteriene.</p>
Kognitive og sosiale bias	<p>“Første taler”-effekten. At den første som sier noe får uforholdsmessig mye å si for utfallet av søknadsbehandlingen (van Arensbergen et al. 2014). Både sosiale og kognitive prosesser bidrar til denne effekten. “Forankringseffekten” er en av flere ting som bidrar. Det er tendensen til å legge for stor vekt på den første informasjonen man mottar før man skal ta en beslutning (Tversky and Kahneman 1974; Wilson et al. 1996)</p> <p>Tilgjengelighetseffekten. Lett tilgjengelig informasjon, tillegges større vekt enn mindre tilgjengelig informasjon. (Gilovich et al. 2002)</p> <p>Bekreftelsesskjevhet (“confirmation bias”). Tendens til å bekrefte våre egne antakelser.</p> <p>Likhetseffekten (“homophily”). Tendens til å foretrekke personer som likner oss selv (Brezis and Birukou 2019; Helmer et al. 2017).</p> <p>Eksposeringseffekten (“mere exposure effect”). Tendens til at man liker noe bedre bare fordi man har blitt eksponert for det ofte (Harrison 1977).</p> <p>Gruppetenkning. At vurderinger og avgjørelser i grupper styres av et ønske om harmoni og enighet innad gruppen heller enn en grundig vurdering av søknaden (van Arensbergen et al. 2014; Asch 1956). En nært beslektet effekt er “the social influence effect” (Lorenz et al. 2011).</p> <p>Autoritet. Når autoriteter blir tillagt en spesiell vekt bare i kraft av at de er autoriteter og ikke på grunn av kvaliteten deres argumenter (Teplitskiy et al. 2019).</p>
Kilder til støy	<p>Rekkefølgeeffekter. Hvilken rekkefølge søknadene behandles i kan spille en rolle for utfallet (Chen et al. 2016).</p> <p>Beslutningstretthet. En serie av mange eller kompliserte vurderinger og beslutninger kan føre til endringer i måten vi gjør vurderingene på (Chernev et al. 2015)</p> <p>Tidspress. Hvor mye tid vi har til rådighet når vi skal gjøre vurderinger, påvirker vurderingene (Dhar and Nowlis 1999; Rieskamp and Hoffrage 1999)</p> <p>Humør. Humøret vårt kan påvirke vurderingene våre. (Johnson and Tversky 1983; Slovic and Peters 2006)</p>

* Skillet mellom disse områdene er ikke alltid klart. Systematiske og forutsigbare tendenser (bias) hos den enkelte, kan bli til uforutsigbare og ukontrollerbare årsaker til at fagutvalgsmedlemmene vurderer søknadene ulikt (støy).

3 Søknadsbehandlingen i våre programmer

Stiftelsen Dam deler ut rundt 300 millioner kroner årlig. Alle midlene går til tidsavgrensede prosjekter rettet mot målgrupper i Norge, og som bidrar til å fremme levekår, fysisk og psykisk helse, mestring, livskvalitet eller sosial deltakelse. Midlene deles ut gjennom søknadsprosesser der forhåndsgodkjente søkerorganisasjoner må stå som søker.

Tabell 5 viser de fire ulike programmene man kan søke midler gjennom.

Tabell 5: Stiftelsen Dams programmer

Program	Årlig sum	Utlysninger per år	Uavhengige vurderinger per søknad	Antall søknader per utlysning	Søknadssum (snitt)	Innvilgelsesandel
<u>Helse</u>	ca 180 mill	2	3 (5**)	ca 550	ca 750.000	ca 20 %
<u>Forskning*</u>	ca 80 mill	1	5	ca 300	ca 2.000.000	ca 13 %
<u>Utvikling*</u>	ca 20 mill	1	4	ca 100	ca 1.500.000	ca 9 %
<u>Ekspress</u>	ca 12 mill	2	2	ca 800	ca 25.000	ca 30 %

* Programmer med en prekvalifiseringsrunde. Innvilgelsesandelen baserer seg på innvilgelser relativt til antallet søknader i prekvalifiseringsfasen og blir derfor spesielt lav.

** Rundt 25 prosent av søknadene behandles i to runder. Disse får minst fem uavhengige vurderinger.

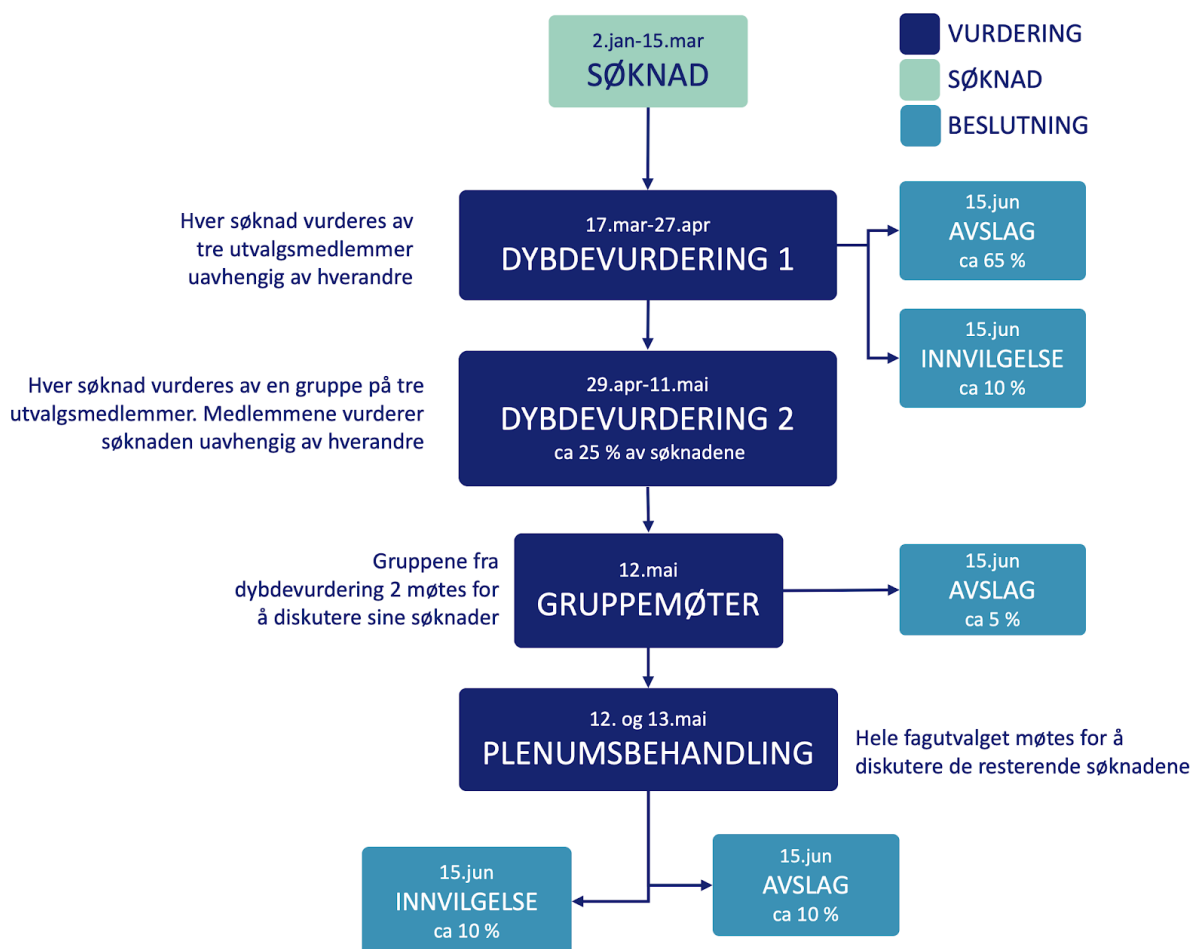
Som tabellen viser retter programmene seg mot ulike typer prosjekter av ulike størrelser. Dette stiller også ulike krav til grundigheten i søknadsbehandlingen. I det følgende beskrives hovedtrekkene i søknadsbehandlingen i de ulike programmene.

3.1 Helse

I programmet Helse foregår søknadsbehandlingen i to runder. I **runde 1** fordeles søknadene til tre utvalgsmedlemmer. Søknadens tema og kompetanseprofilen til fagutvalgsmedlemmene er utslagsgivende for hvem som får søknaden til vurdering. Utvalgsmedlemmene vurderer søknaden uavhengig av hverandre og det skjer elektronisk, uten at de møtes.

Basert på disse vurderingene avslås cirka 65 prosent av de svakeste søknadene, cirka ti prosent av de sterkeste innvilges og de resterende 25 prosentene går videre til fornyet vurdering i runde to.

I **runde 2** fordeles søknadene til grupper på tre fagutvalgsmedlemmer, der vanligvis en av fagutvalgsmedlemmene har vurdert søknaden i runde 1. Fagutvalgsmedlemmene vurderer søknadene på nytt, uavhengig av hverandre og elektronisk. Etter å ha gjort dette møtes gruppene for å diskutere søknadene og for å sette en felles karakter på søknaden. Basert på dette velger hver gruppe ut noen søknader som går til avslag. Resten av søknadene behandles av hele fagutvalget i plenum.



Figur 2: Søknadsflyt, programmet Helse (datoer for Helse vår 2020-utlysningen)

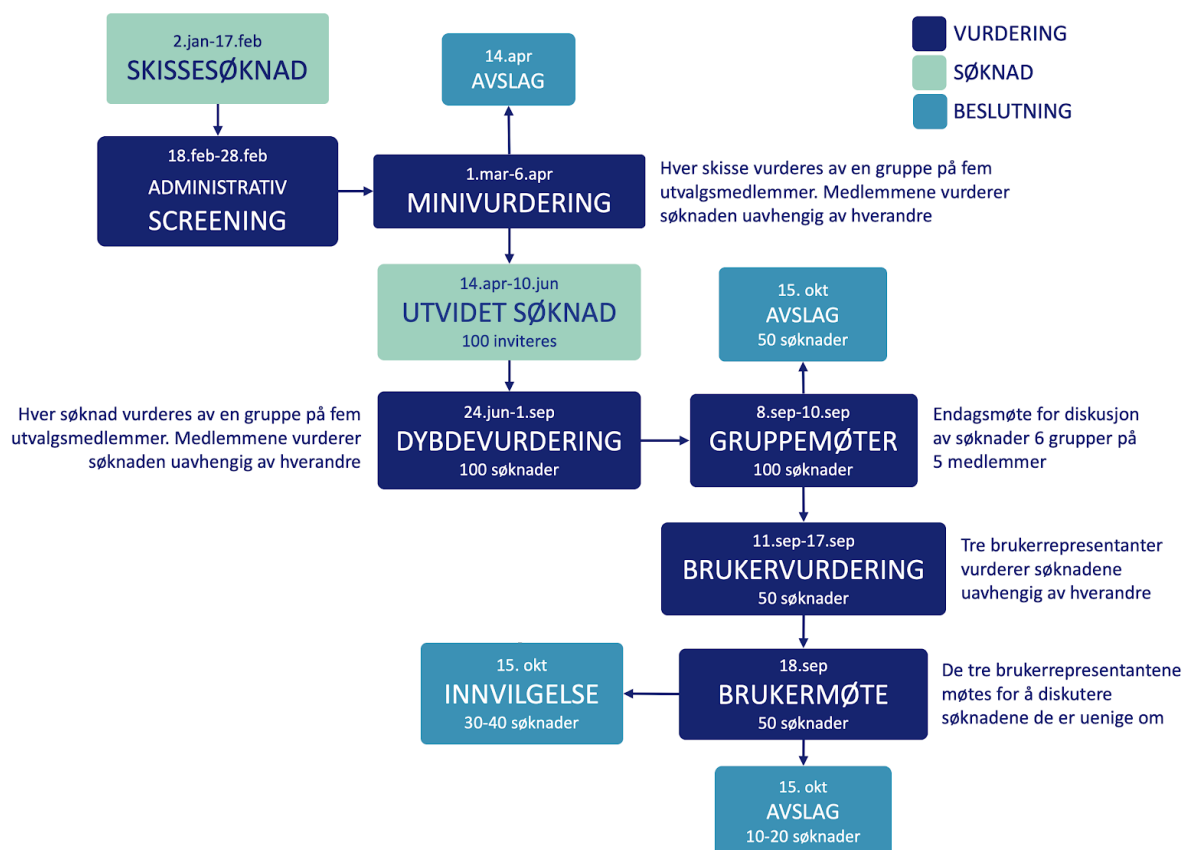
Søknadsbehandlingen for programmet Helse beskrives i detalj i rutinebeskrivelsen Søknadsvurdering, Helse og i fagutvalgets instruks.

3.2 Forskning

Programmet Forskning ble lagt om før utlysningen i 2020 og nå skjer søknadsbehandlingen i to runder: Prekvalifisering og utvidet søknad. I **prekvalifiseringsrunden** sender søker en forenklet projektskisse. Denne skissen fordeles til grupper på fem fagutvalgsmedlemmer, basert på skissens tema og gruppenes kompetanseprofil. De fem medlemmene vurderer søknaden uavhengig av hverandre og uten at de møtes. Basert på disse vurderingene inviteres omtrent en fjerdedel til å sende utvidede søknader. Resten avslås.

De utvidede søknadene vurderes av samme gruppe som vurderte skissen. Først vurderer gruppemedlemmene søknaden helt uavhengig av hverandre, så møtes gruppen for å diskutere søknadens styrker og svakheter. Basert på diskusjonen skal fagutvalgsmedlemmene justere sine individuelle karakterer. Gruppen skal altså ikke sette en felles karakter, slik det gjøres i programmet Helse. Søknadene som er aktuelle for innvilgelse går til vurdering hos tre brukerrepresentanter.

Brukerrepresentantene vurderer om brukermedvirkningen er ivaretatt. I tilfeller der søknadene står likt, er det karakteren fra brukerrepresentantene som avgjør.



Figur 3: Søknadsflyt i programmet Forskning (datoer for 2020-utlysningen)

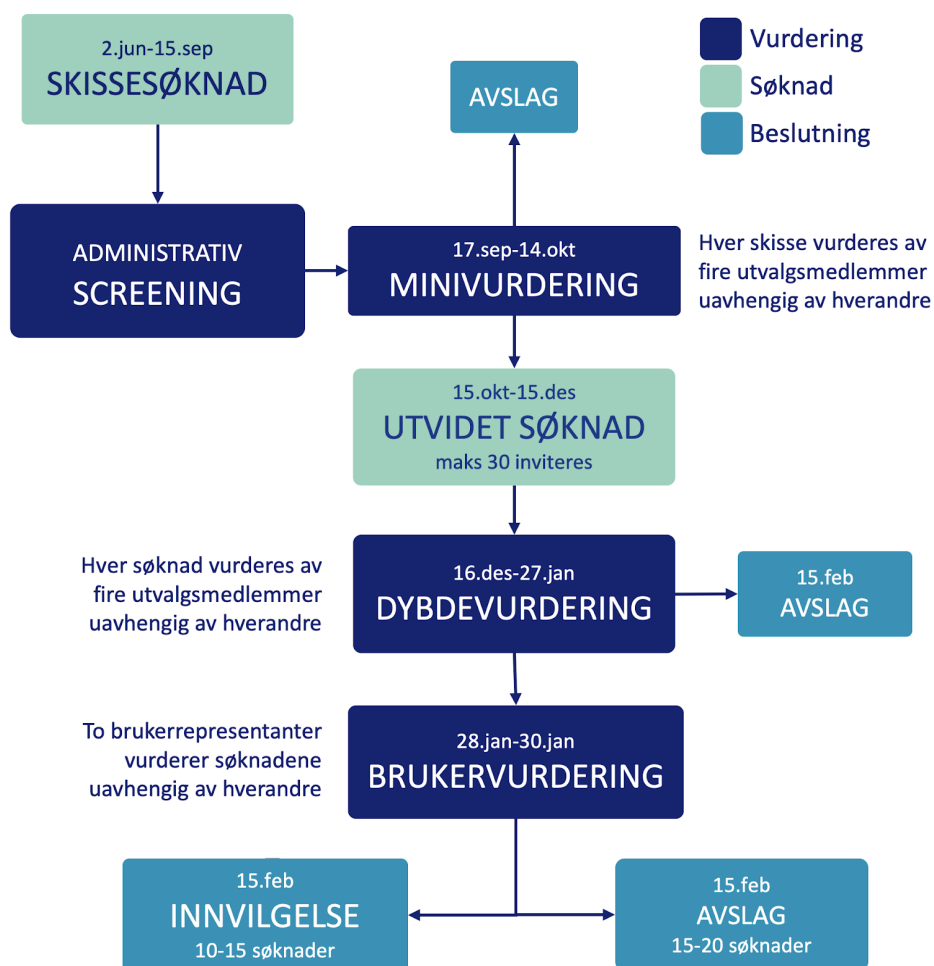
Søknadsbehandlingen for programmet Forskning beskrives i detalj i rutinebeskrivelsen [Søknadsvurdering, Forskning](#) og i [fagutvalgets instruks](#).

