

# **Finn Hansson og Frode Frederiksen**

**Institut for ledelse, politik og filosofi  
Handelshøjskolen i København  
Blaagaardsgade 23 B  
DK-2200 København N**

## **Forskningsvotning eller forskningsudvikling?<sup>1</sup> Working Paper 21/2002**

Dette working paper er bygget op over en kritisk diskussion af ”Et forsøg på Benchmarking. En analyse af de samfundsvidenskabelige og juridiske institutter ved de højere læreanstalter.” Rapport fra Analyseinstitut for Forskning, 2001/6 og afsluttes med nogle overvejelser fra litteraturen om forskningsevaluering om andre måder at gennemføre sammenlignende forskningsevalueringer på.

### **1. Indledning**

Den evaluering af forskningens kvalitet, som altid har været en integreret del af videnskaben (peer reviews) har i de senere år fået følgeskab af en anden form for evaluering, der udspringer af et udefra kommende politisk krav om dokumentation for effektivitet og produktivitet (accountability). Peer reviews har ikke fuldstændigt eller alene kunne tilvejebringe den viden, som kunne opfylde kravene til synliggørelse og kontrol (Hansson 2001). Der er derfor i de senere år blevet lavet en række forsøg med at udvikle enklere (og billigere) metoder til at vurdere forskningen hos den enkelte forsker, på fagområder, institutionsområder, nationer mm. Kravene til offentlige forskningsinstitutioner er ikke længere forskellige med hensyn til de krav, der i dag stilles til alle andre offentlige institutioners aktiviteter (value for money). Forskningsinstitutionernes synliggørelse af og begrundelse for sine aktiviteter overfor andre myndigheder - i sidste instans offentligheden - er en proces, der er rettet udad, men det er også en proces, som med fordel kan anvende for at rejse interne diskussioner om udviklingshastighed og retning, dvs. en proces der rummer muligheder for organisationsudvikling.

Det hidtil seneste bud på en metode til en kvantitativ vurdering af ”produktivitet i en forskningsinstitution”, finder vi i ny rapport fra Analyseinstitut for Forskning om benchmarking af samfundsvidenskabelige institutter i Danmark (Damm 2001)<sup>2</sup>. Rapporten er bygget op som en benchmarking af de forskellige institutters forskningsproduktivitet base-

---

<sup>1</sup> En kortere udgave af dette working paper er trykt som en kommentar i Dansk Sociologi, vol. 12, nr. 2001.

<sup>2</sup> Rapporten kan hentes fra instituttets hjemmeside: <http://www.afsk.au.dk/>

ret på DEA-modellens beregninger af input-output relationer. DEA-modellen bruges af finansministeriet<sup>3</sup> til at vurdere offentlige virksomheders produktivitet og forsøges her tilpasset forskningsinstitutter, eller med rapportens egne ord:

”Rapporten illustrerer en metode til evaluering af forskning, der kan hjælpe med at finde mulige indsatsområder for de institutter, som ikke opnår status som best practices, således at disse institutter kan anvende viden fra de bedste institutter til styrkelse af deres egen forskningsplanlægning.”(Damm 2001: 8)

Rapporten vil således forstås som et bidrag til evaluering af forskning, men bevæger sig i virkeligheden også ind i områder som organisationsudvikling og videnoverførsel mellem forskningsinstitutioner, nemlig i det omfang der er tale om hvordan man skal finde frem til og bruge erfaringer fra ’best practice’ institutter.

Rapporten fremhæver at den valgte metode ikke kan stå alene i vurdering af forskning, men den rejser nogle interessante problemstillinger og giver derfor anledning til kritik og diskussion af centrale emner som hvad er et godt forskningsmiljø?, hvordan finder man et godt forskningsmiljø? og hvordan ændres et forskningsmiljø til det bedre? Rapporten veksler mellem at præsentere sig som et metodestudie og som en konkret benchmarking. I et efterfølgende notat fremhæver Analyseinstituttet, at der ikke er tale om et metodestudie hvis man accepterer modellens præmisser, herunder produktionsmålene og publikationsopdelingerne. Under disse forudsætninger skal undersøgelsen opfattes som en regulær benchmarking (Mortensen 2001: 3-4).

I næste afsnit vil vi kommentere den konkrete rapport og debatten om den anvendte metode. Dernæst vil vi folde fire centrale problemstillinger mere ud. Disse er a) grundantagelser ved brug af benchmarking, b) tidsdimensionen i modellen med henblik intentionerne om læring af ”best practice”, c) institutternes mange opgaver i sammenhæng med fokusering på publikationer og endelig d) forholdet mellem publicering og kvalitet med henblik på den konkrete opdeling af publikationerne. Afslutningsvist vil vi pege på nogle mere lovente tilgange til sådanne større sammenlignende forskningsevalueringer, som bl.a. kan findes i udlandet.

## **2. Benchmarking, produktivitet og vurdering af forskning.**

For at finde frem til et ’best practice’ institut foretages der en benchmarking af produktivitet hos de samfundsvidenskabelige og juridiske institutter ved de højere læreanstalter i perioden 1995-1997. Undersøgelsens centrale empiri er et instituts videnskabelige publikationer opgjort i årsberetninger i forhold til forskningsårsværk. Undersøgelsens teoretiske grundlag er DEA-modellen, som anvendes til at lave forskellige beregninger, hvor vægtingen mellem de forskellige publikationstyper ændres relativt for derved at sætte de enkelte institutters score i indbyrdes relationer<sup>4</sup>.