3.3 Utvikling

I pilotprogrammet Utvikling skjer søknadsbehandlingen i to runder: Prekvalifisering og utvidet søknad. I **prekvalifiseringsrunden** sender søker en forenklet projektskisse. Denne skissen fordeles til fire utvalgsmedlemmer som vurderer søknaden uavhengig av hverandre og uten at de møtes. Basert på disse vurderingene inviteres omtrent 25 prosent til å sende utvidede søknader. Resten avslås.

De utvidede søknadene vurderes av fire fagutvalgsmedlemmer (ikke nødvendigvis av de samme som vurderte skissen) uavhengig av hverandre. Deretter går alle søknadene videre til vurdering hos to **brukerrepresentanter** som også vurderer søknadene uavhengig av hverandre.

Etter dette rangeres alle søknadene etter snittkarakteren fra fagutvalget (avrundet til nærmeste halve karakter). I tilfeller der søknadene står likt, er det karakteren fra brukerrepresentantene som avgjør.



Figur 4: Søknadsflyt i programmet Utvikling (datoer for 2020-utlysningen)

Søknadsbehandlingen for programmet Utvikling beskrives i detalj i rutinebeskrivelsen Søknadsvurdering, Utvikling og i fagutvalgets instruks.

3.4 Ekspress

I programmet Ekspress er søknadsbehandlingen svært enkel. Her kan det kun søkes om inntil 30.000 kroner og både utfyllingen av søknadsskjemaet og vurderingen skal gå raskt.

Søknadene vurderes av to fagutvalgsmedlemmer uavhengig av hverandre og det er gjennomsnittskarakteren som avgjør om søknaden innvilges eller ikke.



Figur 5: Søknadsflyt i programmet Ekspress

Søknadsbehandlingen for programmet Ekspress beskrives i detalj i rutinebeskrivelsen [Søknadsvurdering, Ekspress](#) og i [fagutvalgets instruks](#).

3.5 Kavlifondets program for helseforskning

Stiftelsen Dam har siden 2017 også stått for utviklingen og driften av [Kavlifondets program for helseforskning](#), som deler ut 20-25 millioner kroner årlig til forskning på barn og unges psykiske helse. Brukere (pasienter, pårørende og helsepersonell) har en sentral rolle i prioriteringen av hvilke forskningsspørsmål som skal finansieres. I Aftenposten-innlegget ["Hvordan unngå bortkastet helseforskning?"](#) skriver vi om bakgrunnen for programmet (Hesselberg et al. 2017)

Også i dette programmet benytter vi en prekvalifiseringsfase og det er fire faste fagutvalgsmedlemmer som vurderer søknadene.

4 Valgene vi har tatt og hvorfor

De siste årene har Stiftelsen Dam gjennomført store endringer i utformingen av de ulike programmene. Sentrale elementer i søknadsbehandlingen har blitt justert, trukket fra eller blitt lagt til. Vi får ofte spørsmål om hvorfor vi gjør som vi gjør og det er ikke alltid enkelt å gi gode oppsummeringer av flere måneders tankevirksomhet og analyser. Ønsket om å samle forklaringen bak våre viktigste tiltak har vært noe av motivasjonen bak denne rapporten.

Når vi gjør endringer i programmene er det vanligvis, som beskrevet over, for å redusere ressursbruken eller uønsket påvirkning av søknadsbehandlingen. Vi forsøker å jobbe systematisk når vi skal gjennomføre endringer. Gjennom administrasjonen våre egne programmer og Kavlifondets program for helseforskning, har vi bygget et rikt erfaringsgrunnlag. Over tid har vi sørget for å sikre datainnsamling fra de ulike delene av søknadsbehandlingen og har i dag et rikt datagrunnlag (se blant annet dam.no/statistikk) som vi ofte bruker for å veilede beslutningene våre.

Vi ser også ofte til andre finansiørers arbeid og erfaringer. Det finnes et bredt spekter av nasjonale og internasjonale finansiører som fordeler midler på lignende måter som oss og som har testet mange ulike tiltak for å forbedre prosessene. Vi har blant annet hentet mye kunnskap og inspirasjon fra både Norges forskningsråd, EUs Horizon 2020, danske Villum-fonden, tyske VolkswagenStiftung, britiske Wellcome Trust og amerikanske Open Society Foundations.

Vi gjennomfører også regelmessige undersøkelser blant våre interessenter. Gjennom dem gir søkerne, fagutvalgsmedlemmene og søkerorganisasjonene oss viktig tilbakemelding på hvordan de opplever programmene og hvilke endringer vi bør vurdere. Programmene til stiftelsen er avhengige av å være relevante for søkerorganisasjonene og fagmiljøene. Derfor forsøker vi å ha et bredt tilbud og programmer som treffer organisasjonenes behov for kunnskapsutvikling til enhver tid. Blant annet kom pilotprogrammet Utvikling i stand fordi organisasjonene meldte om et behov for forsknings- og utviklingsprosjekter som ikke ble dekket i våre andre programmer.

Sist, men ikke minst gjennomfører vi også egen forskning. Når vi iverksetter tiltak, forsøker vi å gjøre det på en måte som gjør at vi kan få svar på hvilken effekt de har. Da vi vurderte å innføre standardiserte, skriftlige tilbakemeldinger til fagutvalgsmedlemmene våre inviterte vi dem til å delta i en studie der halvparten fikk tilbakemeldingen og den andre halvparten fikk en kontrollintervensjon (se dam.no/feedback for mer informasjon). Studien ble valgt ut til å bli presentert under PEERE International Conference On Peer Review 2020.

Vi gjennomfører også en større studie der vi skal kartlegge hvilke faktorer som påvirker graden av enighet mellom fagutvalgsmedlemmer som vurderer samme søknad. Både Kreftforeningen, Kavlifondet og Norges forskningsråd har bidratt med data til denne studien.

Dette er eksempler på arbeidsmåtene som ligger til grunn for tiltakene som beskrives under og vi håper vi får fram at vi tar dette arbeidet svært seriøst.

4.1 Fire felles vurderingskriterier i alle våre programmer

Vi bruker et felles sett med vurderingskriterier i alle våre programmer. De fire vurderingskriteriene er:

- **Soliditet** – Kvaliteten på de foreslåtte aktivitetene og metodene. I hvilken grad det foreslåtte arbeidet er ambisiøst og nytenkende.
- **Virkning** – Potensiell virkning og nytteverdi av det foreslåtte prosjektet. Formidling og deling av resultater og utnyttelsen av dem.
- **Gjennomføring** – Kvaliteten på prosjektets organisering, styring og ressursbruk (inkludert kompetansen til prosjektleder og prosjektgruppe).
- **Stiftelsen Dams prioriteringer** – I hvilken grad prosjektet oppfyller de spesielle prioriteringene styret har nedfelt i utlysningene. Dette varierer mellom programmene våre, men frivillighet og brukermedvirkning er to elementer som står sentralt i alle programmer.

Det er fem hovedårsaker til hvorfor vi benytter disse fire kriteriene

1. *Kriteriene fanger essensen i alle prosjekter.*
2. *Oversiktlig for søkere og fagutvalgsmedlemmer.*
3. *Mulighet for sammenlikning på tvers av egne programmer og med andre finansiører.*
4. *Økt sannsynlighet for gjenbruk.*
5. *Vide kriterier gir økt fleksibilitet.*
6. *Spørreundersøkelser viste at våre fagutvalgsmedlemmer var positive til å gå over til disse kriteriene.*

4.1.1 Kriteriene fanger essensen i alle prosjekter

Kriteriene er vide, men vi mener likevel de tre første kriteriene fanger opp essensen i de fleste prosjekter.

I alle prosjekter blir *noe* gjort med et *formål* (soliditet og gjennomføring) og det forventes en viss *nytte* av dem (virkning). I tillegg har vi som finansiør noen områder og tema vi ønsker å prioritere. Dette kan løses på flere måter, som for eksempel å avgrense utlysningen eller operere med øremerkede midler til de prioriterte områdene. Vi har landet på en kombinasjon av å avgrense utlysningen og å vektlegge noen tema i et eget kriterium som gir ekstra poeng i søknadsvurderingen.

4.1.2 Oversiktlig for søkere og fagutvalgsmedlemmer

Tidligere hadde hvert program sine egne kriterier. Totalt 41 ulike vurderingskriterier var i bruk. Dette gjorde det uoversiktlig både for søker, fagutvalgsmedlemmene og andre interessenter. Samkjøring av kriteriene på tvers av alle programmene, gjør at søkere og fagutvalgsmedlemmer kan bevege seg på tvers av programmene og likevel kjenne seg igjen.

4.1.3 Mulighet for sammenligninger

Siden programmene ikke kunne sammenliknes med hverandre eller med programmer fra andre finansiører gikk vi også glipp av viktig informasjon.

Vi så til Norges forskningsråd og EU da vi valgte disse kriteriene. De bruker kriteriene *Excellence*, *Impact* og *Implementation*. Hva som inngår i disse kriteriene beskrives grundigere i egne dokumenter (European Union 2019). Slik de beskrives mente vi at de dekker alle sentrale sider ved ethvert prosjekt. I tillegg ønsket vi muligheten til å sammenlikne våre tildelinger med andre tildelinger. EU og Norges forskningsråd står for størstedelen av de tildelingene det er rimelig at vi sammenlikner oss med, derfor var det et naturlig valg.

De tre første kriteriene vi valgte (*Soliditet*, *Virkning* og *Gjennomføring*) sammenfaller med de tre kriteriene som EU og Norges forskningsråd bruker (*Excellence*, *Impact* og *Implementation*). I tillegg la vi til et kriterium som fanger opp det som er særtrekk ved våre programmer.

4.1.4 Økt sannsynlighet for gjenbruk

Siden vi har valgt å bruke de samme kriteriene som EU og Norges forskningsråd, vil søkerne enklere kunne gjenbruke de samme søknadene også disse stedene. Av samme grunn bruker vi også de samme malene for prosjektbeskrivelse og CV og vi benytter også den samme karakterskalaen som Norges forskningsråd.

4.1.5 Vide kriterier gir økt fleksibilitet

Dersom kriteriene blir veldig spesifikke, er det stor risiko for at de må endres når andre omstendigheter forandrer seg. For eksempel hadde vi tidligere et kriterium som het *Klinisk, pasientnær forskning*. Utfordringen med et slikt spesifikt kriterium er at, dersom vårt styre bestemmer seg for å revurdere dette som et satsningsområde, så må kriteriene endres. Nå ligger dette punktet under kriteriet *Stiftelsens prioriteringer* og vi trenger ikke endre kriteriet og scoringsverktøyene vi bruker i datasystemet vårt dersom programmets innretning skulle endre seg.

4.1.6 Våre fagutvalgsmedlemmer var positive

Våre ulike fagutvalg ble presentert for forslaget over og totalt var 85 prosent enig eller helt enig i at vi burde endre til disse kriteriene.

Tabell 6: Hvordan utvalgsmedlemmene stiller seg til påstanden *Jeg tror det generelt sett er positivt om Stiftelsen Dam endrer til det foreslåtte kriteriesettet*

Fagutvalg	Uenig/helt uenig	Verken enig/uenig	Enig/helt enig
Helse	12 % (n=2)	6 % (n=1)	83 % (n=15)
Forskning	0 % (n=0)	13 % (n=1)	88 % (n=7)
Ekspress	12 % (n=1)	0 %	88 % (n=7)
Totalt	9 % (n=3)	6 % (n=2)	85 % (n=29)

Antall mottakere = 81 (svarprosent på 46)

4.2 Sjudelt karakterskala

Vi bruker samme skala i all vår søknadsbehandling. Den går fra 1 til 7, der 7 er beste karakter. Før høsten 2019 brukte vi flere ulike karakterskalaer, avhengig av hvilket program det var snakk om. Behovet for å samkjøre karakterskalaen, gjorde at vi våren 2019 kartla hvilken karakterskala som ville egne seg best for oss.

En slik karakterskala skal fylle flere behov som ikke alltid går godt sammen. Den bør for eksempel være enkel å forstå for de som skal bruke den, samtidig som den bør ha høy nok oppløsning (ha mange nok trinn) til å skille mellom prosjektene. En tredelt skala er kanskje enkel å forstå, men har for dårlig oppløsning.

Slike motstridende forhold, og det at det kan være problematisk å endre skalaer man allerede har tatt i bruk, er noe av årsaken til at ulike finansiører har landet på ulike skalaer og at det har vært vanskelig å enes om én.

I sin gjennomgang av alternative skalaer landet en gruppe ved National Institutes of Health (NIH) i USA på at en **sjudelt** skala var det beste alternativet. NIH bestemte seg til slutt likevel for en **nidelt** skala fordi de ønsket høyere oppløsning (National Institutes of Health 2009). I Storbritannia bruker National Health Service en seksdelt skala (National Institute of Health Research 2019).

I Norge bruker de regionale helseforetakene en **femdelt skala** (Nasjonalt program for klinisk behandlingsforskning i spesialisthelsetjenesten 2019), mens Kreftforeningen (Norwegian Cancer Society 2019), Nasjonalforeningen for folkehelsen og Norges forskningsråd bruker en **syvdelt** skala. Norges forskningsråd skal muligens endre sin til den **seksdelte** skalaen som EU bruker (Oxley 2016; Nasjonalt program for klinisk behandlingsforskning i spesialisthelsetjenesten 2019; European Union 2019).

Vi har valgt den sjudelte skalaen av to hovedårsaker:

1. *Balansen mellom oppløsning og enkelhet er god.*
2. *Sjudelt skala muliggjør sammenlikning med andre finansiører.*

4.2.1 Balansen mellom oppløsning og enkelhet er god

I programmet Forskning har vi en innvilgelsesprosent på rundt 13. Det er med andre ord få prosjekter som blir innvilget og derfor trenger vi en skala som er finmasket nok til å skille mellom de beste og de nest beste søknadene. En tredelt skala (som for eksempel "dårlig", "ok", "bra") ville endt med altfor mange søknader med toppscore. En sjudelt skala gir oss nok oppløsning til å skille mellom prosjektene. I 2018 og 2019 fikk i snitt 4 prosent av forskningssøknadene toppkarakteren 7, mens 21 prosent fikk nest høyeste karakter (6).

En syvdelt skala er også håndterbar i scoringsverktøyet som våre fagutvalgsmedlemmer bruker. Det er viktig for oss at fagutvalgsmedlemmenes arbeid blir så enkelt som mulig og vi ønsker ikke at de skal måtte bruke for mye tid på å taste inn tall eller velge dem fra nedtrekksmenyer. Derfor bruker vi

en scoringsmatrise, der utvalgsmedlemmene bare klikker på den scoren de ønsker å gi. Det vil raskt bli uoversiktlig dersom vi bruker en skala med for mange ledd.

4.2.2 Sjudelt skala muliggjør sammenlikning med andre finansiører

Som nevnt over bruker både Norges forskningsråd, Kreftforeningen og Nasjonalforeningen for folkehelsen sjudelte skalaer. Alle disse aktørene deler ut midler til liknende forskning som vi støtter og det er en fordel å kunne sammenlikning våre vurderinger med deres. Vi vil for eksempel lettere kunne se om fagutvalgsmedlemmene våre bruker skalaen veldig forskjellig og om det er ulik grad av enighet mellom fagutvalgsmedlemmer som vurderer samme søknad.

4.3 Beregnet total karakter på søknaden basert på kriterie karakterene

Når vurderingen til våre fagutvalgsmedlemmer er gjennomført, *beregner* vi en gjennomsnittskarakter basert på de fire kriteriene. Alle kriteriene vektet likt og gjennomsnittskarakteren er ikke synlig for fagutvalgsmedlemmene.

Det skiller seg fra hva vi gjorde tidligere. Før august 2019 satte fagutvalgsmedlemmene karakterer på de ulike kriteriene, men de satte også en total karakter som var uavhengig av kriteriene. Det var denne total karakteren som ble utslagsgivende for den videre behandlingen.

4.3.1 Vi ønsker størst mulig fokus på våre kriterier

Vi endret til beregnet total karakter fordi vi mistenker at muligheten til å sette denne uavhengig av kriteriene, øker sannsynligheten for at fagutvalgsmedlemmene vektlegger andre faktorer enn vi ønsker. Det er *kriteriene* de skal vurdere, ikke noe annet. Slik håper vi å demme opp for noen av de uønskede skjevhetene som vi beskrev under Hva vi bekymrer oss for.

4.4 Faste ekspertgrupper, ingen eksterne eksperter

I dag oppnevner vi faste grupper av eksperter, kalt fagutvalg, for perioder på fire år. Gruppene settes sammen basert på forhåndsbestemte kriterier som beskrives i rutinebeskrivelsen "Oppnevning av medlemmer til Stiftelsen Dams fagutvalg" (se dam.no/oppnevning-fagutvalg). Der stilles det krav til gruppens samlede faglige kompetanse og erfaringsbakgrunn.

Et fast fagutvalg betyr også at søknadene må fordeles til de fagutvalgsmedlemmene vi har til enhver tid. Mange andre finansiører henter i stedet, eller i tillegg, inn eksperter *etter* at søknadene er mottatt og håndplukker dem til den enkelte søknad. Det er flere gode grunner til å gjøre det slik. Viktigste er at de ekspertene med mest kompetanse innen temaet søknaden omhandler, må antas å være best i stand til å vurdere søknaden.