---

<sup>3</sup> Benchmarking i den offentlige sektor – nogle metoder og erfaringer, Marts 2000. Findes på ministeirets hjemmeside: [http://www.fm.dk/udgivelser/publikationer/benchmarking\\_00/index.htm](http://www.fm.dk/udgivelser/publikationer/benchmarking_00/index.htm)

<sup>4</sup> De anvendte beregningsmodeller er beskrevet i rapportens kapitel 3 og appendiks A.

Brug af produktivitetsoverretede modeller til at belyse forskningsaktiviteten har været forsøgt før – også her i landet. Bl.a. har Jennergren og Obels (1985) anvendt DEA-modellen til at benchmarke forskningen ved 22 økonomiske institutter. Denne benchmarking blev i sin tid kritiseret for at dels at antage, at de 22 udvalgte økonomiske institutter var homogene nok til at kunne sammenlignes ud fra denne produktivitetsoverretede model, og dels for at der bag deres argumentation for benchmarking af de økonomiske institutter ved brug af DEA-modellen lå en ikke klart formuleret antagelse om, at produktivitetssammenligninger på outputkategorier også var et gyldigt udtryk for kvalitetsforskelle. Ikke mindst Andersen og Foss Hansen (1985) bidrog med en omfattende gennemgang af for- og bagdele ved forskellige tilgange til forskningsevaluering, som især præciserede det begrænsede evalueringsperspektiv, som DEA-modellens effektivitetsoverretning lægger op til.

Men fremfor at tage afsæt i de problemer som blev rejst i debatten fra midten af 80'erne, tager rapporten ukritisk Jennergren og Obels arbejde til indtægt for ”at allerede i 1985 fremhæves DEA som en egnet metode til vurdering af forskning i Danmark” (Damm 2001: 110). Den nye rapport går et afgørende skridt videre i forhold til Jennergren og Obels undersøgelse ved at inddrage et kvalitetselement i selve modellen. Nu anvendes en klassificering af publikationer efter vigtighed og betydning, hvor to afgørende kriterier er *international publicering* og *publicering i peer review'ede tidsskrifter*. Men anvendelse af denne opdeling eller rangorden af publikationer efter kvalitet/anerkendelse, som kan findes mange steder, rejser en række metodeproblemer om kvalitetsmåling, som ikke tages op i rapporten.

Men heller ikke input-beregningerne er en enkel sag. For at nå frem til et enkelt tal til modellen antager rapporten, at antal forskningsårsværk gange 40% er det bedst mulige inputmål. Baggrunden må være at arbejdstiden for forskere på universiteter beskrives ud fra en norm på 40-50-10 (forskning-undervisning-administration) som dog aldrig er anerkendt af forskningsministeriet. Problemet med at bruge dette grove gennemsnit er, at der med de mange nye typer stillinger (ph.d., forskningsassistenter, -lektorer,-professorer mm.) er kommet meget store lokale variationer mht. forskningstid ind i billedet, hvorfor brugen af en gennemsnitlig procent kun har mening, hvis man antager at der er samme struktur i stillingsmønstret på alle institutter (Barré 2001: 2262).

### **3. Hvad kan og skal man sammenligne?**

En af de fordele der hyppigst fremhæves ved benchmarking metoder er, at de muliggør en organisatorisk læringsproces, hvor erfaringer fra 'best practice' søges overført på andre forskningsorganisationer (Barré 2001). Det indebærer ikke nødvendigvis at de enheder der sammenlignes på alle væsentlige områder skal ligne hinanden, men det fremhæves dog ofte i litteraturen at et vist mål af homogenitet skal være til stede. I universitetsverdenen betyder det, at hvis man ønsker at benchmarke institutters forskningsproduktivitet (alene forstået som antal publikationer i forhold til forskningsressourcer) er det en form for benchmarking, der kan gennemføres på en ret grov afgrænsning af fagområder eller hele institutioner eller lande som i en ny rapport fra EU<sup>5</sup>.

<sup>5</sup> European Commission Research Directorate General (2001).

Men på dette punkt er rapporten fra ASFK mere ambitiøs. Udover at beregne produktivitetssammenligninger ud fra DEA-modellen har rapporten ambitioner om at udvide analysen med beregninger, hvor produktivetsmål kombineres med et kvalitetsmål (Damm 2001: 53, 74). Rapporten søger at løse det meget komplicerede problem med at kvantificere kvalitetsmål ved at opstille en form for rangorden af publikationer efter type og tildele disse forskellige vægte. Denne måde at lave kvalitetsvurdering på er imidlertid ikke uden problemer og de løses ikke af, at rapporten ”kun (undersøger) institutter fra erhvervsøkonomiske, juridiske og samfundsvidenskabelige fakulteter.”

Problemet består i sin enkelthed i, at det øjeblik at benchmarkingen ændres fra den direkte målbare forskningsproduktivitet (som i princippet kan gennemføres mellem alle mulige fagområder, f.eks. medicin og fysik) til at benchmarke kvaliteten, så er vi tilbage i de helt grundlæggende problemstillinger der er forbundet med at vurdere forskningens kvalitet. Skal den traditionelle videnskabelige kvalitetsvurdering gennem peer review's erstattes af andre metoder kræver det, at de forskellige indikatorer for kvalitet vurderes mht. deres udsagnskraft<sup>6</sup>.

Desuden er det vigtigt at undersøge de områder der sammenlignes for at sikre, at der er et fagligt relevant grundlag og en ensartethed til stede, der gør brug af sammenligningsindikatorer hensigtsmæssig. Til det sidste punkt er der i dag al mulig grund til at være på vagt overfor en opdeling, der baserer sig på traditionelle organisatoriske strukturer som fakulteter og institutter. I en tid hvor diskussionen om grundlæggende forandringer i organiseringen af samfundets vidensproduktion er på dagsordenen med fokus på tværfaglighed og problemorientering i forskningen fremfor discipliners traditionelle faggrænser (Gibbons 1994, Nowotny 2001) må institutter opfattes som en meget formel organisatorisk ramme om vidensproduktion, en ramme som ikke uden videre kan tages til indtægt for at der disse steder er en tilstrækkelig homogenitet mht. faglige områder, publiceringstraditioner og -veje, direkte og indirekte brugere. En afgrænsning til institutter kan være problematisk og forudsætter en ganske høj grad af stabilitet, som ikke er givet ved vidensproduktion i dag. Flere af de institutter der indgår i undersøgelsen er ganske nye, nogle med et gammelt navn men med nyt indhold, andre er siden undersøgelsestidspunktet forsvundet eller rettere er formodentlig indgået i nye organisatoriske strukturer, andre igen har beholdt navnet mens indholdet er blevet ændret (nye centre og projekter kommer og går). Fremover må man forsøge at lokalisere homogene fagområder på tværs af institutionelle opdelinger.

Rapportens udgangspunkt, at DEA-modellen er yderst velegnet til at udpege ’best practice’ institutter, bygger på en række forudsætninger som ikke diskuteres i nævneværdigt omfang. Rapporten arbejder konsekvent ud fra den opfattelse, at evaluering af forskning er en problemstilling, der bedst kan løses med kvantitative værktøjer fra økonomiske produktionsmodeller, cost-efficiency, cost-benefit, eller som her DEA-modellen, og diskuterer ingen steder måling og kvantificering som en særlig forskningsstrategi (Hansson 2001a). Barré (2001: 263) peger i et kritisk studie af benchmarking af forskning på, at forskningsproduktivitet ikke kan opfanges af et enkelt tal, men at indikatorer alene er startpunktet for en øvelse i ’learning-by-comparing’ gennem diskussion.

---

<sup>6</sup> Den bedste diskussion af forskningsindikatorproblemer findes i Foss Hansen, Hanne og Holst Jørgensen (1995).

#### 4. Tidsdimensionen – det lærende element

Et andet forhold vi vil omtale, er rapportens intention om at institutter med best practice kan være forbilleder for institutter der ikke er best practice, men ”som ønsker at forbedre sig i håb om næste år at blive best practice” (Damm 2001: 51). Denne intention vedrører en åbning af ’black box’en i produktionsfunktionen med et input og et output. Det interessante er derfor hvilke processer, hvilken organisering og hvilken ledelse, der medvirker til at frembringe forskning af best practice og lige så interessant, hvordan man forbedrer disse processer og organisations- og ledelsesformer.

Vi har ikke svaret på disse spørgsmål, men er tvivlende overfor om bagudrettede indikatorer er de bedste virkemidler til at forstå disse processer – ud over den ofte implicite antagelse om at miljøer, der tidligere har udført god forskning, også fremover vil være i stand til at udføre god forskning. Problemstillingen kan sammenfattes således. De undersøgte publikationer er fra 1995 til 97. Afhængig af de konkrete forskningsprojekter og publiceringskanal, kan forskningen være udført flere år forinden; f.eks. to, fem eller ti år. De gode råd, som institutterne derefter skal udveksle, skal derefter implementeres, forskningen skal udføres og den skal publiceres. Dette kan også tage f.eks. to, fem eller ti år. Samlet kan der derfor være tale om et timelag på ganske mange år før en givet ønsket ændring kan konstateres. Hermed ikke sagt at institutter og forskningsmiljøer ikke skal lære af hinanden; effekter er blot vanskelige at måle og DEA-modellens fokusering på produktivitet alene er ikke det bedste udgangspunkt.

Der kan findes mange eksempler på at åbne forskningsevalueringer kan være et godt udgangspunkt for interne refleksioner og efterfølgende udvikling i de evaluerede institutioner. Bl.a. kan nævnes forsøgene med forskningsevaluering på Handelshøjskolen i København i midten af 90erne, hvor der blev arbejdet målrettet mod at inddrage de evaluerede faglige miljøer. I den anvendte metode blev der benyttet en blanding af faktiskt baggrundsmateriale og site-visits af internationale peers (Foss Hansen og Borum 1999).