Når vi skal sette sammen fagutvalgene våre, klarer vi aldri å dekke alle fagområdene og temaene i søknadene. Derfor vil mange søknader bli vurdert av fagutvalgsmedlemmer som ikke har inngående og oppdatert kunnskap om det søknaden omhandler. Det kan være et betydelig problem, spesielt for komplekse prosjekter som stiller store krav til kjennskap til kunnskapsstatus og metoder innen det

aktuelle feltet. Noe forskning tyder også på at forskningssøknader kan bli vurdert strengere når de vurderes av fagutvalgsmedlemmer utenfor det fagfeltet søknaden omhandler (Porter and Rossini 1985). En mulig årsak til dette kan være at mennesker har en tendens til å foretrekke ting vi er godt kjent med, også omtalt som eksponeringseffekten (Harrison 1977). I paneldiskusjoner kan dette gi seg utslag i at fagutvalgsmedlemmene "kjemper" litt mindre for søknadene som ikke omhandler det de kan best (Travis and Collins 1991)

Vi har likevel valgt å bruke faste fagutvalgsmedlemmer fordi det:

1. *er forutsigbart hvilke søknader vi får og hvilken kompetanse vi dermed trenger.*
2. *ikke alltid er åpenbart at det gir bedre vurderinger å bruke eksperter som er håndplukket til den enkelte søknad.*
3. *er enklere å lære opp utvalgsmedlemmene, og de får mye "trening" i å vurdere søknader.*
4. *er rimeligere og tidsbesparende.*

4.4.1 Hvilken kompetanse vi trenger er forutsigbart

Vi vet omtrent hvilke søknader vi kommer til å få. Vi behandler årlig rundt 3.000 søknader og selv i det programmet vi får færrest søknader, får vi rundt 300 søknader årlig. Selv om det skjer endringer i hva søknadene omhandler tematisk fra år til år, er endringene relativt begrensede. Det gjør at vi har et godt grunnlag for å vite hvor mange fagutvalgsmedlemmer vi trenger og hvilken kompetanse de bør ha.

4.4.2 Gir ikke åpenbart bedre vurderinger

Selv om det høres rimelig ut at man får bedre vurderinger av å håndplukke eksperter til hver enkelt søknad, er det flere problemer forbundet med det. For det første er det ikke enkelt å finne eksperter som i tillegg er uavhengige. Ofte vil de som har mest kompetanse innen området søknaden gjelder kunne ha egeninteresser som er ugunstig for søknadsvurderingen. For det andre er det vanskelig å lære opp et stort antall eksterne eksperter i hvordan søknadsvurderingen skal gjennomføres. Dette kan i sin tur lede til en avvikende tolkning av kriterier og karakterskala (Hodgson 1995; Thorngate n.d.). Dette er noe av årsaken til at forskeren Nancy E. Mayo og hennes kollegaer har anbefalt "å gi opp søket etter eksterne eksperter" (Mayo et al. 2006).

4.4.3 Opplæringen blir enklere

Som de fleste andre finansører har vi vedtekter og utlysninger som beskriver hva vi skal dele ut penger til. Det tar tid å bli kjent med innholdet i disse dokumentene og faste utvalg gjør opplæringen enklere. Det samme gjelder også alt som følger med selve søknadsbehandlingen, som for eksempel å bli kjent med våre vurderingskriterier og vårt datasystem. Faste grupper av eksperter gjør at vi kan gjennomføre felles opplæring og gjør det også enklere å følge opp de ulike ekspertene.

I tillegg får hvert enkelt fagutvalgsmedlem flere søknader å vurdere, noe som både gir dem mer trening og en bedre referanseramme når karakterene skal settes. Forskning tyder på at dette kanskje kan bidra til en mer enhetlig og mer kostnadseffektiv vurdering av søknadene (Jayasinghe et al. 2006). I en undersøkelse av et vurderingssystem (omtalt som "the new reader system") i det australske forskningsrådet, der ett av elementene var at grupper av eksperter vurderte alle søknadene innenfor sitt felt, konkluderte disse forskerne med følgende:

(...) the new reader system resulted in substantially higher reliabilities: 0.643 and 0.881, respectively. In comparison to the traditional peer review approach, the new reader system is substantially more reliable, timely, and cost efficient – and applicable to many peer review situations.

4.4.4 Det er rimeligere

Å finne nye eksperter til hver søknad, er svært tidkrevende for administrasjonen og dermed også kostbart. I en stor internasjonal spørreundersøkelse blant finansierer, svarte enkelte at de brukte opptil seks timer per søknad for å finne passende eksperter (Hardcastle 2019). Det tar tid, ikke bare fordi det er vanskelig å finne eksperter, men også fordi mange takker nei.

Vi har opprettet en egen database der personer som er interesserte i bistå oss i vurderingen av søknadene kan registrere seg. På dam.no/vurdere-soknader kan hvemsomhelst legge inn detaljer om sin kompetanse og per 20. januar 2020 hadde 342 personer registrert seg. Det har gjort det vesentlig enklere for oss å finne en god balanse i fagutvalgenes samlede kompetanse og nesten ingen svarer nei når vi spør dem.

4.5 Bredt sammensatte fagutvalg

Vi har i dag fire ulike fagutvalg som vurderer søknadene i de fire programmene våre (Helse, Forskning, Utvikling og Ekspres). Det er svært ulik profil på disse fagutvalgene, men felles for alle utvalgene er at de er bredt sammensatt. Både hva gjelder faglig kompetanse, alder, kjønn og geografi. Rutinebeskrivelsen Oppnevning av medlemmer til Stiftelsen Dams fagutvalg gir detaljer om hvordan oppnevningen skjer (se dam.no/oppnevning-fagutvalg).

Det er flere utfordringer knyttet til det å bruke bredt sammensatte fagutvalg. For det første vil opplæringen av utvalget bli mer krevende, da det vil være større variasjon i kompetanse når ulike problemstillinger skal tas opp. For det andre er det stor sannsynlighet for at større faglig diversitet bidrar til lavere enighet mellom fagutvalgsmedlemmer som vurderer samme søknad. Lav enighet er et problem fordi det betyr at hvem som tilfeldigvis vurderer søknaden, får avgjørende betydning for utfallet av vurderingen. Dersom samme søknad kunne ha havnet hos andre fagutvalgsmedlemmer – slik tilfellet er hos oss – betyr lav enighet at samme søknad sendt flere ganger med høy sannsynlighet vil få ulikt utfall. Vi skriver mer om dette problemet under "Veien videre".

Vi har likevel valgt å fortsette med bredt sammensatte fagutvalg fordi:

- *Vi tror diversitet kan ha positiv effekt på produktivitet og innovasjon.*
- *Vi tror det øker sannsynligheten for å oppdage styrker og svakheter.*
- *Våre brede utlysninger gjør at vi trenger bredt sammensatte fagutvalg.*

4.5.1 Diversitet kan ha positiv effekt på produktivitet og innovasjon

Vi bruker ikke bare fagutvalgsmedlemmene våre til søknadsbehandling. De er også viktige i stiftelsens utvikling gjennom å gi oss tilbakemelding på prosessene de er en del av. Da er det også viktig at fagutvalgsmedlemmene er ulike mennesker med ulike perspektiver. Studier tyder på at dette kan gi både økt produktivitet og mer innovasjon (Mannix and Neale 2005; Jeppesen 2007).

4.5.2 Større sannsynlighet for å oppdage styrker og svakheter

Poenget med ulike perspektiver, gjelder også for selve søknadsbehandlingen. Ulike personer bringer med seg ulike måter å se verden på. Det gjenspeiler seg også i søknadsvurderingen. Både faglige og personlige kunnskaper, holdninger og interesser gjør at både søknadens styrker og svakheter, vurderingskriteriene og karakterskalaen vil bli tolket ulikt. Et bredt sammensatt utvalg øker sannsynligheten for at flest mulig av disse ulike perspektivene ivaretas når søknaden vurderes.

4.5.3 Brede utlysninger gjør at fagutvalgene må ha stor bredde i kompetansen

Siden medlemsorganisasjonene som står bak stiftelsen favner svært bredt, skal også utlysningene til Stiftelsen Dam favne bredt. Det gjør at vi får mange ulike søknader, både hva gjelder tema og størrelse. Derfor vil vi alltid være avhengig av å ha fagutvalgsmedlemmer fra mange ulike fagfelt.

4.6 Flere uavhengige vurderinger på hver søknad

Antallet vurderinger søknadene får, avhenger av hvilket program de faller inn under og hvor mange ledd søknaden går gjennom i hvert program. Oppsummert ser det slik ut:

Tabell 7: Antall vurderinger per søknad per program

Program	Antall uavhengige vurderinger per søknad	Søknadssum (snitt)	Innvilgelsesandel
<u>Helse</u>	3-6	ca 750.000	ca 20 %
<u>Forskning*</u>	5	ca 2.000.000	ca 13 %
<u>Utvikling*</u>	4	ca 1.500.000	ca 9 %**
<u>Ekspress</u>	2	ca 25.000	ca 30 %

* Programmer med en prekvalifiseringsrunde

** Basert på antall skissesøknader (Innvilgelsesandel basert på antall utvidede søknader var 34 %).

Som det fremgår av tabellen er antallet vurderinger tett knyttet til størrelsen på prosjektene det søkes midler til. I Ekspress søker de største prosjektene om 30.000, mens i Forskning er den gjennomsnittlige søknadssummen på cirka to millioner kroner.

Hovedårsaken til at vi bruker flere vurderinger av en søknad, er at vi ønsker å

1. jevne ut uheldige variasjoner mellom fagutvalgsmedlemmene.
2. øke sannsynligheten for å avdekke styrker og svakheter.

4.6.1 Jevner ut variasjoner mellom fagutvalgsmedlemmene

Det er en kjent sak at det ofte er stor uenighet mellom de som vurderer prosjektsøknader (i mange ulike sammenhenger). Stor uenighet betyr at tilfeldigheter (i motsetning til egenskaper ved prosjektet) får større betydning i søknadsbehandlingen og den såkalte reliabiliteten² til utvelgelsesprosessen går ned (Mayo et al. 2006). Slike uheldige utslag kan reduseres ved å få flere personer til å vurdere søknaden (Cicchetti 1991). Da vil også reliabiliteten øke.

Flere har advart mot å støtte seg på vurderingene til kun to personer (Mayo et al. 2006; Cicchetti 1991). European Science Foundation anbefaler *minimum* tre eksperter per forskningssøknad (European Science Foundation 2011) for å redusere den tilfeldige variasjonen. Det samme gjorde en sveitsisk forskergruppe i 2012 (Mutz et al. 2012), og i 2015 kom en artikkel som konkluderte med at fem eksperter gir den beste balansen mellom kostnader og nytte (Snell 2015).

Disse tallene baserer seg på vurderinger av forskningssøknader, der det ofte er store summer involvert, men selv i de tilfellene er det vanlig å tilpasse antallet vurderinger til søknadens type og omfang. Cancer Research UK uttaler for eksempel at *For most grants, and depending on the size of the award, we seek anywhere between three and six written reviews* (Shailes 2017).

² Reliabilitet brukes om konsistens eller stabilitet i målinger. Se snl.no/reliabilitet

I programmet Ekspress avviker vi fra dette og gir søknaden kun to vurderinger. Det er fordi alle søknadene ligger mellom 5.000 og 30.000 kroner og at søknadsbehandlingen skal være enkel.

4.6.2 Øke sannsynligheten for å avdekke styrker og svakheter

Økt reliabilitet er ikke den eneste årsaken til at vi vanligvis gir søknadene tre til seks vurderinger. Reliabilitet alene sier ingenting om vurderingene er gyldige og faktisk plukker ut de beste prosjektene (har god validitet³). God reliabilitet er en nødvendig, men ikke tilstrekkelig, forutsetning for god validitet. Reliabiliteten kan være god uten at validiteten er det. Håpet er imidlertid at flere vurderinger også øker validiteten, ved å øke sannsynligheten for å identifisere styrker og svakheter ved prosjektet. Dette er imidlertid en antakelse. Å måle validiteten (hvor godt vurderingene treffer), er svært komplisert og noe som gjøres i liten grad blant andre finansører. Vi jobber for å få bedre validitetsmål på plass og kommer tilbake til det i senere rapporter.

4.7 Prekvalifiseringsrunde

I de to programmene Forskning og Utvikling bruker vi en prekvalifiseringsrunde der søkerne først sender en svært forenklet skissesøknad. Selve søknaden fylles i sin helhet ut i et elektronisk skjema (det lastes ikke opp noen prosjektbeskrivelse) og søkeren gis totalt 5.800 tegn fordelt på ulike forhåndsdefinerte tekstfelt. Det tilsvarer en Aftenposten-kronikk. Basert på fagutvalgets vurdering av denne skissen, inviteres et begrenset utvalg til å sende utvidede søknader i runde to.

Det var noen viktige grunner til at vi innførte skissesøknader:

1. Det reduserer tidsbruken for søker.
2. Svartiden blir kortere.
3. Det blir rimeligere, noe som ga oss mulighet til å øke antallet vurderinger per søknad.
4. Våre interessenter var positive til det.
5. Vi antok at skissesøknaden kunne være nok til å presentere prosjektets potensial.

4.7.1 Redusere tidsbruken for søker

I programmet Forskning har innvilgsandelen ligget rundt 13 prosent. Det er bekymringsverdig lavt.

Den største andelen av ressursbruken knyttet til søknadsprosesser bæres nettopp av søkerne. En undersøkelse fra National Health and Medical Research Council i Australia viste at forskningsorganisasjonene stod for 85 prosent av den totale ressursbruken, nettopp fordi søknadsskrivingen tar så mye tid (Graves et al. 2011). Selv om arbeidet med søknaden kan være nyttig, selv om den ikke innvilges, er vår holdning at det er en del av vårt ansvar å sørge for at søknadsprosessen er så lite tidkrevende som mulig. Da vil belastningen på søkerne og dermed det norske samfunnet bli minst mulig.

Både våre og undersøkelser fra Norges forskningsråd viser at det går med svært mye tid til å søke. Som Tabell 8 viser, brukte våre søkere i gjennomsnitt 8,2 ukeverk per søknad. En undersøkelse

³ Validitet sier noe om i hvilken grad man ut fra resultatene av et forsøk eller en studie kan trekke gyldige slutninger om det man har satt seg som formål å undersøke. Se snl.no/validitet

gjennomført av NIFU for Norges forskningsråd viste at gjennomsnittet hos dem lå på 9,0 ukeverk (Ramberg 2017).

Tabell 8: Søkernes selvrapporterte tidsbruk i ukeverk

Gjennomsnittlig tidsbruk (median i parentes)	Norges forskningsråd	Forskning 2019
Ukeverk ved egen institusjon	6,8 (4,0)	6,6 (4,0)
Ukeverk ved eksterne partnerinstitusjoner	2,2 (1,0)	1,6 (0,6)
Totalt	9,0 (5,0)	8,1 (4,6)

Både NIFUs og våre undersøkelser viser for øvrig at søknadene som sendes inn i begrenset grad bygger på tidligere søknader. Som Tabell 9 viser er det kun 14 prosent av våre søknader som i stor grad er gjenbruk av tidligere søknader.

Tabell 9: Gjenbruk blant søkere

Andel gjenbruk	Norges forskningsråd	Forskning 2019
Bygget ikke direkte på arbeid gjort for noen tidligere søknad	44 %	47 %
I noen grad gjenbruk av tidligere søknad	38 %	36 %
I stor grad gjenbruk av tidligere søknad	15 %	14 %

NIFUs undersøkelse viste også at søkerne selv mener at en prekvalifiseringsrunde kan spare tid. I undersøkelsen ble søkere presentert for ni forskjellige tiltak for å redusere tiden som går med til å skrive søknader og alternativet "En prekvalifiseringsrunde med en enkel skisse/forenklet søknad" var det mest populære tiltaket (Ramberg 2017).

En undersøkelse som så nærmere på konsekvensene av at et program i britiske National Institute for Health Research (NIHR) gikk fra en ett- til totrinnsprosess, konkluderte med at totrinnsprosesser var vesentlig mindre belastende for søker (Morgan et al. 2020).

Med bakgrunn i disse undersøkelsene og vårt ønske om å begrense belastningen på søker, besluttet vi å innføre en slik prekvalifiseringsrunde for programmene Utvikling og Forskning, henholdsvis fra og med utlysningene i 2019 og 2020. I utlysningen for Forskning fikk vi inn 364 skissesøknader, hvorav hundre ble invitert til å sende utvidede søknader. Som Tabell 10 viser, ga den nye søknadsbehandlingen betydelig reduksjoner i både tidsbruk per søknad, total tidsbruk og i svartiden.

For de rundt 70 prosent av søkerne som får avslag i prekvalifiseringen, faller gjennomsnittlig tidsbruk med 58 prosent, fra 8,1 til 3,4 ukeverk. Den gjennomsnittlige tidsbruken per søker faller med 42 prosent fra 8,1 til 4,7 ukeverk, og den totale tidsbruken for utlysningen faller med 23 prosent fra 47,8 til 36,7 årsverk. Anslagsvis 11,1 årsverk ble med andre ord spart etter overgangen til prekvalifiseringsfasen. Dette til tross for at utlysningen i 2019 fikk uvanlig få søknader.

Tabell 10: Sammenligning av tidsbruk og svartid før (2019) og etter (2020) innføring av prekvalifisering i programmet Forskning

Tidsbruk og svartid	2019	2020 prekvalifisering	2020 utvidet søknad	2020 totalt*
Ukeverk per søknad (median)	8,1 (4,6)	3,4 (1,4)	5,2 (4,4)	4,7 (2,5)
Totalt antall årsverk	47,8	26,7	10,0	36,7
Dagsverk per 100.000 bevilgede kroner	12,7	-	-	11,8
Gjennomsnittlig svartid	136 dager	57 dager	127 dager	88 dager
Andel som er fornøyd/svært fornøyd med søknadsprosessen totalt sett	93 %	85 %	91 %	-
Andel som er misfornøyd/svært misfornøyd med søknadsprosessen totalt sett	3 %	2 %	1 %	-
Antall søknader	271	366	90	366

* Beregningene her baserer seg på tidsbruken forbundet både med skissesøknadene i prekvalifiseringsrunden og de utvidede søknadene i runde to.

4.7.2 Raskere svar

En annen positiv effekt av å ha en prekvalifiseringsrunde er at flertallet av søkerne vil få raskere svar på søknaden sin. Før vi begynte med skissesøknader var søknadsfristen 1. juni og offentliggjøringen 15. oktober. Dermed måtte alle søkerne vente minst 136 dager på svar. Med innføringen av en prekvalifiseringsrunde, ble søknadsfristen for denne runden 15. februar, med tilbakemelding 5. april. For rundt tre fjerdedeler av søkerne blir dette slutten på søknadsprosessen. Dermed vil rundt 75 prosent få svar innen 57 dager.

Den fjerdedelen som går videre må (som tidligere) vente til 15. oktober for å få svar. Selv om disse må vente noe lengre, vil gjennomsnitt svartid bli 92 dager. Det er 32 prosent mindre enn tidligere.

4.7.3 Rimeligere

Bruk av tottrinnsprosesser har vist seg ikke bare å spare søkeren for tidsbruk. I den nevnte undersøkelsen av søknadsprosesser i NIHR konkluderer forfatterne at kostnadene forbundet med programmet gikk med fra GBP 148.908 for ettrinnsprosessen til GBP 105.342 for tottrinnsprosessen (Morgan et al. 2020). Det er en besparelse på 29 prosent.

Stiftelsen Dam har også hatt betydelig besparelser, men siden vi samtidig økte antallet individuelle vurderinger per søknad fra to til fem, har de totale kostnadene for programmet blitt omtrent de samme. Vi kan imidlertid sammenlikne kostnaden forbundet med en enkelt individuell vurdering *før* og *etter* vi begynte med prekvalifiseringsprosessen.

I gjennomsnitt fikk vi 298 søknader for hver utlysning fra 2010-2019. I dag godtgjøres fagutvalgsmedlemmene 1,5 times arbeid etter statens satser for deltakelse i utvalgsmøter. Legger vi

2019-satsene til grunn blir det 218.666 kroner for to vurderinger for den gjennomsnittlige søknadsrunden (1,5 timer * 478 kr * 298 søknader).

I 2020 startet vi med prekvalifisering og fikk 366 skissesøknader. Fagutvalgsmedlemmene godtgjøres med 20 minutters arbeid for vurderingen av disse søknadene. Dette utgjør 58.316 kroner (0,33 timer * 478 kr * 366 skissesøknader). I tillegg er det lagt opp til at 100 utvidede søknader skal behandles i runde to, noe som gir en kostnad på 71.700 kroner (1,5 timer * 478 kr * 100 søknader). Til sammen gir det en total kostnad for to individuelle vurderinger på 130.016 kroner.

Nedgangen fra 218.666 til 130.016 kroner gir en besparelse på 39 prosent.

4.7.4 Positive tilbakemeldinger fra våre interessenter

Forslaget om å innføre skissesøknader ble presentert både for fagutvalget, alle medlemsorganisasjonene som hadde sendt forskningssøknader siste to år og alle søkere som deltok i søknader i 2018. I en spørreundersøkelse høsten 2018 fikk de muligheten til å si sin mening om forslaget.

Som det fremgår av tabellen er de forskjellige gruppene interessenter gjennomgående positive til forslagene om skissesøknader. En "prosjektkoordinator" er i denne sammenhengen representanter tilknyttet stiftelsen Dams søkerorganisasjoner som har koordineringsansvaret for organisasjonens søknader og prosjekter.

Tabell 11: Hvordan ulike grupper interessenter stiller seg til påstanden "Jeg synes Stiftelsen Dam bør gå over til skissesøknader"

Gruppe	Uenig/helt uenig	Verken enig/uenig	Enig/helt enig
Prosjektkoordinatorer (n=49 av 120)	4 % (n=2)	6 % (n=3)	90 % (n=44)
Søker (n=116 av 306)	23 % (n=27)	7 % (n=8)	70 % (n=81)
Fagutvalg (n=13 av 21)*	8 % (n=1)	0 % (n=0)	92 % (n=12)
Totalt (n=178 av 447)	17 % (n=30)	6 % (n=11)	77 % (n=137)

Total svarprosent var 40 (178 av 447).

* For fagutvalget var påstanden "Med tanke på søkerne tror jeg det er positivt om Stiftelsen Dam begynner med skissesøknader/prekvalifisering"

4.7.5 Skissesøknadene gir antakelig nok informasjon til å vurdere potensialet

En viktig forutsetning for å bruke skissesøknader er at de gir et tilstrekkelig godt bilde av potensialet i det foreslåtte prosjektet. Dersom skissesøknaden blir så enkel at det ikke er plass til den mest sentrale informasjonen, får ikke søkerne vist hva de er gode for og fagutvalgsmedlemmene får for lite informasjon til å gjøre vurderingene sine. Basert på erfaringer fra andre finansiører og våre egne erfaringer fra arbeidet med Kavlifondets program for helseforskning, har vi laget et søknadsskjema som er kort, men som vi tror gir søkeren rom for å vise de viktigste delene av prosjektet.

Da vi spurte våre fagutvalgsmedlemmer om de mente at skissesøknaden var tilstrekkelig for å vurdere prosjektets potensial, svarte 88 prosent at de var litt eller helt enig. I kommende utlysninger vil vi spørre søkerne om det samme.

4.8 Siling av søknadene

I de tre programmene Helse, Utvikling og Forskning siles vi søknadene i løpet av søknadsbehandlingen (se Søknadsbehandlingen i våre programmer). Det vil si at noen søknader avslås basert på en innledende vurdering (basert på tre til fem uavhengige vurderinger), mens de resterende går videre til en enda grundigere vurdering. Det er mellom 65 og 75 prosent av søknadene som avslås.

En ulempe med at så mange søknader siles ut er at søknader med potensiale vil bli silt ut. Noen søknader som *kunne* ha endt opp med positiv vurdering i en vurderingsrunde nummer to, og i tillegg kanskje ha endt opp med å bli gode prosjekter, vil dessverre bli luket ut. Det er problematisk for oss, som "mister" et godt prosjekt, og for søkerne, som legger ned mye tid i søknadsarbeidet.

Vi gjennomfører likevel en slik utsiling fordi:

- vi ønsker å begrense ressursbruken forbundet med vurderingene.
- vurderingene er bedre egnet til å luke ut svake søknader, fremfor å skille mellom de beste.
- fordi denne utsilingen frigjør tid til å diskutere gråsonetilfellene grundigere.

4.8.1 Begrenser ressursbruk

Ideelt sett skulle alle søknadene våre fått samme grundige behandling. Søknadsbehandling er imidlertid en kostbar prosess og som andre deler av denne rapporten belyser, ønsker vi å begrense kostnadene forbundet med vurderingene. En måte å gjøre dette på, er å bruke mindre ressurser på å vurdere søknadene som har lav sannsynlighet for å bli innvilget.

4.8.2 Vurderinger best egnet til å luke ut svake søknader

Mye tyder på at vurderingene i fagutvalget egner seg godt til å luke ut de svakeste søknadene. Blant annet konkluderer forfatterne av en kjent studie fra 1991 med at "*there is (...) much more agreement on rejection than acceptance*" (Cicchetti 1991) og i en studie fra 2015 oppfordrer forskerne utsiling av både de beste og de svakeste søknadene:

Results support aggressive high- and low-end stratification or triaging of applications for subsequent stages of review, with the proportion and set of mid-range submissions to be retained for further consideration being dependent on overall success rate. (Snell 2015)

Det er mange forbehold knyttet til akkurat dette. Som sitatet viser bør utsilingen betinges av hvor stor andel av søknadene som kan innvilges. I tillegg vil spredningen i kvaliteten på søknadene ha mye å si. Dersom alle søknadene holder veldig lik kvalitet, blir grunnlaget for å skille mellom dem dårligere enn om det er stor spredning.

4.8.3 Mer tid på gråsonetilfellene

Tidligere ble alle søknadene behandlet i to runder. Innføringen av utsilingen har frigjort tid til å bruke mer tid på å diskutere gråsonetilfellene. Det gir søknadene som er aktuelle for innvilgelse en ekstra grundig behandling og det gir fagutvalgene bedre mulighet til å lære av hverandre.

4.9 Mindre grupper av fagpersoner, fremfor stort plenum

I programmet Forskning gjorde vi det tidligere slik at søknadene i siste instans ble vurdert av hele fagutvalget samlet (omtalt som "plenum"). I dette plenumet satt 18 fagutvalgsmedlemmer, en leder for fagutvalget og to brukerrepresentanter. Der ble alle søknadene som var aktuelle for innvilgelse diskutert. Under siste utlysning der vi gjorde dette, ble 117 søknader behandlet over tre dager i dette plenumet. I snitt hadde plenumet fire minutter til diskusjon per søknad.

Før utlysningen i 2020 besluttet vi å endre denne delen av søknadsbehandlingen. Plenumet ble erstattet av seks grupper bestående av fem fagutvalgsmedlemmer hver.

Denne endringen ble gjort av følgende grunner:

- Plenumsdiskusjonene påvirket i liten grad søknadsbehandlingen.
- Mer tid til diskusjoner.
- Bedre grunnlag for deltakelse i diskusjonene.
- Rimeligere.

4.9.1 Små bevegelser i store plenum

Enkelte tidligere studier har vist at paneldiskusjoner har begrenset betydning. Blant annet viste Carpenter et al. at paneldiskusjoner bare påvirket finansieringsbeslutningen (innvilgelse eller avslag) i ~~ti~~ 10 prosent av tilfellene, sammenliknet med hva resultatet ville ha blitt om man bare baserte seg på de individuelle vurderingene til ekspertene (Carpenter et al. 2015).

Vi har gjort liknende analyser og har sett på hvilke søknader som ville ha blitt innvilget og avslått på bakgrunn av vurderingene som fagutvalgsmedlemmene gjør før plenumsbehandlingen. Med andre ord: Hva resultatet ville ha blitt dersom vi hadde droppet plenumsbehandlingen.

Som Tabell 12 viser, ville det for tildelingen i Forskning i 2018 ha resultert i at 97 prosent av beslutningene ville ha blitt identiske. Kun fem av 164 søknader ville ha fått et annet utfall dersom vi hadde bestemt oss uten å gjennomføre plenumsbehandlingen.

Tabell 12: Forskning 2018. Hvordan søknadsbehandlingen ville ha slått ut dersom den siste plenumsbehandlingen ikke hadde blitt gjennomført

	Avslag, etter plenumsbehandling	Innvilgelse, etter plenumsbehandling	Totalt
Avslag, før plenumsbehandling	81 % * (n=133)	2 % (n=3)	83 % (n=136)
Innvilgelse, før plenumsbehandling	1 % (n=2)	16 % * (n=26)	17 % (n=28)
Totalt	82 % (n=135)	18 % (n=29)	100 % (n=164)

* Totalt ville 97 prosent av beslutningene blitt de samme dersom plenumsbehandlingen *ikke* ble gjennomført.

Vi har gjort de samme analysene i programmet Helse der vi fortsatt bruker en slik plenumsbehandling og også der viser det seg at 97 prosent av beslutningene ville ha blitt identiske. Kun 16 av 513 søknader ville ha fått et annet utfall dersom plenumsbehandlingen hadde blitt droppet.

Tabell 13: Helse vår 2019. Hvordan søknadsbehandlingen ville ha slått ut dersom den siste plenumsbehandlingen ikke hadde blitt gjennomført

	Avslag, etter plenumsbehandling	Innvilgelse, etter plenumsbehandling	Totalt
Avslag, før plenumsbehandling	79 % * (n=404)	2 % (n=9)	81 % (n=413)
Innvilgelse, før plenumsbehandling	1 % (n=7)	18 % * (n=93)	19 % (n=100)
Totalt	80 % (n=411)	20 % (n=102)	100 % (n=513)

* Totalt ville 97 prosent av beslutningene blitt de samme dersom plenumsbehandlingen *ikke* ble gjennomført.

Hvorfor plenumsbehandlingen har gitt lite utslag hos oss er uklart, men vi tror to faktorer antakelig spiller en vesentlig rolle:

1. Det er lite tid til diskusjon i plenum.
2. De fleste i plenum har ikke lest den fulle søknaden (i beste fall et sammendrag) og har dårlig grunnlag for å argumentere mot dem som har hatt ansvaret for å vurdere søknaden i rundene før plenum.

I de nye gruppemøtene forsøker vi å demme opp for begge disse utfordringene (se under).

4.9.2 Mer tid til diskusjoner

Slik plenumsbehandlingen er lagt opp i dag, har plenum i snitt omtrent fire minutter til å behandle hver forskningsøknad. Det er et godt stykke under hva andre finansører opererer med. For EU-programmet Research and Innovation Actions er det avsatt to timer per søknad for diskusjon og hos Norges forskningsråd brukes normalt mellom 30 og 45 minutter per søknad (noe avhengig av

hvilke program det dreier seg om). Norges forskningsråd vurderer å gjøre det til et krav å ha 30 minutter til rådighet per søknad – etter at de svakeste søknadene er ekskludert.

Årsaken til at de ønsker å sikre denne tidsbruken er at forskning viser at arbeid under tidspress i større grad fører til at tilfeldigheter blir utslagsgivende og *reduserer ekspertenes ambisjoner om og evne til å gjennomføre grundige vurderinger (Oxley 2016).*

Våre gruppemøter skal være en arbeidsdag og gruppene vil få cirka 15 søknader til diskusjon. Det betyr i snitt 30 minutter til diskusjon for hver søknad.

4.9.3 Bedre grunnlag for deltakelse i diskusjonene

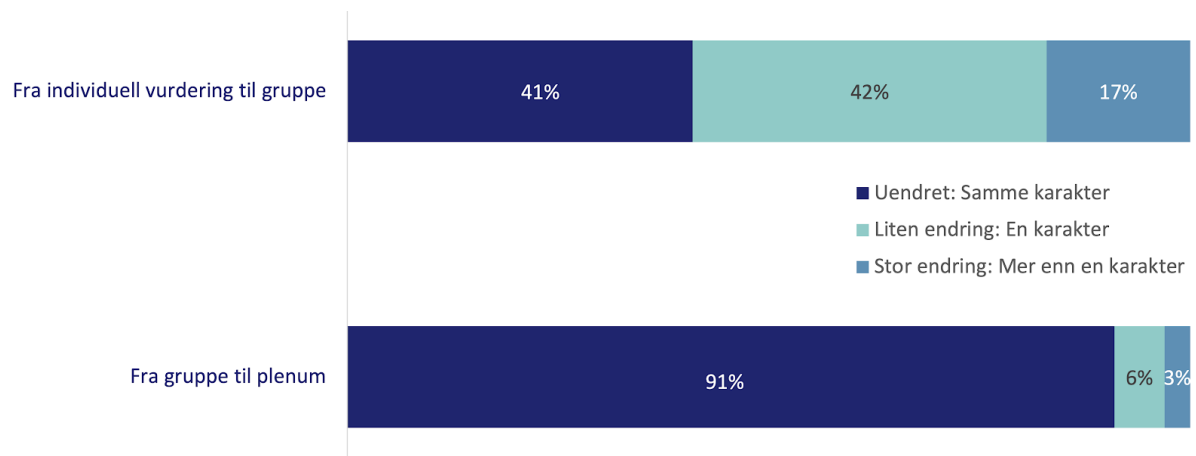
I den nye modellen skal gruppe medlemmene vurdere søknadene på like vilkår. Ingen har rollen som “saksordfører” med et spesielt ansvar for å sette seg inn i eller presentere søknaden for de andre. Ansvar ligger på alle og alle må sette sine egne karakterer i gruppemøtet. Med det sikrer vi at alle har et like godt grunnlag for å kommentere på søknadene.

Dette er også i tråd med hva forskningen viser. I et notat fra Norges forskningsråd skriver seniorrådgiver Kristin Oxley at *“Forskning viser at saksordførermodellen er særlig lite egnet for å produsere konsistente vurderinger. Modeller som legger en mer likestilt behandling av panelmedlemmene til grunn gir mer robuste vurderinger” (Oxley 2016).*

Våre erfaringer fra programmet Helse gir også grunnlag for å tro at deltakelsen vil øke. I andre runde av søknadsbehandlingen (etter en utsiling) i dette programmet møtes fagutvalgsmedlemmene i grupper på tre for å diskutere søknader alle har vurdert individuelt på forhånd. Sammen skal de sette en felles karakter og vår opplevelse er at kombinasjonen av at alle har lest søknadene og at de har bedre tid enn i plenum, fører til grundigere diskusjoner enn i plenum.