#### 5. Forskning er mere end publicering.

Spørgsmålet er om fokuseringen på publikationer (som vi ved rummer mange *vanskeligheder*) også er den *rigtige* måde at vurdere universitetsinstitutter. Det fremhæves ganske rigtigt med eksempler, at det er mange af institutionernes eget ønske at øge publiceringen. Og flere andre outputmål bliver diskuteret (men ikke inddraget). De har alle tilknytning til en af universiteternes andre hovedopgaver – undervisningen. Men med til billedet af institutterne som forskningsmiljøer hører andre væsentlige opgaver end publicering. På den forskningspolitiske dagsorden bliver emner som nytten for samfundet (både erhvervslivet, det politisk-administrative system og befolkningen i øvrigt) mere centrale (Gibbons 1994, Nowotny 2001). Så kort fortalt bør aktiviteter som samarbejdsprojekter med og øvrige relationer til andre sektorer, arrangement af konferencer, deltagelse i høringer, efteruddannelse af forskellige samfundsgrupper, videnformidling i bred forstand medregnes ved siden af publikationer. Dertil kommer uddannelse af kommende forskere og udbygning og vedligeholdelse af nationale såvel som internationale netværk som en sikring af de fremtidige betingelser for god forskning.

Skiftet i sigtet med og det overordnede syn på forskning med øget fokus på nytten og anvendelsen og den større grad af samarbejde mellem institutioner og sektorer får ofte betydeligheden af et skift fra modus-1 forskning til modus-2 forskning. Denne udvikling har også betydning for kvalitetskontrollen i forskningen. Hovedreglen for modus-1 forskning er det anonyme peer review med fokus på den individuelle videnskabelige produktion, mens kvaliteten for modus-2 forskning er en kontinuerlig forhandling af behov og kvalitet mellem grupper af interessenter (Jacob 2000). Den daglige praksis for de fleste universitetsmiljøer er ikke et spørgsmål om enten-eller, men findes snarere i et kontinuum mellem disse yderpunkter. Og sammenfattende kan en mis-match mellem opgaver og bedømmelse udtrykkes som et krav om at mange institutter får til opgave at udføre modus-2 forskning (om end dette formål også i høj grad er til debat), mens de efterfølgende primært bliver bedømt på deres indsats indenfor modus-1 forskning.

Et bredere spektrum af benyttede indikatorer vil også ligge i god tråd med advarslen (Damm 2001: 111) om at institutterne – og institutionerne – vil forsøge at tilpasse sig og tænke strategisk i prioritering af forskningsområder for at øge publiceringen. Jævnfør også debatten om det britiske evalueringssystem (Trow 1996, Geuna & Martin 2001).

## 6. Publicering og kvalitet

Rapporten diskuterer kort vanskelighederne ved at måle kvalitet som et ikke umiddelbart kvantificerbart område. Uden at gå nærmere ind i denne diskussion, som i årevis har været et stående tema i videnskabssociologien og informationsvidenskaben, vælger rapporten delvis at bruge nogle af institutionernes egne opdelinger i typer af materiale<sup>7</sup>. En sammenfatning af rapportens definitioner ses her:

*Tidsskriftsartikler* bliver opdelt i hhv. *referee tidsskriftsartikler*, *ikke referee tidsskriftsartikler* og *populærartikler*. *Bøger* opdeles alt efter om bogen er udgivet af et *forlag*, udgivet af et *ministerium* o. lign. eller om der i teksten står, at det er en *lærebog*. Derudover findes *konferenceindlæg*, *proceedings*, *indlæg i større værker*, *indlæg i årsberetninger*, *working paper* og *andet* (rapporter, anmeldelser mm.).

I den konkrete model bliver flere af disse kategorier dog slået sammen til en gruppe for henholdsvis nordisksprogede og fremmedsprogede referee tidsskriftsartikler, en gruppe for andre artikler, ligeledes henholdsvis nordisksprogede og fremmedsprogede (grupperne dækker alt fra artikler i forlagsudgivne antologier til indlæg i aviser), endelig en gruppe for bøger, danske såvel som udenlandske samt en restgruppe. Det giver i alt seks kategorier og den *eneste* forskel fra den førnævnte undersøgelse af Jennergren og Obel er dermed grupperne med nordisksprogede og fremmedsprogede referee tidsskriftsartikler, der er valgt på grund af muligheden for at give ekstra vægt til "kvalitetsartikler" (Damm 2001: 48). Med andre ord er opdelingen udtryk for et kvalitetsmål og centralt for forståelsen af "best practice" er derfor "kvalitetsartikler" i tidsskrifter med referee<sup>8</sup>.

<sup>7</sup> For en sammenfatning og diskussion af publikationsopgørelser i et skandinavisk perspektiv se Frederiksen (1995).

<sup>8</sup> Damm tildeler de peer reviewede tidsskrifter de højeste vægte i en række modelberegninger for at bemarkere 'kvaliteten' i produktionen (s. 53, 69, 74)

Med tilføjelsen af denne kvalitative sondring ser rapporten derfor bort fra, at indenfor mange fagområder findes de mest anerkendte, originale og benyttede bidrag i f.eks. bøger, artikler i antologier eller proceedings.