En indikasjon på dette er også at det er en del bevegelse i karaktersetningen i gruppene. I Figur 6 sammenliknes gjennomsnittet av de tre individuelle karakterene (avrundet til hele karakterer) med den samlede karakteren gruppen ble enig om. I 59 prosent av søknadene (78 av 132) endret karakteren seg, og i 17 prosent av tilfellene (22 av 132) endret den seg med mer enn én karakter (de røde feltene i tabellen).

Dette er vesentlig større endringer enn de som skjer under plenumsbehandlingen. Som Figur 6 viser, endret kun ni prosent av søknadene (13 av 131) karakteren og bare tre prosent endret mer enn én karakter.



Figur 6: Helse vår 2019. Karakterendringer fra individuell vurdering til gruppens vurdering, og fra gruppens vurdering til plenums vurdering

4.9.4 Rimeligere

Å samle 21 personer pluss medlemmer fra sekretariatet i tre dager er svært kostbart. Kostnaden kan forsvares hvis fagfellenes kompetanse kommer til god nytte, men erfaringen viser at det er veldig variabelt hvor mange som engasjerer seg i behandlingen av hver søknad og forskning tyder på at slike plenumsbehandlinger spiller en begrenset rolle for utfallet og at kostnadene ikke står i stil med nytten.

4.10 Ingen konsensuskarakter og forbud mot karaktersnakk i gruppemøter

I programmet Forskning avsluttes søknadsbehandlingen med et gruppemøte der fagutvalgsmedlemmene møtes i mindre grupper for å diskutere søknadene (som alle har vurdert i forkant). I diskusjonen skal de kun trekke frem hva de mener er styrker og svakheter ved søknaden og de får eksplisitt beskjed om *ikke* å si noe om hvilken karakter de tenker å gi. Etter diskusjonen skal alle justere sine karakterer uavhengig av hverandre og uten at de andre fagutvalgsmedlemmene får vite det.

En åpenbar svakhet ved denne tilnærmingen er at fagutvalgsmedlemmene mister en god mulighet til å kalibrere seg mot de andre fagutvalgsmedlemmene. En slik åpen diskusjon rundt karakterene kunne også være ledet av administrasjonen, slik at man også får muligheten til å kalibrere karaktersettingen på tvers av ulike grupper av fagutvalgsmedlemmer. I både EU og i Norges forskningsråd diskuteres karaktersettingen åpent og panelene skal komme frem til en felles karakter for hvert av kriteriene som skal vurderes.

En annen ulempe ved vår tilnærming er at søknadens vurderingsgrunnlag og dermed også grunnlaget for tilbakemeldingen til søker, består av mange ulike vurderinger. Det er ofte stor uenighet mellom fagutvalgsmedlemmene og derfor får søkerne ofte motstridende tilbakemeldinger som er vanskelige å tolke og som kan skape frustrasjon.

Vi har likevel ikke valgt å legge opp til en slik konsensuskarakter fordi:

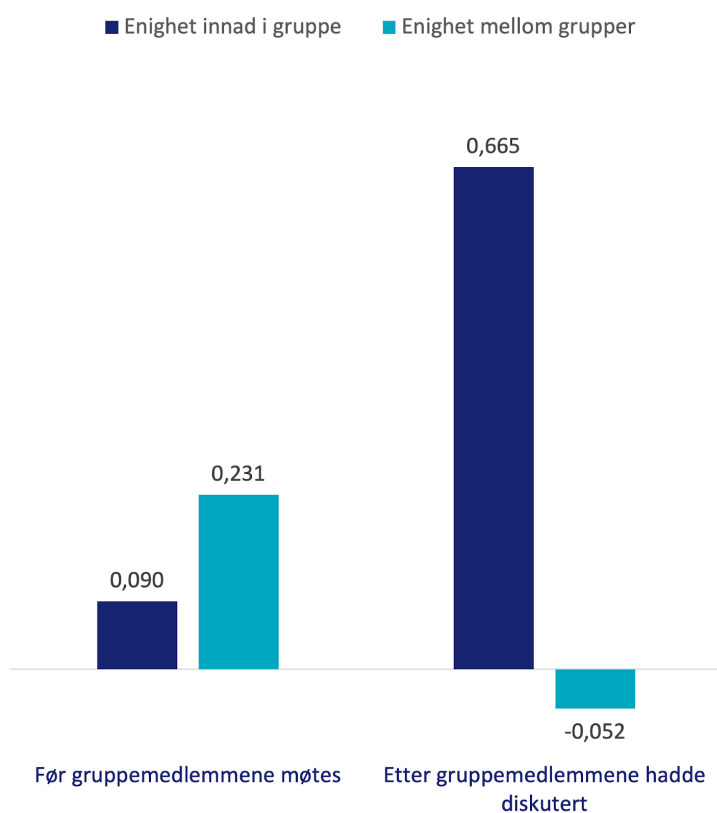
1. Det er uklart hvilken tilleggsverdi det har å bruke konsensuskarakter.
2. Vi vil begrense konformitetspress.
3. Vi vil begrense "første taler"-effekten og forankringseffekten.
4. Vi vil begrense sosial loffing.
5. Det er mulig å kalibrere karaktersettingen på andre måter.

4.10.3 Uklart hvilken tilleggsverdi det har å bruke konsensuskarakter

Å la gruppen komme fram til en felles karakter (konsensuskarakter) har den klare fordel at hver søknad får en enhetlig vurdering av søknaden, og ikke flere, potensielt sprikende vurderinger. Det er lettere å håndtere, både for oss som finansiør og for søker.

Vi mener imidlertid at det er svært uklart hvorvidt denne måten å vurdere søknadene på fører til en *bedre* vurdering av søknadene. Fører det egentlig til at man velger bedre prosjekter enn om man bare bruker gjennomsnittet fra de individuelle vurderingene? Slik vi ser det, virker svaret på dette å være at vi ikke vet.

Flere studier viser at eksperter har en klar tendens til å bli enige når de møtes til diskusjon, men at det ikke nødvendigvis betyr at de kommer nærmere en slags "sann" kvalitet ved søknaden (Forsythe et al. 2018; Pier et al. 2017). I en studie fra 2017 fikk en stor gruppe eksperter de samme søknadene. Så satte de en karakter uavhengig av hverandre, før de ble delt inn i ulike grupper som skulle diskutere søknadene. Som Figur 7 viser økte enigheten *innad* i gruppene, mens enigheten *mellom* gruppene gikk ned. Med andre ord kom gruppene som hadde lest de samme søknadene til *mer* forskjellige konklusjoner enn om man bare hadde brukt vurderingene de ga før de møttes i grupper.



Figur 7: Enighet i og mellom grupper målt ved Krippendorffs alfa, før og etter gruppemøte

Dette tyder på at det er andre ting enn egenskaper ved søknaden som forklarer den økte enigheten i gruppene. Vi mener slike tendenser er et godt argument for å være tilbakeholden med å lage beslutningsprosesser der ulike personer må bli enige.

4.10.2 Vi vil begrense konformitetspress

Mennesker er sosiale dyr og sosialpsykologisk forskning har ettertrykkelig vist at det kan være svært ubehagelig å bryte med det man oppfatter som gruppens meninger. Det samme gjelder det å være den som vurderer søknadene annerledes enn resten av gruppen. Vi ønsker ikke at dette konformitetspresset skal bli utslagsgivende. Av denne grunn ønsker vi at hver enkelt skal få muligheten til å si sitt (ved å sette en egen karakter), og derfor er også den individuelle karaktersettingen anonym overfor de andre fagutvalgsmedlemmene.

4.10.1 Vi vil begrense “første taler”-effekten og forankringseffekten

En ting som bidrar til å etablere “gruppens mening” og dermed også bidrar til konformitetspress (se over), er hva som blir sagt først. I en gjennomgang av forskningen relatert til gruppeprosesser i søknadsbehandling, skriver forfatterne at:

What the first speaker starts the discussion with is found to be decisive [...] The strong or weak points of the applicant mentioned first are strongly supported by other panelists (van Arensbergen et al. 2014).

Nært beslektet med dette er den såkalte forankringseffekten. Vi vet at tydelige utsagn om karakterer (*Jeg har gitt denne søknaden toppkarakter 7*) kan fungere som et forankringspunkt for de andre medlemmene og forskning tyder på at de da kan risikere å trekkes mot denne karakteren, uavhengig av om det er gode grunner til det eller ikke (Gilovich et al. 2002; van Arensbergen et al. 2014).

Det bør understrekes at dette alltid er en utfordring når beslutninger skal tas i grupper, enten man tillater karaktersnakk eller ikke. Styrker og svakheter skal også diskuteres og kan også virke ledende. Dette er også noe av årsaken til at vi har innført maler for fremlegg av fagutvalgsmedlemmenes vurderinger når disse skal presenteres for andre fagutvalgsmedlemmer i grupper eller plenum.

4.10.2 Vi vil begrense sosial loffing

Begrepet “sosial loffing” betegner det at “folk yter mindre i grupper enn de ville gjort individuelt” (Svartdal 2020). Vi tror at bruk av konsensuskarakter øker faren for at hvert enkelt fagutvalgsmedlem kjenner på mindre ansvar for å gjøre en grundig vurdering av søknaden. Bruk av konsensuskarakter *trenger* ikke å utløse sosial loffing og det er mulig å ansvarliggjøre hver enkelt ekspert på andre måter, men vi frykter at vissheten om at det til syvende og sist skal settes en konsensuskarakter og at den individuelle karakteren og vurderingen ikke blir stående, gjør at faren øker.

4.10.3 Mulig å kalibrere på andre måter

Et argument *for* å la fagutvalgsmedlemmene snakke om karakterene sine er som nevnt at de enklere vil få en felles forståelse av hvordan karakterskalaen skal brukes. Beskrivelsen av karakterskalaer gir alltid rom for tolkning og ulike fagutvalgsmedlemmer vil bruke den ulikt. Veldig stor variasjon i

tolkningen er uheldig og derfor er det rimelig å sørge for at fagutvalgsmedlemmene tolker skalaen så likt som mulig.

Vi tror imidlertid at dette kan gjøres tilfredsstillende på andre måter. Som for eksempel ved å gi fagutvalgsmedlemmene en oversikt over deres karaktersetting sammenliknet med de andre fagutvalgsmedlemmene (basert på den totale karaktersettingen, uten at det knyttes til enkeltsøknader). En slik løsning har vi testet i en egen studie (se dam.no/feedback) og basert på fagutvalgsmedlemmenes ønsker vil slike tilbakemeldinger bli fast praksis.

4.11 Administrativ fremfor faglig møteledelse

Vi benytter ansatte i administrasjonen og ikke fagutvalgsmedlemmer til å lede plenums- og gruppemøter. Det er kun i programmene Helse og Forskning vi bruker slike møter.

4.11.1 Frigjør fagutvalgsmedlemmenes kapasitet til primæroppgaven

Noe av årsaken til at vi gjør det slik, er at det å overlate ledelsen av gruppemøtene til administrasjonen gjør at fagutvalgsmedlemmene kan fokusere på sin primæroppgave, som er å vurdere søknadens kvalitet.

4.11.2 Administrasjonen har mest kompetanse på prosessene og rutinene knyttet til stiftelsens søknadsbehandling

Sekretariatet har som en av sine hovedoppgaver å legge til rette for god søknadsbehandling. Som rapporten viser, innebærer det også å holde seg oppdatert på hva forskningen sier om beslutningsprosessene som finner sted i gruppemøtene. Dette kombinert med sekretariatets inngående kjennskap til stiftelsens rutiner og vedtekter, gjør at vi mener sekretariatet er best skikket til å lede møtene.

4.12 Brukerrepresentanter i søknadsvurderingen

Brukermedvirkning står sentralt i Stiftelsen Dam, noe som er synlig både i tildelingene våre og vårt offentlige engasjement (Hesselberg et al. 2017; Svege and Hesselberg 2018; Ormstad et al. 2019; Hesselberg 2017) Vi er en stiftelse av og for de frivillige helseorganisasjonene i Norge, og nesten alle er enten brukerorganisasjoner eller har et sterkt fokus på brukere. Hos oss sikres brukermedvirkningen først og fremst ved at alle søknader må sendes fra en av våre godkjente søkerorganisasjoner. Det vil si at organisasjonen må bidra i utformingen av og stille seg bak søknaden og selv kan sikre at brukermedvirkningen ivaretas.

I tillegg krever vi brukermedvirkning i forskningsprosjekter som finansieres med våre midler.

Vi har imidlertid også sørget for at alle søknader i programmet Forskning og pilotprogrammet Utvikling som er aktuelle for innvilgelse blir vurdert av tre brukerrepresentanter uavhengig av hverandre. Disse vurderingene får betydning for hvilke søknader som innvilges. For det første vil

søknader som får dårlige vurderinger av alle tre brukerrepresentantene bli avslått. For det andre vil søknader som har fått samme karakter i kvalitetsvurderingen i fagutvalget bli skilt ved karakteren fra brukerrepresentantene. Det vil si at for alle søknader som har fått lik karakter i fagutvalget, vil brukerrepresentantenes vurderinger bli avgjørende⁴.

Å benytte brukerrepresentanter i søknadsvurderingen er fortsatt ikke vanlig blant forskningsfinansiører. Årsaken er antakelig at det er praktisk vanskelig å gjennomføre, at det ikke er sikkert at det faktisk gjør forskningen mer nyttig for sluttbrukerne og at det ikke er åpenbart hvem som skal representere brukerne (spesielt i så brede programmer som vi har).

Vi har likevel valgt å ha dette leddet i behandlingen av våre forskningssøknader av følgende to hovedårsaker.

4.12.1 Etisk begrunnet

Det viktigste argumentet er etisk begrunnet. Beslutninger om hvilke forskningsprosjekter som skal få midler, påvirker også sluttbrukerne av kunnskapen fra disse prosjektene. Vi mener at de som påvirkes av beslutningene også bør høres når beslutningene skal tas. Et sitat som brukes ofte og som tilskrives Valerie Billingham er *“Nothing about me without me”*. Å inkludere brukerrepresentanter i søknadsvurderingen som har det mål for øye å sikre at brukerne blir hørt i de ulike prosjektene, mener vi bidrar til dette.

4.12.2 Antatt økt nytteverdi av forskningen

I forbindelse med opprettelsen av Kavlifondets helseforskningsprogram (som Stiftelsen Dam driver), skrev vi et innlegg i Aftenposten om rasjonale bak å involvere brukere (Hesselberg et al. 2017). Innledningen kan tjene som bakgrunn også her:

“I 2009 publiserte Iain Chalmers og Paul Glasziou artikkelen «Avoidable waste in the production and reporting of research evidence» i det prestisjetunge tidsskriftet The Lancet (Chalmers and Glasziou 2009). Der dokumenterer de at anslagsvis 85 % av helseforskningen er bortkastet og at det kunne vært unngått.

Årsakene til dette høye tallet er mange og viktigst er at veldig mange studier aldri offentliggjør resultatene sine. Chalmers og Glasziou peker imidlertid også på at forskningen ikke alltid oppleves som nyttig av dem som er sluttbrukerne av kunnskapen.

For eksempel viste en studie av forskningsprioriteringer blant pasienter med kneartrose at bare ni prosent av pasientene ønsker mer forskning på medikamentell behandling, mens 80 prosent av de randomiserte kontrollerte studiene var studier på effekten av medikamentell behandling (Crowe et al. 2015).

Selv om det foreløpig er dårlig dokumentert at involveringen av brukere i vurderingen av forskningssøknader, fører til at forskningen blir nyttigere, mener vi at det er gode grunner til å anta at fokuset på brukernes stemmer øker sannsynligheten for at den blir det.

⁴ Detaljene i søknadsbehandlingen beskriver i *“Rutine for søknadsvurdering, Forskning”*: dam.no/soknadsvurdering-forskning

4.13 Beregnet rangering etter gruppebehandling

I vårt program Forskning avsluttes søknadsbehandlingen med seks ulike gruppemøter. I hvert av disse møtene samles fem fagutvalgsmedlemmer for å diskutere sine søknader (cirka 15 søknader i hver gruppe, cirka 90 søknader totalt). De setter individuelle karakterer og snittet av de fem karakterene, blir søknadens endelige total karakter. I tillegg vurderer tre brukerrepresentanter søknadene som er aktuelle for innvilgelse (se [Brukerrepresentanter i søknadsvurderingen](#)).

Deretter blir søknadene fra *alle* gruppene rangert etter denne total karakteren og brukerrepresentantenes vurderinger.

Det er antakelig på dette punktet vi skiller oss mest fra andre finansiører. Den vanlige fremgangsmåten dersom man har flere grupper av eksperter som bare vurderer en andel av søknadene, er at man har et avsluttende vurderingstrinn etter gruppebehandlingen. I dette trinnet møtes noen (for eksempel gruppelederne eller et styre) for å gå gjennom alle søknadene som er behandlet i de ulike gruppene, og for å fatte en endelig beslutning om hvilke søknader som skal innvilges eller avslås.

I Forskning bruker vi *ikke* dette trinnet og legger kun til grunn karaktersetningen fra hver av de seks gruppene. Hver søknad vil få fem uavhengige vurderinger og basert på denne beregner vi en snittkarakter som rundes av til halve karakterer. Denne danner grunnlaget for en rangering av alle søknadene fra alle gruppene. Det vil si at dersom du får snittkarakter 5,5 i gruppe én, så havner du foran en søknad som får 5,0 i gruppe fire.

En vanlig innvending er at dette kan slå uheldig ut dersom faggruppene har veldig ulik tilnærming til å sette karakterer. Kanskje er det for eksempel slik at noen grupper nesten bare bruker karakterene 6 og 7, mens andre nesten bare bruker karakterene 3 og 4. Våre analyser viser også at det er stor variasjon i karaktersetningen mellom fagutvalgsmedlemmene og annen forskning viser at dette ikke er uvanlig (Zupanc and Štrumbelj 2018).

Det er imidlertid noen gode argumenter for hvorfor dette problemet *ikke* løses ved å innføre et avsluttende vurderingstrinn som går gjennom hele søknadsbunken. De følger under, men grovt sett handler de om at:

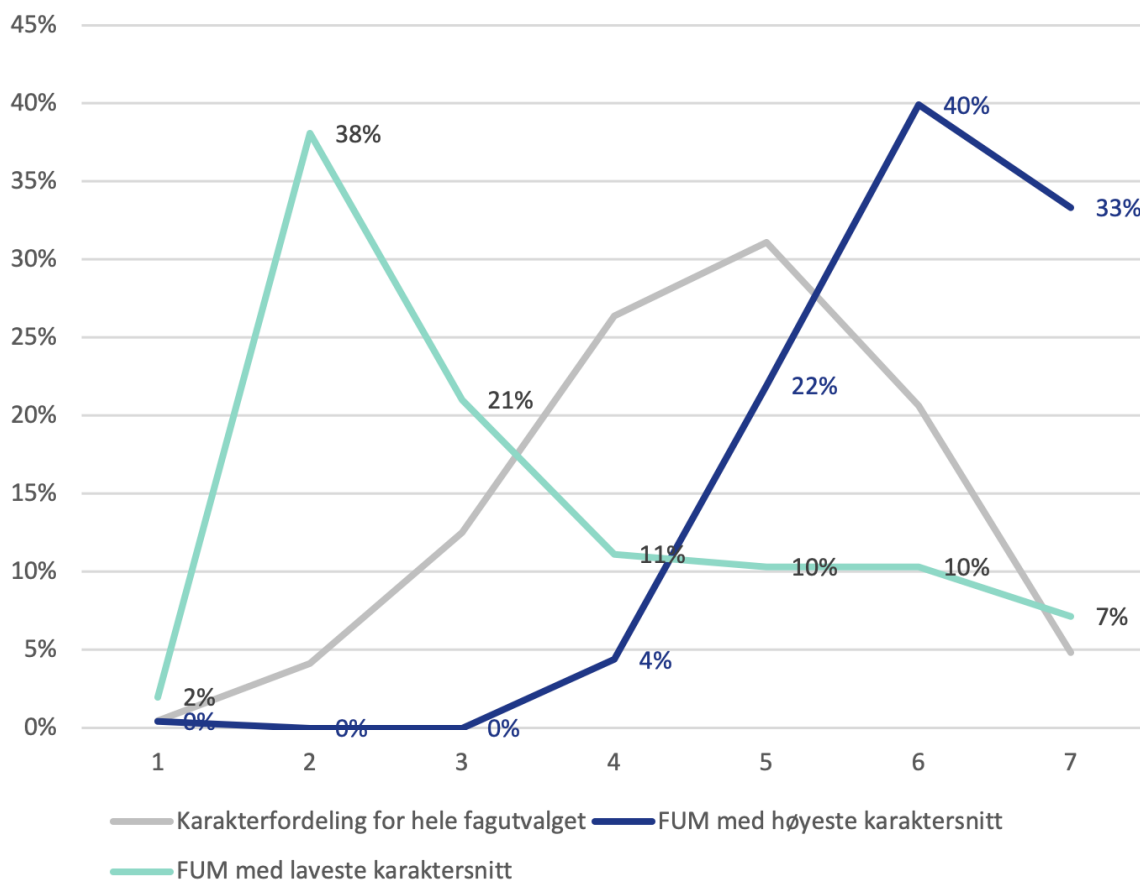
- *utfordringene ikke nødvendigvis blir mindre med et slikt avsluttende trinn.*
- *trinnet tilfører nye problemer.*

4.13.1 Fem uavhengige vurderinger jevner ut forskjellene

Det viktigste argumentet for vår gruppebehandling er at det er fem fagutvalgsmedlemmer som vurderer søknadene *uavhengig av hverandre* og at dette vil utjevne noen av forskjellene mellom fagutvalgsmedlemmene (som for eksempel at noen er "snille" og noen er "strenge").

I programmet Forskning i 2020 var 30 fagutvalgsmedlemmer involvert i vurderingen av de 361 skissesøknadene. Alle søknadene ble vurdert på en karakterskala fra 1 til 7 (se [Sjuddelt karakterskala](#)). Karaktersnittet for det "stengeste" fagutvalgsmedlemmet var på 3,5, mens det snilleste var på 6,0.

Figur 8 viser hvor ulikt de to utvalgsmedlemmene brukte karakterskalaen, sammenliknet med den gjennomsnittlige karakterfordelingen i hele fagutvalget.



Figur 8: Karakterfordeling i prekvalifiseringsrunden i Forskning 2020, totalt og for fagutvalgsmedlemmene med høyeste og laveste karaktersnitt.

Disse to utvalgsmedlemmene har vurdert ulike søknader og er dermed ikke direkte sammenlignbare. Begge skiller seg imidlertid også markant ut fra de utvalgsmedlemmene som har vurdert de samme søknadene som dem. De tilhører grupper med fire andre utvalgsmedlemmer som har vurdert akkurat de samme søknaden og når vi fjerner vurderingene gjort av de to fagutvalgsmedlemmene fra sine respektive grupper, blir karaktersnittene i de to gruppene svært like. 4,6 i gruppen til den “strenge” og 4,8 i gruppen til den “snille”.

Tabell 14 gir en oversikt over karakterspriket i hver av de seks gruppene. I disse gruppene vurderer de fem fagutvalgsmedlemmene akkurat de samme søknadene, men likevel er det stor avstand mellom utvalgsmedlemmene med laveste og høyeste snitt i hver gruppe. Spriket i det totale karaktersnittet mellom gruppene blir heldigvis og naturlig nok vesentlig mindre. Som tabellen viser har gruppen med lavest karaktersnitt et snitt på 4,4 og den med høyest snitt har 5,0. På dam.no/bedre-soknadsbehandling kan du se et boxplot med detaljert informasjon om karaktersettingen til hvert enkelt fagutvalgsmedlem.

Tabell 14: Antall søknader, karaktersnitt, spriket mellom laveste og høyeste snitt og antall søknader som gikk videre til runde 2 per gruppe i prekvalifiseringsrunden i Forskning 2020

Gruppe	Søknader	Totalt snitt	Laveste snitt	Høyeste snitt	Sprik	Videre til runde 2
1	63	4.4 ± 0.72	3,5 ± 1,51	5,1 ± 0,75	1,6	13 (21 %)
2	55	4.7 ± 0.65	4,2 ± 0,96	5,4 ± 0,69	1,1	19 (35 %)
3	60	5.0 ± 0.42	4,3 ± 0,71	6,0 ± 0,62	1,7	22 (37 %)
4	62	4.6 ± 0.51	4,5 ± 0,89	4,9 ± 1,43	0,4	12 (19 %)
5	62	4.5 ± 0.48	4,0 ± 0,81	5,4 ± 0,76	1,4	12 (19 %)
6	59	4.7 ± 0.58	4,2 ± 0,81	5,0 ± 0,74	0,8	20 (34 %)
Totalt	361	4.6 ± 0.60	3,5 ± 1,51	6,0 ± 0,62	2,5	98 (27 %)

± angir standardavviket. Karakterskalaen går fra 1 (lavest) til 7 (høyest). Totalt snitt er snittet for alle karakterene satt i de ulike gruppene. Laveste og høyeste snitt gjengir snittet til det fagutvalgsmedlemmet med det laveste/høyeste snittet i gruppen. "Sprik" angir differansen mellom laveste og høyeste snitt.

4.13.2 Umulig å si om det er karaktersettingen eller søknadsbunken som er skjevfordelt

Vi fordeler ikke søknadene tilfeldig til våre fagutvalgsmedlemmer. Søknadene fordeles til noen vi tror har kompetanse til å vurdere søknaden. Gruppene settes sammen for å gjenspeile søknadsbunken og derfor vil gruppene få ulike søknader. En gruppe kan for eksempel få mange søknader som omhandler kardiologi, mens en annen gruppe kanskje i hovedsak får søknader som omhandler psykisk helse. Dersom det viser seg at kardiologigruppen setter høyere karakterer enn psykisk helse-gruppen, vet vi ikke om det er fordi fagutvalgsmedlemmene der har fått bedre søknader eller om de bare er snillere. Dette problemet kommer ingen helt unna.

4.13.3 Små endringer av karakterer i plenumsbehandlingen i våre programmer

Å utfordre karaktersettingen til fem uavhengige fagpersoner som har satt seg grundig inn i en søknad er krevende. Plenum har ikke tid til å gå grundig inn i hver søknad og dette tror vi er grunnen til at det ofte skjer veldig få og små justeringer under slike plenumsbehandlinger. Under punktet "Mindre grupper av fagpersoner" kan du lese mer om våre erfaringer og analyser av dette. Vår antakelse er at det samme skjer i et slikt avsluttende trinn.

4.13.4 Plenum introduserer nye kilder til uønsket påvirkning

I delen "Hva vi bekymrer oss for" beskrives kilder til uønsket påvirkning i søknadsbehandlingen. Et eksempel på en uønsket skjevhet kan for eksempel være at menn kommer heldigere ut av søknadsbehandlingen, selv om søknadene ellers er helt like. Disse tendensene forsvinner ikke, bare fordi man møtes i et slikt plenum.

Som vist under “Ingen konsensuskarakter” er det også slik at enigheten som oppstår i slike møter ikke nødvendigvis handler om at man har kommet nærmere “sannheten” om søknaden, men at det like godt kan være et resultat av gruppeprosesser som kan gjøre vurderingen dårligere.

Slike faktorer utgjør allerede en potensiell utfordring i gruppemøtene hvor tre til fem fagutvalgsmedlemmer møtes for å diskutere styrker og svakheter ved søknadene de har vurdert, men innføringen av et avsluttende plenum gir slike faktorer nok en arena hvor de kan påvirke det endelige utfallet.

Vår holdning er derfor at, dersom det ikke er godt dokumentert at nye ledd fører til bedre søknadsbehandling, så skal vi ikke bruke dem. At det *virker* fornuftig å gjøre det eller at det er vanlig å gjøre det slik, bør ikke være tilstrekkelig.

4.13.5 Justeringer kan gjøres administrativt og på en forutsigbar måte

Et argument for et avsluttende plenum er at de kan korrigere åpenbare skjevheter mellom gruppene. Men *hvordan* skal de gjøre det? Det er ingen enkel oppgave å lage forutsigbare rutiner for det. Dessuten er det også mulig å gjøre slike korrigeringer rent administrativt. Med fem uavhengige karakterer for hver søknad kan vi både se på variasjonen mellom gruppene og mellom fagutvalgsmedlemmene.

Selv om vi aldri kan si om en gruppe er veldig “snille” eller bare tilfeldigvis har fått knallsterke søknader, så kan vi undersøke karaktersetningen til de enkelte utvalgsmedlemmene. I et tilfelle så vi for eksempel at et fagutvalgsmedlem nesten utelukkende brukte de beste karakterene (5, 6 og 7). Hvis dette er et gjentakende mønster, er det et problem. Søknader som havner hos dette fagutvalgsmedlemmet vil automatisk få økt sannsynligheten sin for innvilgelse. Slikt er det imidlertid mulig å korrigere for og noen finansierer gjør det rutinemessig. Vi har også gjort det i noen svært spesielle tilfeller.

Detaljene her er ikke viktige. I stedet er hovedpoenget at dette antakelig er den beste og mest forutsigbare måten å gjøre slike korrigeringer på og det er *ikke nødvendig å bruke plenum for å gjøre det*.

4.14 Tilbakemelding til søker

Stiftelsen Dam er ikke en offentlig institusjon og er ikke forpliktet til å begrunne sine avslag. Dette er heller ikke vanlig blant stiftelser. Vi har likevel ønsket å gi tilbakemelding til søker, med tanke på den innsatsen som nedlegges i søknadsskriving. En del søknader er støtteverdige, men når ikke opp på grunn av begrensninger i tilgjengelige midler. Andre vurderes ikke som gode nok. Denne dialogen med søkerne har blitt tatt godt imot og har blitt videreutviklet de senere årene.

Vi gir ikke en sammenfattet begrunnelse, men har i stedet valgt å dele tilnærmet hele beslutningsgrunnlaget med søker. Det vil si at søker får oversikt over karaktersettingen til fagutvalget og kommentarene fra fagutvalgsmedlemmene. Vi deler også navnene på dem som sitter i fagutvalget, men søkeren får ikke navnet på hvilke fagutvalgsmedlemmer som har vurdert den enkelte søknad.

Vi har valgt å gjøre det slik fordi vi mener:

1. åpenhet og transparens i slike prosesser både er riktig og utviklende.
2. tilbakemeldingene blir riktigere.
3. det ikke påvirker kvaliteten på tilbakemeldingene.
4. det går raskere.

4.14.1 Åpenhet

I mars 2020 vedtok stiftelsens styre egne retningslinjer for åpenhet i Stiftelsen Dam (se dam.no/retningslinjer-for-apenhet). Deling av informasjon om søknadsvurderingen er en sentral del av disse retningslinjene og i dem står det blant annet at *“Det skal være mulig som utenforstående på egenhånd å orientere seg om hvordan søknadene behandles og hva som fører til innvilgelser og avslag i de ulike programmene.”*

Vi gjør ikke dette bare fordi vi mener søkerne bør få innblikk i våre prosesser, men også fordi vi mener det er utviklende for oss. Innsyn og kritiske blikk fra primærbrukerne av våre systemer er avgjørende for at vi skal kunne videreutvikle oss på best mulig måte. Det er også noe av bakgrunnen for denne rapporten.

4.14.2 Riktigere tilbakemeldinger

Karakterene som fagutvalgene setter danner grunnlaget for våre beslutninger om innvilgelser og avslag. Derfor mener vi det er viktig at søkerne får tilgang til hvilke karakterer som er satt av de enkelte fagutvalgsmedlemmene og ikke bare får en gjennomsnittskarakter.

Resultatet av dette blir at søkerne i mange tilfeller vil se at både karakterene og kommentarene fra de ulike fagutvalgsmedlemmene spriker. Det kan være frustrerende og vanskelig å forstå hvordan man skal jobbe videre dersom man ønsker å forbedre søknaden.

Andre finansiører løser dette noen ganger ved at administrasjonen sammenfatter en tilbakemelding basert på kommentarene og karakterene. Vi mener dette skaper et feilaktig inntrykk av

søknadsbehandlingen. En slik sammenfatning maskerer den uenigheten som nesten alltid er tilstede og hindrer søker i å få et direkte innblikk i det faktiske grunnlaget for søknadsbehandlingen.

4.14.3 Forskning tyder ikke på at tilbakemeldingene blir dårligere

Det er lett å tenke seg å tenke seg to delvis motstridende utfall av å gå over til å dele fagutvalgmedlemmenes kommentarer til søknaden. På den ene siden kan man tenke seg at kvaliteten heves fordi fagutvalsmedlemmene legger mer i vurderingen og kommentarene når de vet at andre skal lese det. På den andre siden er det mulig at kvaliteten synker fordi fagutvalgmedlemmene modererer seg fordi de ikke ønsker å støtte søkeren.

Hos oss får søker som sagt ikke vite nøyaktig hvilke fagutvalgmedlemmer som har vurdert den enkelte søknaden, men de får vite hvem som sitter i fagutvalget. Dermed er vår løsning verken helt anonym eller helt åpen.

Oss bekjent finnes det ikke forskning som omhandler effekten av å være åpen om identiteten til eksperter som vurderer søknader. Det er imidlertid gjort forskning på dette for ekspertvurderinger (fagfellevurderinger) av vitenskapelige artikler, og selv om disse vurderingene er ulike, har de også klare likhetstrekk. I denne forskningen er det foreløpig ingen tydelige tegn til at det skjer mye med kvaliteten dersom man bruker åpen fagfellevurdering (van Rooyen et al. 2010). Det vil si at vi verken kan forvente bedre eller dårligere vurderinger.

4.14.4 Raskere

Da vi først begynte å gi tilbakemeldinger i programmet Helse i 2018, var det to personer fra sekretariatet som gikk gjennom fagutvalgmedlemmenes vurderinger og basert på dem sammenfattet en begrunnelse. Dette ble gjort for rundt 1.100 søknader i året og arbeidet tok anslagsvis to månedsverk.