I forhold til den ret detaljerede beskrivelse af de enkelte kategorier og af selve DEA-modelberegningerne, er beskrivelsen af hvilke procedurer der er brugt til at frembringe og klassificere dette empiriske råmateriale meget kortfattet. Spørgsmålet om udvælgelsen af ”kvalitetstidsskrifter” er central i modellen. I fraværet af en direkte peer review-baseret kvalitetsbedømmelse, opdeles artiklerne efter om pågældende tidsskrift bruger peer review til den redaktionelle udvælgelse eller ej. Det fremgår ingen steder i rapporten, at dette i virkeligheden er en ganske kompliceret proces og selv for førende tidsskrifter er der ingen entydig praksis eller standard for hvor udførligt og formelt den redaktionelle proces er beskrevet, en anden og endnu mere kompliceret proces er hvordan den daglige praksis er<sup>9</sup>. Det er i dag vanskeligt at forestille sig et videnskabeligt tidsskrift, der vil bekræfte at de ikke bruger en udvælgelsesprocedure baseret på peer review. Og hvorledes kan man så skelne mellem disse tidsskrifter<sup>10</sup>?

Når man i bilagene ser den konkrete klassificering af de enkelte tidsskrifter (Damm 2001: 157ff.) dukker der en række spørgsmål op, som der ingen steder i teksten kan findes forklaringer på. I kategorien *Nordisksprogede tidsskrifter m. referee*, dvs. med den højeste kvalitet eller vægt i modellen, finder vi følgende tidsskrifter:

Økonomisk Debatt, Ledelse og Erhvervsøkonomi, Nationaløkonomisk tidsskrift, Nordisk Administrativt Tidsskrift, Ugeskrift for Læger, Ugeskrift for Retsvæsen.

Fra den omfattende kategori *Nordisksprogede tidsskrifter uden referee* har vi valgt nedenstående samfundsvidenskabelige tidsskrifter ud, fordi det alle er tidsskrifter vi kender noget til og som bruger en form for peer review ved den redaktionelle behandling af modtagne artikler:

Dansk Sociologi, Den jyske Historiker, Filosofiske Studier, GRUS, Historisk Tidsskrift, Kritik, Philosophia, Politica, Retfærd, Samfundsøkonomen, Sociologisk forskning, Statsvetenskaplig Tidsskrift, Økonomi og Politik, Økonomistyring og informatik.

Da ovenstående tidsskrifter kun er registreret i kategori 2 vil institutter, hvis forskere hyppigt skriver i disse tidsskrifter, få en meget lavere score mht. best practice. Holder vi os inden for de skandinaviske sprog kan man sige, at sociologer har noget vanskeligere ved at

---

<sup>9</sup> Der er foretaget en del forskning i selve reviewprocessen, se ex. Hemlin, og Boe (1999), og Hemlin (1999). Også problemet om pålidelighed i denne proces har været genstand for omfattende forskning, en god oversigt giver Cicchetti (1991). En sammenfatning af problemer og styrker ved peer review systemet kan findes i Hansson (2001) og i Chubin og Hackett (1990).

<sup>10</sup> Mængden af forsøg på at vurdere og rangordne videnskabelige tidsskrifter er meget omfattende. Se f.eks. Andersen og Frederiksen (1995). Et eksempel på en løbende service gives af Emerald Reviews, der inden for området management har udvalgt godt 400 centrale tidsskrifter og ladet dem bedømme af en accrediteringsgruppe (peer reviews) opdelt i 10 fagområder (Accounting & finance, Human resource management osv.). Se <http://www.emeraldinsight.com/reviews/journals.htm>

få en høj score end jurister, da der ikke er et eneste nordisk sociologitidsskrift, der efter rapportens vurdering har en kvalitetskontrol!

Alene på baggrund af at rapporten uden nærmere forklaring har valgt at se bort fra en række anerkendte (og peer review'ede) samfundsvidenskabelige tidsskrifter, er det næppe så overraskende, at rapporten kan konkludere, at det er blandt juridiske og økonomiske institutter man kan finde best practice! (Damm 2001: 50). Den arbitrære opdeling af nordiske tidsskrifter i henholdsvis peer review'ede og ikke peer review'ede, ser ud til på forhånd have lagt en struktur i data der afspejler resultatet, jf. rapportens sammenfatning (Damm 2001: 83).

Tilsvarende gælder for udvælgelsen af den specielle gruppe af fremmedsprogede tidsskrifter med referee. Ved klassificering af internationale tidsskrifter anvendes ofte det kriterium om de bliver indekseret i ISI's citationsdatabaser. Denne opdeling er dog også ofte genstand for diskussion og kritik, da det hverken er entydigt eller gennemsigtigt hvilke tidsskrifter, der bliver optaget i disse databaser. Rapporten har da heller ikke benyttet denne metode, og i begge grupper (med eller uden peer review) findes mange tidsskrifter, der bliver indekseret i ISI's baser. Derimod bliver det ikke diskuteret hvilke kriterier, der så ligger til grund for opdelingen. Men da de fleste fremmedsprogede tidsskrifter er placeret i kategorien med peer review og omfatter en langt bredere faglig profil end tilfældet var med nordisksprogede tidsskrifter, har opdelingen formodentlig mindre systematisk betydning her.

## **7. Konklusion på kritikken af benchmarking.**

Vi har i denne artikel påpeget nogle principielle fejl og begrænsninger i rapportens empiriske grundlag (især vedr. udvælgelse af de peer review'ede tidsskrifter) og vi har påpeget, at en række forhold af betydning for evaluering af forskning, som de senere tiårs videnskabs-sociologi og evalueringslitteratur har fremhævet som væsentlige, hverken medtages eller diskuteres i denne benchmarking.