I dag får søkerne en e-post som henter alle kommentarene til fagutvalgmedlemmene og i tillegg gir en oversikt over de gjennomsnittlige karakterene på de ulike vurderingskriteriene (se [Fire vurderingskriterier](#)). Innføringen av denne e-posten har medført noe utviklingskostnader, men nå som dette er gjort, går denne måten å gi tilbakemeldingene på helt automatisk. Det betyr også at søkerne kan få tilbakemeldingene i det øyeblikket søknadsbehandlingen er ferdig.

5 Veien videre

Til tross for omfattende endringer og tiltak i vår søknadsbehandling de siste årene, er forbedringspotensialet fortsatt tilstede.

Det er spesielt tre områder vi ønsker å jobbe videre med i de kommende årene:

- Redusere ressursbruk
- Etablere gode validitetsmål
- Støyreduksjon

5.1 Fokusområder framover

5.1.1 Redusere ressursbruk

Som beskrevet i innledningen under "[Hva vi bekymrer oss for](#)" er ressursbruken forbundet med søknadsbehandlingen en av de viktigste årsakene til endringene vi har gjort. Ressursbruken i stiftelsen er bare en liten del av dette. Tiden søkerne bruker på å skrive søknader utgjør den desidert største delen av de totale ressursene som brukes i våre søknadsprosesser.

Arbeidet med å redusere ressursbruken til både søker, fagutvalgsmedlemmene og sekretariatet vil fortsette også i de kommende årene. Forutsetningen er naturligvis at dette balanseres mot kvaliteten på søknadsbehandlingen. Enklere søknadsskjema gjør arbeidet lettere for søker, men det hjelper lite dersom skjemaet blir så enkelt at fagutvalgsmedlemmene mister grunnlaget for å gjennomføre vurderingen.

5.1.2 Øke reliabilitet

Vi kartlegger regelmessig hvor enige våre fagutvalgsmedlemmer er når de vurderer samme søknad. Programsjef Jan-Ole Hesselberg har skrevet om dette i sin spalte i Magasinet Forskningsetikk (Hesselberg 2020):

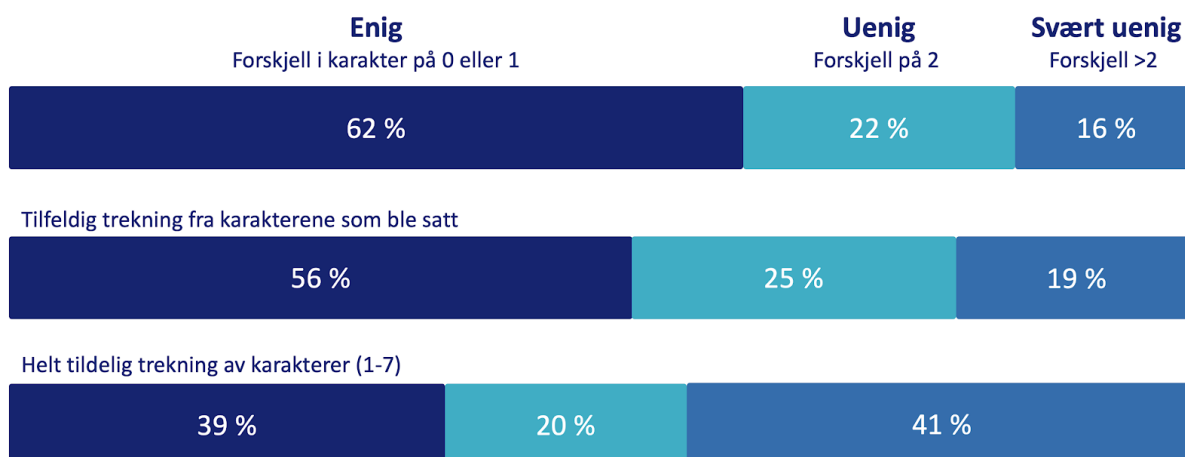
“Hver søknad som sendes til oss, vurderes av to til fem eksperter, uavhengig av hverandre. Selv om kriteriene de skal forholde seg til, er de samme for alle, må vi forvente at vurderingene vil sprike. Ekspertene er tross alt mennesker med ulike preferanser, og ikke alle leser søknadene på nøyaktig samme måte.

Men det finnes grader av uenighet. Vi har kartlagt enigheten mellom par av eksperter som vurderer samme søknad i et av våre programmer.

I seks av ti tilfeller var ekspertene «enige» om karakteren. I denne sammenhengen var det definert som en forskjell på maksimalt én karakter. Det virker jo bra, tenker du kanskje. Men hadde ekspertene trukket karakterene tilfeldig, ville fordelingen sett nesten helt lik ut.”

I artikkelen refereres det til våre analyser av 2018-utlysningen i programmet Forskning og Figur 9 gir et bilde av den beskrevne situasjonen. I 62 prosent av tilfellene var fagutvalgsmedlemmene enige om

karakteren, men en tilfeldig trekning som tar hensyn til at noen karakterer brukes oftere enn andre ville gitt en enighet på 56 prosent og helt tilfeldig trekning av karakterer ville gitt en enighet på 39 prosent.



Figur 9: Enighet i 2020-utlysningen i programmet Forskning, sammenlignet med enighet ved to måter å trekke karakterer tilfeldig på.

Det er viktig å understreke at det ikke nødvendigvis er noe galt med lav enighet. Om man ønsker at en søknad skal vurderes av eksperter med ulike perspektiver, er det rimelig å forvente lav enighet. Noen forskere har av den grunn tatt til orde for at det til og med kan være ønskelig med lav enighet (Bailar and Patterson 1985), nettopp fordi det kan være et tegn på at man har valgt eksperter med ulike perspektiver. Det er heller ikke slik at høy enighet er et kvalitetstegn i seg selv. Hadde man gitt fagutvalgsmedlemmene beskjed å gi søknadene toppkarakter dersom bokstaven A dukket opp i søknaden, ville alle søknadene fått samme karakter og enigheten ville blitt hundre prosent, uten at det ville ha sagt spesielt mye om søknadens kvalitet.

Vi mener likevel det er grunn til å bekymre seg over for lav enighet. Problemet med for lav enighet er at vurderingsprosessen blir inkonsistent og ustabil. Man kan ikke ha lav enighet og samtidig klare å plukke ut de beste prosjektene på en systematisk måte. Man kan heller ikke vite hvorvidt den lave enigheten er et resultat av at ekspertene bringer med seg ulike og nyttige perspektiver, eller om det bare er et tegn på støy i beslutningsprosessene (se [Støy](#)). Uansett vil lav enighet medføre at samme søknad kan sendes flere ganger, men få ulikt utfall.

Det finnes egne mål på enighet som tar hensyn til at man kan bli enige ved ren tilfeldighet. Vi bruker i hovedsak målet "intra-klasssekorrelasjon" (Hesselberg and Svege 2018). Dette målet sier noe om hvor stor andel av variasjonen i vurderingen av søknadene (altså karakterene) som kan tilskrives kvaliteter ved søknaden, og ikke andre faktorer. I eksempelet over var denne på 0,21, det vil si at 21 prosent av variasjonen i karakterer kunne tilskrives søknaden. Det er lavt.

Tabell 15 gir en oversikt over enigheten i vurderingene som ble gjennomført i stiftelsens programmer i fjor. ICC-verdiene strekker seg opp til 1,0 og det er vanlig å beskrive verdiene på følgende måte (Koo and Li 2016):

- < 0.50 Lav reliabilitet (poor reliability)
- 0.50-0.75 Moderat reliabilitet (moderate reliability)
- 0.75-0.90 God reliabilitet (good reliability)
- > 0.90 Svært god reliabilitet (excellent reliability)

Tabell 15: Enigheten i vurderingene, målt ved intraklassekorrelasjon (ICC) og gjennomsnittlig avvik karakterer

Program	Antall søknader	Antall vurderinger per søknad (k)	ICC (1,k) snitt	Gjennomsnittlig avvik i karakter*
Helse høst 2019	431	3	0,32 (n=372)	1,15 (n=1.165)
Ekspress høst 2019	589	2	0,45 (n=515)	1,35 (n=519)
Forskning 2020 (prekvalifisering)	366	5	0,55 (n=193)	1,10 (n=3.180)
Utvikling 2020 (prekvalifisering)	126	4	0,63 (n=118)	1,09 (n=772)
Kavlifondet 2019 (prekvalifisering)	24	4	0,82 (n=24)	1,01 (n=144)

* Baserer seg på søknadens total karakter (som er et gjennomsnitt av karakterene på de ulike kriteriene). Skalaen strekker seg fra 1 til 7. Differansene baserer seg på forskjeller mellom par av fagutvalgsmedlemmer som vurderer samme søknad. Ved helt tilfeldig trekning av karakterer, ville det gjennomsnittlige avviket blitt 2,3.

5.1.3 Etablere mål på virkning

Vi kan redusere redusere ressursbruken og sikre perfekt reliabilitet, men likevel feile totalt i oppgaven vi er satt til å gjøre: Å velge de beste prosjektene.

Å måle gyldigheten – eller validiteten – av vurderinger som våre, er en nøtt finansierer over hele verden forsøker å knekke (Guthrie et al. 2017). I en studie fra 2016 så forskerne på sammenhengen mellom fagutvalgsmedlemmers vurderinger og hvordan det gikk med prosjektene etter at de fikk finansiering (målt ved hvor ofte de vitenskapelige artiklene ble sitert). Utvalgsmedlemmenes vurderinger hang svært dårlig sammen med hvordan det gikk med prosjektene (Fang et al. 2016).

Selv hvis vi holder oss til vårt mest ensartede program, Forskning, støtter vi på utfordringer dersom vi ønsker å finne ut av om ekspertene våre faktisk velger de beste forskningsprosjektene. For hva er et “godt forskningsprosjekt”?

Er det antallet siteringer? Patenter? Retweets? Nyhetsoppslag? Endring av retningslinjer for behandling?

Virkningen av forskning kan måles på et utall måter som alle har store mangler. Og det blir enda mer utfordrende med de andre programmene våre. I Helse søkes det om alt fra 5.000 kroner til å lage en

gapahuk til tre millioner kroner i delfinansiering til et TV-program. Hvordan måle virkningen av så ulike prosjekter?

En forutsetning for slike målinger er i tillegg at de kan automatiseres eller er enkle å gjennomføre. Vi kunne ha sendt ut tre eksperter til hvert prosjekt og latt dem bruke fire uker på å kartlegge alt prosjektet oppnådde, men med rundt tusen prosjekter som avsluttes hver år, ville det ha resultert i 261 årsverk i slike evalueringer hvert år. Til tross for at det ville ha gitt oss interessant informasjon, vil noen bakstreberske mennesker antakelig mene at "vinningen går opp i spinningen".

Vi er foreløpig usikre på hvilke mål vi skal og bør bruke, men har kommet til et punkt i utviklingen av programmene våre der vi skal begynne å kartlegge mulighetene grundigere.

6 Hva alle som driver søknadsbehandling bør gjøre

Stiftelsen Dam deler ut hundretalls millioner kroner og både kan og bør bruke prosesser som er altfor ressurskrevende for de fleste som driver søknadsbehandling. Basert på vårt forsknings- og kvalitetssikringsarbeid, har vi imidlertid tanker om hva som bør utgjøre et minimum når man skal lage en søknadsprosess.

Vi mener alle bør:

- ... bruke etablerte eller godt designede kategorier i søknadsskjemaet.
- ... la minst to personer vurdere søknadene, uavhengig av hverandre.
- ... sette et tall på vurderingene.

6.1 Bruke etablerte eller godt designede kategorier

Tenk grundig på hvilke data som skal inn i søknadsskjemaet.

De fleste som deler ut penger har behov for å si noe om hvem søkerne er, hva de søker om og hva det deles ut midler til. Skal man besvare disse spørsmålene på en god måte, starter arbeidet med søknadsskjemaet. Hvis ikke informasjonen du trenger går inn i søknaden, er det heller ikke mulig å hente den ut. Og med mindre du har spisskompetanse på kvalitative studier og mye tid til overs, må du sørge for at informasjonen kategoriseres.

Ikke bare be søkeren beskrive seg selv og sin kompetanse, sørg for å lage avkrysningsmuligheter for det som er viktig for deg å vite noe om. Og bruk kategorier som også brukes av andre. Ingen kan stoppe deg i å lage egne definisjoner av utdanningsnivåer, men om du bruker [Norsk standard for utdanningsgruppering](#) åpner det seg mange muligheter for å sammenligne dine søkere med andre grupper. Legg inn et obligatorisk stopp på ssb.no/klass. Der finner du offisielle kategorier for alt fra yrker og lønn til skipstyper og fødselsvekt. Kanskje finner du til og med noen kategorier du ikke visste at du hadde bruk for.

Tenk også på om det er noe det er spesielt viktig å få tilbakemelding på fra søkeren. Vi spør for eksempel alltid om hvor mye tid søkeren anslår å ha brukt på søknaden og vi har hentet spørsmålet fra en større undersøkelse i Norges forskningsråd. Resultatet er at vi både kan sammenligne tiden våre søkere bruker med noe annet, og det gir oss muligheten til å se om endringene vi gjør fører til mindre tidsbruk hos søkerne. Slik kunne vi dokumentere en reduksjon på 42 prosent i den gjennomsnittlige tidsbruken i programmet Forskning, etter at vi hadde gjort store endringer før utlysningen i 2020 (se [Prekvalifiseringsrunde](#)).

6.2 La minst to personer vurdere, uavhengig av hverandre

Mange som behandler søknader bruker bare én person i vurderingen av dem. Vi vet at det øker kostnadene å la flere vurdere hver søknad, men mener likevel det bør prioriteres. Årsaken er at man mister viktig informasjon om både gyldigheten ([validiteten](#)) av og stabiliteten ([reliabiliteten](#)) til vurderingsprosessen. Hvis to eller flere personer vurderer hver søknad, vil man kunne måle hvor enige de er. Dersom de er veldig uenige og den såkalte [interrater-reliabiliteten](#) er lav, har man et stort

problem. Hvis ikke vurderingene er konsistente og stabile, kan man heller ikke på en systematisk måte klare å velge de beste søknadene (uavhengig av hvordan man definerer det).

Alle ville ha kvittet seg med et termometer som i tre målinger på rad viser 37, 41 og 39. Paradokset er at vi fra forskningen vet at dette ofte skjer i søknadsvurderinger, men at veldig få faktisk undersøker om dette er tilfellet.

Dersom man ikke lar flere søknadene få flere vurderinger, er det ikke mulig å vurdere graden av enighet. Det er ikke nødvendig å innføre flere vurderinger som en fast prosedyre, men i det minste bør det gjøres for et utvalg søknader, slik at man får en indikasjon på denne reliabiliteten. Det finnes flere statistiske metoder man kan bruke, men κ ⁵ er blant de enkleste og mest brukte. Kappa måler hvor stor andel enighet man har, *utover det man kan forvente ved ren tilfeldighet*.

En forutsetning dersom man skal gjennomføre slike analyser, er at de som vurderer samme søknad faktisk gjør det helt uavhengig av hverandre. Aller helst bør ikke de som vurderer søknadene vite hvem de andre er. Dersom dette ikke lar seg gjøre, er det et minimum at de ikke snakker sammen og ikke får se hverandres vurderinger. Årsaken til dette er at vi påvirkes av å få vite hvordan andre tenker og flere studier viser at det fører til kunstig høy enighet (Hesselberg and Svege 2018).

6.3 Sett et tall på vurderingene

Menneskelige vurderinger er kvalitative i sin natur og i mange tilfeller føles det unaturlig å skulle sette et tall på dem – som man gjør når søknader gis en karakter. Samtidig kan ikke alle søknadene innvilges og derfor må de må de settes opp mot hverandre.

Det er i hovedsak to måter å gjøre det på:

1. Sette en karakter
2. Rangere søknadene

Å “sette en karakter” betyr ikke annet enn å gi en verdi til søknaden. Vi bruker samme sjudelte skala som Norges forskningsråd bruker, men en enkel todelt skala som “støtteverdig” og “ikke støtteverdig” kan også gjøre nytten. Hvis ikke alle søknadene vurderes av de samme personene, er det best å bruke karakterer. Dersom alle som bruker skalaen har samme forståelse av den, har karakterer den fordelen at de er sammenlignbare på tvers av dem som vurderer søknaden og på tvers av ulike utlysninger. Slik fungerer karakterene i videregående skole. En femmer ved Nittedal videregående skole i 2018 sidestilles med en femmer ved Frogn videregående skole i 2020.

Utfordringen er naturligvis at verken sensorene i videregående skoler eller våre eksperter har *nøyaktig* samme forståelse av skalaen. Vi ser store forskjeller blant våre eksperter. Noen er veldig “snille” og gir nesten bare toppkarakterene 6 og 7, mens andre eksperter (som leser nøyaktig de samme søknadene), nesten bare bruker bunnkarakterene 1 og 2. Det gjør at det blir vanskelig å sammenlikne karakterene som settes av ulike eksperter på ulike søknader.

⁵ Fra snl.no: “(Kappa) brukes i forskning som betegnelse for en reliabilitetskoeffisient (reliabilitet), det vil si et tall som sier noe om samsvar mellom to tester. Kappa-verdien vil variere mellom 0 og 1. En kappaverdi <0,4 vil vanligvis indikere så lav samstemmighet at vurderingene i liten grad kan brukes. En kappaverdi >0,7 regnes som god.”

Dersom alle søknadene vurderes av de samme ekspertene, kan en løsning på disse problemene være å la ekspertene *rangere* søknadene. Da eliminerer man problemet med snille og strenge eksperter. Utfordringen er at man ikke vet hva ekspertene mener om søknadene. Det kan for eksempel godt være at en ekspert mener at alle søknadene er svake og at ingen bør innvilges. Dette vil ikke bli fanget opp i en rangering alene. Av den grunn brukes rangering ofte i kombinasjon med karaktersetting (som for eksempel "Støtteverdig", "Grenseland" og "Ikke støtteverdig").

6.4 Mange muligheter

Gjennomfører man de foreslåtte tiltakene, åpner det seg mange muligheter. Ikke bare får man mer informasjon å basere utvelgelsen av søknadene på. Det vil også bli vesentlig enklere å få svar på spørsmål som er viktige både i rapporteringen av hvilke søknader som innvilges og hvordan det går med dem. Og det gir helt andre muligheter i kvalitetssikringen av søknadsbehandlingen. Er det for eksempel en sammenheng mellom alder på søker og karakter/rangering? Eller søkerens kjønn og karakter? Er kanskje yngre eksperter strengere enn eldre? Har fagfeltet til eksperten noe å si for karakterene og hvor stor uenigheten blir?

De tre nevnte tiltakene åpner for å kunne svare på slike spørsmål. Spørsmål som det i mange tilfeller er helt nødvendig å få svar på om man skal kunne si at søknadsbehandlingen skjer på en forsvarlig måte.

Og det trenger ikke koste mye. Å få en ekstra vurdering for søknader som vanligvis bare får én vurdering, vil naturlig nok koste litt, men utover det er kostnadene små. Søknadsskjemaer må bygges og vurderingene må gjøres, enten man gjør det på en god eller dårlig måte, og i dag finnes gratis datasystemer og -programmer som dekker de grunnleggende behovene for både søknadsskjemaer, vurderingsskjemaer og analyser av dataene.

Vi tror alle som driver søknadsbehandling har nytte av at vi tenker likt om datainnsamling og analyse av søknadsprosessene, og at vi er åpne om valgene vi tar og utfordringene vi står i. Derfor har vi laget denne rapporten.

7 Referanser

van Arensbergen, P., van der Weijden, I. and van den Besselaar, P. 2014. The selection of talent as a group process. A literature review on the social dynamics of decision making in grant panels. *Research Evaluation* 23(4), pp. 298–311.

Ariely, D., Loewenstein, G. and Prelec, D. 2003. “coherent arbitrariness”: stable demand curves without stable preferences. *The quarterly journal of economics* 118(1), pp. 73–106.

Asch, S.E. 1956. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied* 70(9), pp. 1–70.

Bailar, J.C. and Patterson, K. 1985. The need for a research agenda. *The New England Journal of Medicine* 312(10), pp. 654–657.

Bornmann, L., Mutz, R. and Daniel, H. 2007. Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics* 1(3), pp. 226–238.

Brezis, E. and Birukou, A. 2019. Arbitrariness in the Peer Review Process. In: PEERE.

Busetta, G., Fiorillo, F. and Visalli, E. 2013. Searching for a job is a beauty contest. *SSRN Electronic Journal*.

Carpenter, A.S., Sullivan, J.H., Deshmukh, A., Glisson, S.R. and Gallo, S.A. 2015. A retrospective analysis of the effect of discussion in teleconference and face-to-face scientific peer-review panels. *BMJ Open* 5(9), p. e009138.

Chalmers, I. and Glasziou, P. 2009. Avoidable waste in the production and reporting of research evidence. *The Lancet* 374(9683), pp. 86–89.

Chen, D.L., Moskowitz, T.J. and Shue, K. 2016. Decision Making Under the Gambler’s Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires. *The quarterly journal of economics* 131(3), pp. 1181–1242.

Chernev, A., Böckenholt, U. and Goodman, J. 2015. Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology* 25(2), pp. 333–358.

Cicchetti, D.V. 1991. The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences* 14(01), pp. 119–135.

Crowe, S., Fenton, M., Hall, M., Cowan, K. and Chalmers, I. 2015. Patients’, clinicians’ and the research communities’ priorities for treatment research: there is an important mismatch. *Research Involvement and Engagement* 1, p. 2.

Danziger, S., Levav, J. and Avnaim-Pesso, L. 2011. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences of the United States of America* 108(17), pp. 6889–6892.

Dhar, R. and Nowlis, S.M. 1999. The effect of time pressure on consumer choice deferral. *The Journal of consumer research* 25(4), pp. 369–384.

Englich, B., Mussweiler, T. and Strack, F. 2006. Playing dice with criminal sentences: the influence of irrelevant anchors on experts’ judicial decision making. *Personality and social psychology bulletin* 32(2), pp. 188–200.

European Science Foundation 2011. *European Review Guide: Integrating Policies and Practices into Coherent Procedures*. Strasbourg.

European Union 2019. Eligibility and evaluation criteria - H2020 Online Manual [Online]. Available at: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/grants/from-evaluation-to-grant-signature/evaluation-of-proposals/elig_eval_criteria_en.htm [Accessed: 22 May 2019].

Fang, F.C., Bowen, A. and Casadevall, A. 2016. NIH peer review percentile scores are poorly predictive of grant productivity. *eLife* 5.

Forscher, P.S., Cox, W.T.L., Brauer, M. and Devine, P.G. 2019. Little race or gender bias in an experiment of initial review of NIH R01 grant proposals. *Nature human behaviour* 3(3), pp. 257–264.

Forsythe, L.P., Frank, L.B., Tafari, A.T., et al. 2018. Unique review criteria and patient and stakeholder reviewers: analysis of pcori's approach to research funding. *Value in Health* 21(10), pp. 1152–1160.

Gilovich, T., Griffin, D.W. and Kahneman, D. 2002. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, U.K.; New York: Cambridge University Press.

Graves, N., Barnett, A.G. and Clarke, P. 2011. Funding grant proposals for scientific research: retrospective analysis of scores by members of grant review panel. *BMJ (Clinical Research Ed.)* 343, p. d4797.

Guthrie, S., Ghiga, I. and Wooding, S. 2017. What do we know about grant peer review in the health sciences? [version 1; peer review: 1 approved, 1 approved with reservations]. *F1000Research* 6, p. 1335.

Hardcastle, J. 2019. Old-fashioned peer review is still seen as the best way to allocate grants, but reviewers deserve greater recognition [Online]. Available at: <https://blogs.lse.ac.uk/impactofsocialsciences/2019/12/10/old-fashioned-peer-review-is-still-seen-as-the-best-way-to-allocate-grants-but-reviewers-deserve-greater-recognition/> [Accessed: 22 February 2020].

Harrison, A.A. 1977. Mere Exposure. In: *Advances in experimental social psychology volume 10*. Advances in experimental social psychology. Elsevier, pp. 39–83.

Helmer, M., Schottdorf, M., Neef, A. and Battaglia, D. 2017. Gender bias in scholarly peer review. *eLife* 6.

Hesselberg, J.-O. 2017. Brukermedvirkning vektlegges. *Dagens Medisin*. Available at: <https://www.dagensmedisin.no/artikler/2017/03/25/brukermedvirkning-vektlegges/> [Accessed: 14 May 2020].

Hesselberg, J.-O. 2020. Uenige eksperter. *Magasinet Forskningsetikk*. Available at: <https://www.etikkom.no/Aktuelt/Fagbladet-Forskningsetikk/arkiv/2020/2020-1/arlign-talt/> [Accessed: 9 May 2020].

Hesselberg, J.-O. and Svege, I. 2018. Uenige sensorer og jakten på forskningsmidler [Online]. Available at: <https://forskning.no/forskningsfinansiering-kronikk-skole-og-utdanning/uenige-sensorer-og-jakten-pa-forskningsmidler/1213215> [Accessed: 14 May 2020].

Hesselberg, J.-O., Svege, I. and Mjøs, A. 2017. Hvordan unngå bortkastet helseforskning? *Aftenposten*. Available at:

<https://www.aftenposten.no/meninger/debatt/i/Aj1GE/hvordan-unngaa-bortkastet-helseforskning-ja-an-ole-hesselberg-ida-svege-og-aksel-mjoes> [Accessed: 12 May 2020].

Hodgson, C. 1995. Evaluation of cardiovascular grant-in-aid applications by peer review: influence of internal and external reviewers and committees. *The Canadian Journal of Cardiology* 11(10), pp. 864–868.

Jayasinghe, U.W., Marsh, H.W. and Bond, N. 2006. A new reader trial approach to peer review in funding research grants: An Australian experiment. *Scientometrics* 69(3), pp. 591–606.

Jeppesen, L.B. 2007. Getting Unusual Suspects to Solve R. 85(5).

Johnson, E.J. and Tversky, A. 1983. Affect, generalization, and the perception of risk. *Journal of personality and social psychology* 45(1), pp. 20–31.

Kahneman, D., Rosenfield, A.M., Gandhi, L. and Blaser, T. 2016. Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making. *Harvard Business Review*. Available at: <https://hbr.org/2016/10/noise> [Accessed: 24 March 2020].

Koo, T.K. and Li, M.Y. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine* 15(2), pp. 155–163.

Kramer, L.A. and Weber, J.M. 2012. This is your portfolio on winter. *Social psychological and personality science* 3(2), pp. 193–199.

Langfeldt, L. 2002. Decision-making in expert panels evaluating research: Constraints, processes and bias. Doctoral dissertation. NIFU – Norsk institutt for studier av forskning og utdanning.

Lorenz, J., Rauhut, H., Schweitzer, F. and Helbing, D. 2011. How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences of the United States of America* 108(22), pp. 9020–9025.

Mannix, E. and Neale, M.A. 2005. What Differences Make a Difference? The Promise and Reality of Diverse Teams in Organizations. *Psychological science in the public interest: a journal of the American Psychological Society* 6(2), pp. 31–55.

Marsh, H.W., Bond, N.W. and Jayasinghe, U.W. 2007. Peer review process: Assessments by applicant-nominated referees are biased, inflated, unreliable and invalid. *Australian psychologist* 42(1), pp. 33–38.

Marsh, H.W., Jayasinghe, U.W. and Bond, N.W. 2008. Improving the peer-review process for grant applications: reliability, validity, bias, and generalizability. *The American Psychologist* 63(3), pp. 160–168.

Mayo, N.E., Brophy, J., Goldberg, M.S., et al. 2006. Peering at peer review revealed high degree of chance associated with funding of grant applications. *Journal of Clinical Epidemiology* 59(8), pp. 842–848.

Midtbøen, A.H. and Rogstad, J. 2012. *Diskrimineringens omfang og årsaker: Etniske minoriteters tilgang til norsk arbeidsliv*. Institutt for samfunnsforskning.

Morgan, B., Yu, L.-M., Solomon, T. and Ziebland, S. 2020. Assessing health research grant applications: A retrospective comparative review of a one-stage versus a two-stage application assessment process. *Plos One* 15(3), p. e0230118.

Mrdjenovich, A.J. 2020. Authors Disclosing Their Theistic Orientation in Journal Articles on Religion

and Health? Infrequent, Informal, and Mostly Inconsistent with Conflict of Interest. *Journal of religion and health*.

Mutz, R., Bornmann, L. and Daniel, H.-D. 2012. Heterogeneity of inter-rater reliabilities of grant peer reviews and its determinants: a general estimating equations approach. *Plos One* 7(10), p. e48509.

Nasjonalt program for klinisk behandlingsforskning i spesialisthelsetjenesten 2019. Søknadsvurdering – Program for klinisk behandlingsforskning [Online]. Available at: <http://kliniskforskning.rhf-forsk.org/utlysning/vurdering/> [Accessed: 22 May 2019].

National Institutes of Health 2009. NOT-OD-09-024: Enhancing Peer Review: The NIH Announces New Scoring Procedures for Evaluation of Research Applications Received for Potential FY2010 Funding [Online]. Available at: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-09-024.html> [Accessed: 22 May 2019].

National Institute of Health Research 2019. Reviewer Assessment Form: Guidance for providing a review [Online]. Available at: <https://www.nihr.ac.uk/funding-and-support/documents/reviewers-for-reviewing/professional-reviewer-guidance-notes-EME.pdf> [Accessed: 22 May 2019].

Norwegian Cancer Society 2019. Application Guidelines Open Call 2019 [Online]. Available at: <https://nettsoknad.kreftforeningen.no/> [Accessed: 22 May 2019].

Ormstad, H., Svege, I. and Jamtvedt, G. 2019. Bortkastet forskning. *Dagsavisen*. Available at: <https://www.dagsavisen.no/debatt/bortkastet-forskning-1.1597515> [Accessed: 14 May 2020].

Oxley, K. 2016. *Søknadsbehandling i Forskningsrådet*. Internal report at The Norwegian Research Council: unpublished.

Pier, E.L., Raclaw, J., Kaatz, A., et al. 2017. “Your comments are meaner than your score”: score calibration talk influences intra- and inter-panel variability during scientific grant peer review. *Research Evaluation* 26(1), pp. 1–14.

Porter, A.L. and Rossini, F.A. 1985. Peer review of interdisciplinary research proposals. *Science, Technology, & Human Values* 10(3), pp. 33–38.

Ramberg, I. 2017. *Tids- og ressursbruk for søkning til Norges forskningsråd i 2016*. NIFU.

Rieskamp, J. and Hoffrage, U. 1999. When do people use simple heuristics, and how can we tell? In: Gigerenzer, G. and Todd, P. M. eds. *Evolution and cognition. Simple heuristics that make us smart*. Oxford University Press, pp. 141–167.

Rom, M.C. and Musgrave, P. 2014. Political outcome bias in grading: identifying problems and suggesting solutions. *Journal of Political Science Education* 10(2), pp. 136–154.

van Rooyen, S., Delamothe, T. and Evans, S.J.W. 2010. Effect on peer review of telling reviewers that their signed reviews might be posted on the web: randomised controlled trial. *BMJ (Clinical Research Ed.)* 341, p. c5729.

Saunders, E.M. 1993. Stock Prices and Wall Street Weather on JSTOR. *The American economic review*.

Shailes, S. 2017. To fund or not to fund? *eLife* 6.

Slovic, P. and Peters, E. 2006. Risk Perception and Affect. *Current Directions in Psychological Science*

15(6), pp. 322–325.

Smith, R. and Blazeby, J. 2018. Why religious belief should be declared as a competing interest. *BMJ (Clinical Research Ed.)* 361, p. k1456.

Snell, R.R. 2015. Menage a quoi? Optimal number of peer reviewers. *Plos One* 10(4), p. e0120838.

Svartdal, F. 2020. Sosial loffing [Online]. Available at: https://snl.no/sosial_loffing [Accessed: 16 June 2020].

Svege, I. and Hesselberg, J.-O. 2018. Mengden bortkastet forskning må reduseres. *Morgenbladet*. Available at: <https://morgenbladet.no/ideer/2018/02/mengden-bortkastet-forskning-ma-reduseres> [Accessed: 14 May 2020].

Teplitskiy, M., Acuna, D., Elamrani-Raoult, A., K rding, K. and Evans, J. 2018. The sociology of scientific validity: How professional networks shape judgement in peer review. *Research Policy* 47(9), pp. 1825–1841.

Teplitskiy, M., Ranu, H., Grey, G., Menietti, M., Guinan, E. and Lakhani, K.R. 2019. Do Experts Listen to Other Experts? Field Experimental Evidence from Scientific Peer Review. *HBS Working Paper Series*.

Thorngate, W. Mining the archives: Analyses of CIHR research grant adjudications.

Travis, G.D.L. and Collins, H.M. 1991. New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System. *Science, technology & human values* 16(3), pp. 322–341.

Tuk, M.A., Trampe, D. and Warlop, L. 2010. Inhibitory spillover: increased urination urgency facilitates impulse control in unrelated domains. *SSRN Electronic Journal*.

Tversky, A. and Kahneman, D. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185(4157), pp. 1124–1131.

Wilson, T.D., Houston, C.E., Etling, K.M. and Brekke, N. 1996. A new look at anchoring effects: basic anchoring and its antecedents. *Journal of Experimental Psychology. General* 125(4), pp. 387–402.

Witteman, H.O., Hendricks, M., Straus, S. and Tannenbaum, C. 2019. Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *The Lancet* 393(10171), pp. 531–540.

Zupanc, K. and Štrumbelj, E. 2018. A Bayesian hierarchical latent trait model for estimating rater bias and reliability in large-scale performance assessment. *Plos One* 13(4), p. e0195297.

Rapport 2020:1

Utgitt av Stiftelsen Dam
Adresse: Akersgata 28, 0158 Oslo, Norge

Kontaktperson: Programsjef Jan-Ole Hesselberg, joh@dam.no

Bilde [Max Rovensky](#) på [Unsplash](#)

ISBN 978-82-691947-0-8

Copyright Stiftelsen Dam: [CC BY-NC 4.0](#)

www.dam.no

Stiftelsen Dam

The Dam Foundation

www.dam.no