En konsekvens heraf er, at rapportens tilgang til og diskussion af benchmarking er snævert fokuseret på brug af DEA-modellens produktivitetsmål, og vi er ikke blevet overbevist om det relevante og hensigtsmæssige ved at bruge DEA-modellen til en benchmarking af dansk samfundsvidenskabelig forskning.

Det rejser naturligvis spørgsmål ved den anvendte benchmarking-models præmisser, men med Peter Mortensens ord så er resultatet af benchmarkingen givet når præmisserne accepteres: "Resultaterne af de kvantitative beregningen skal imidlertid **ikke** opfattes som et metodestudie, hvis man vælger at acceptere en given models præmisser.." (Mortensen 2001: 3, notatets fremhævelse).

De vigtige spørgsmål omkring vurdering af den offentlige forsknings kvalitet ud fra en bredere forståelse af produktivitetmåling og med inddragelse af de miljømæssige og foranderlige institutionelle rammer er det meget svært at finde blot antydning af svarmuligheder på i rapporten. Ikke desto mindre finder vi at det er vigtigt at fastholde, at kritikken af DEA-modellen og de anvendte data ikke er en afvisning af at bruge sammenfattende kvan-



titative oversigter og beregninger til forskningsevaluering. Det er tværtimod en kritik, som forsøger at rejse mere principielle spørgsmål om vanskelighederne ved at bruge meget forenkede mål (med lav pålidelighed) som udtryk for kvalitetsindikatorer.

Hvad skal man gøre fremover? Er der andre muligheder for at lave større tværgående evalueringer af forskningskvalitet og produktivitet i dansk forskning? Det er en problemstilling, der ikke tages op i ASFK-rapporten, men som uden tvivl må få stor opmærksomhed i fremtidige undersøgelser af forskningskvalitet. Hvis man retter blikket udover landets grænser findes en række muligheder for inspiration til andre, nye evalueringsmetoder.

## **8. Andre perspektiver på forskningsevaluering.**

I den omfattende litteratur om forskningsevaluering er det muligt at pege på to andre typer eller tilgange til evaluering af forskningsinstitutioner, som bryder med den meget produktivitetsoverrettede økonomiske model for evaluering, som ligger bag den type benchmarking vi har kritiseret foran.

Den ene retning kan man beskrive som den samtidige brug af mange forskellige indikatorer, en tilgang som Ben Martin (1996) under betegnelsen 'multiple indicators in assessment' har argumenteret for som en strategi for SPRU's tilgang til forskningsevaluering. Fra dette udgangspunkt har Geuna og Martin (2001) i en ny rapport fra SPRU undersøgt erfaringer med brug af forskningsevaluering på nationalt niveau til styring af universitetsforskningens budgetter. Der er her grund til at fremhæve den omfattende oversigt over for og bagdele ved 'performance-based' forskningsbudgetteringssystemer (Geuna og Martin 2001: 25-26), baseret på erfaringer fra en række lande. Det grundlæggende spørgsmål de rejser er: "does the benefits of a performance-based system outweigh the cost?" og deres konklusion som er baseret på en omfattende gennemgang af erfaringer fra en række lande er værd at citere udførligt – ikke mindst fordi den også peger på nogle andre initiativer og erfaringer (fra Holland) med andre tilgange.

“Given the substantial (and perhaps increasing) costs of a full performance-based system for allocating research resources to universities, it is worth concluding by pointing to the advantages of a hybrid system, that is one based partially on performance-based (incentive-creating) approaches and partially on educational size (cost-minimising) approaches. Among the countries analysed, The Netherlands, Finland and Denmark have various forms of hybrid systems. Of particular interest is the Dutch system. In both the UK and The Netherlands, research assessment is based mainly on informed peer review. However, unlike the UK which focuses on a one-dimensional concept of quality and links the exercise directly to funding, in The Netherlands the quality of research performance is assessed in relation to four dimensions (scientific quality, scientific productivity, scientific relevance and long-term viability of the research) and there is no direct link to funding. The greater breadth of the quality measurement encourages diversity in the research conducted by universities and in the strategies adopted to improving their research. At the same time, the lack of direct connection with funding results in significantly lower organisational and administrative costs though the assessment create competition based on reputation that can originate most of the advantages identified in Table 2. For those

countries contemplating a shift towards a performance-based funding system, such a hybrid may offer a better trade-off than the more expensive approach of the UK research assessment exercise.” (Geuna og Martin 2001: 33-34)

Geuna og Martin fremhæver især de hollandske erfaringer med en fler-dimensionel model for forskningsevaluering, der foretages af kommiteer med udenlandske eksperter (peer reviews) men med omfattende lokal indflydelse fra institutionerne (bottom-up) og uden direkte forbindelse med budgetlægning, som det bedste eksempel på denne mange-dimensionelle tilgang til forskningsevaluering<sup>11</sup>.

Den anden tilgang der kan trækkes frem i litteraturen kan betegnes som organisationsperspektivet eller 'social capital' perspektivet på forskningsevaluering. Denne tilgang flytter evalueringens fokus væk fra de målbare produktivitsdata og resultater fra programmer til de organisatoriske og sociale dimensioner i de processer, der frembringer viden. Denne tilgang er meget inspireret af dele af den nyere litteratur om knowledge management og den fokusering på person-, proces-, og organisationsaspekterne i vidensfrembringelsen med begreber som tacit knowledge, human og social capital (Bozeman, Dietz og Gaughan 2001, Rodgers og Bozeman 2001). En væsentlig faktor bag denne udvikling er den udvidelse af forskningsevalueringerne, både med hensyn til hvem der deltager (inddragelse nye grupper i selve evalueringsprocessen) og hvilke institutioner der skal evalueres, som man har kunnet iagttage i USA i de senere år.

I USA har offentlige forskningsinstitutioner siden vedtagelsen i 1993 af GPRA<sup>12</sup> været underkastet et lovkrav om offentlig fremlæggelse af forskningens præstationer og kvalitet. Susan Cozzens (1997, 2000) har i flere artikler fra de senere år undersøgt de amerikanske erfaringer med årlige resultatorienterede forskningsevalueringer gennem en årrække. Et måske overraskende resultat er, at

”In the area of result indicators, publications could have been used as an output indicator. Not a single U.S. agency has chosen to do so. The U.S. university system<sup>13</sup> has steered away from the use of publication counts as an evaluative measure, because it encourages quantity rather than quality. The NSF consciously chose not to base performance goals on output measures because of the possibility of distortion of behaviour. (Cozzens 2000: 18)”

I det hele taget er kvantitative resultatmål sjældne blandt amerikanske forskningsinstitutioner og udviklingen har klart gået i retning af at opbygge systemer med paneler af eksterne interessenter, fra industrien men også fra NGO'er, som deltager i vurderingen af opfyldelsen af institutionernes strategiske planer. Strategiske planer har tilsvarende fået større be-

---

<sup>11</sup> Der kan også henvises til Foss Hansen og Borum's analyse af en tilsvarende mange-dimensionel evaluering på Handelshøjskolen (1999).

<sup>12</sup> Government Performance and Results Act kræver at alle amerikanske regeringsinstitutioner fremlægger årlige rapporteringer af aktiviteter (Cozzens 1997, 2000)

<sup>13</sup> Det kan også nævnes at en af de mest kendte universitetsrankings i USA foretaget hvert år af US News benytter 16 forskellige parametre. Ingen af disse er baseret på publikationer, se <http://www.usnews.com/usnews/edu/college/rankings/ranknatudoc.htm>. US News rankings og metoder er naturligvis også genstand for diskussion og kritik. For interesserede kan der henvises til artikler i The Washington Monthly: <http://www.washingtonmonthly.com/features/2000/norc.html> og <http://www.washingtonmonthly.com/features/2000/0009.thompson.html>

tydning mht. at præcisere institutionens målsætninger men også mht. at stille nye krav til forskningens offentlige begrundelse, eller med Cozzens formulering: "Even basic research needs its public justification" (Cozzens 2000: 19). Det billede som Cozzens tegner af 8 års amerikanske erfaringer med løbende performance evaluering af offentlige forskningsinstitutioner er, at det har været nødvendigt at 'opfinde' nye måder at foretage denne evaluering på – måder som hverken er baseret på snævre produktivitetsmålninger eller på forskersamfundenes interne evalueringstraditioner (peer reviews). En af disse nye modeller er paneler med interessenter i de enkle program- eller forskningsområder, repræsentanter for det politiske system og institutionsrepræsentanter. Disse paneler vurderer hvorledes institutionen opfylder sine strategiske planer og baserer en del af vurderingen på peer reviews af programmer og projekter, men i modsætning til tidligere er disse peer reviews ikke længere beskyttet mod en kritisk offentlighed. Cozzens (1997, 2000) påpeger at denne form for offentlig forskningsevaluering også har indbygget nogle klare interessekonflikter, som kun i meget begrænset omfang er kommet frem i den amerikanske diskussionen af GPRA. På sigt vil professionelle evaluatoreer få en ny og vigtig rolle i arbejdet med at inddrage interessenter i evalueringsprocesserne og herunder at udvikle metoder der kan der kan oversætte og synliggøre forskningens 'public benefit' for ikke-eksperter.

"Instead, the new research evaluator needs a new set of skills, like facilitationg strategic planning, which helps organizations articulate their goals; stakeholder involvement, which reaffirms the role of the public; and logic modelling, to clarify how programmes bring about their results."( Cozzens 2000: 20)

Afslutningsvis vil vi fremhæve at initiativer, der synliggør forskningsindsatsen på de danske universiteter, er meget ønskelige. Det er nyttig information, både for universiteterne og for det omgivende samfund. Og en synliggørelse kan sikkert bidrage til en bedre forskningspolitisk debat i Danmark; jævnfør de ofte tilbagevendende diskussioner om "nul-forskere". Men det er også afgørende, at sigtet med synliggørelsen er åbenhed og gennemsigtighed, og her kan benchmarking ved brug af DEA modellen næppe bruges med sin snævre fokusering på produktivitet og forsøg på rangordning. Erfaringer fra lande som Holland og USA peger på, at evaluering af forskningskvalitet, der både skal have værdi i organisationen og accepteres af denne og samtidig skal have værdi for administrative myndigheder og den politiske offentlighed stiller krav til udvikling og brug af nye og utraditionelle metoder.

## Referencer:

- Andersen, Heine og Lars Frode Frederiksen 1995. "Internationale tidsskrifter i samfundsvidenskaber.myter og realiteter." *Samfundøkonomen* Vol 5, no. 1: 14-22
- Andersen, Ria og Hanne Foss Hansen 1985. "Forskningsevaluering og forskningspolitik." *Økonomi og Politik* no. 4: 257-268
- Barré, Rémi 2001. "Sense and nonsense of S&T productivity indicators." *Science and Public Policy* Vol. 28, no. 4: 259-266
- Bozeman, B., Dietz, J. S., & Gaughan, M. (2001). Scientific and technical human capital: an alternative model for reserach evaluation, *International Jorunal of Technology Management*, vol. 22, (7/8), 716-740.
- European Commission Research Directorate General (ed.) 2001. "Toward a European Research Area. Key Figure 2001. Special Edition. Indicators for benchmarking of national research policies." Bryssel: European Commission.
- Chubin, Daryl E. og Edward J. Hackett 1990. "Peerless Science. Peer review and U.S. science policy." Albany, N. Y.: State University of New York Press.
- Cicchetti, Dominic 1991. "The Reliability of Peer Review for Manuscript and Grant Submission." *Behavioral and Brain Sciences* no. 14: 119-186
- Cozzens, Susan E. 1997. "The Knowledge Pool: measurement Challenges In Evaluating Fundamental Research Programs." *Evaluation and Program Planning* Vol. 20, no. 1: 77-89
- Cozzens, Susan E. 2000. "Result Oriented Management: A New Ethical Context for Research Evaluation." IPTS Report no. 50: 15-21
- Damm, Bettina 2001. "Et forsøg på Benchmarking. En analyse af de samfundsvidenskabelige og juridiske institutter ved de højere læreanstalter." Rapport, Aarhus: Analyseinstituttet for Forskning.
- Finansministeriet 2000: "Benchmarking i den offentlige sektor – nogle metoder og erfaringer", København: Finansministeriet
- Foss Hansen, Hanne og Birte Holst Jørgensen 1995. "Styring af forskning.kan forskningsindikatorer anvendes?" Copenhagen: Samfundslitteratur.
- Foss Hansen, Hanne og Finn Borum 2000. "The Local Construction and Entactment of Standards for Research Evaluation. The case of the Copenhagen Business School." *Evaluation* Vol. 6, no. 3: 281-299
- Frederiksen, Lars Frode: Publicering, publikationsopgørelser og forskningsevaluering, i: Proceedings of the 13th Nordic Conference on Business Studies. Vol. 1, København, 1995.
- Geuna, Aldo og Ben Martin 2001. "University Reserach Evaluation and Funding: An International Comparison." SPRU-Science and Technology Policy Research, University of Sussex.
- Gibbons, Michael, Camille Limoges, Helga Nowotny, Simon Scott Peter Schwartzman, og Martin Trow 1994. "The new Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies." London: Sage Pubications.
- Hansson, Finn 2001. "Kan man have tillid til peer reviews? Evaluering af universitetsforskningen ved brug af peer reviews." *Vest* Vol. 14, no. 2: 7-30
- Hansson, Finn 2001a. "Tillid til samfundsforskningen - om at skabe tillid gennem kvantificering." in Wen-

- neberg, Søren Barlebo og Anders Bordum (eds.) Det handler om tillid. København: Samfundslitteratur.
- Hemlin, Sven og Ole Boe 1999. "Hur bedömer lektörer manuskript till vetenskapliga tidskrifter?" Vest Vol. 12, no. 4: 27-46
- Hemlin, Sven 1999. "(Dis)Agreement in Peer Review." in Juslin, Peter og Henry Montgomery (eds.) Judgment and decision making: neo-Brunswikian and process-tracing approaches. Mahwah, N.J.: Erlbaum Associates.
- Jacob, Merle 2000. "'Mode 2' in Context: The Contract Researcher, the University and the Knowledge Society." in Jacob, Merle og Tomas Hellström (eds.) The Future of Knowledge Production in the Academy. Buckingham: Open University Press.
- Jennergren, L. P og B. Obel 1985. "Forskningsevaluering - eksemplificeret ved 22 økonomiske institutter." Økonomi og Politik no. 2: 86-97
- Martin, B. (1996). The Use of Multiple Indicators in the Assessment of Basic Research, *Scientometrics*, vol. 36, (4), 343-362.
- Mortensen, Peter S. 2001. "Notat vedr. kritikken af rapporten "Et forsøg på Benchmarking" Aarhus: Analyseinstituttet for Forskning.
- Nowotny, Helga, Michael Gibbons, og Peter Scott 2001. "Re-thinking science. Knowledge and the Public in an Age of Uncertainty." Oxford: Polity Press.
- Rodgers, J. D. & Bozeman, B. (2001). "Knowledge Value Alliances": An Alternative to the R&D Project Focus in Evaluation, *Science, Technology & Human Values*, vol. 26, (1), 23-55.
- Trow, Martin 1996. "Trust, markets and accountability in higher education: a comparative perspective." Higher Education Policy Vol. 9, no. 4: 309-324