

***Peer reviews i forskningsevaluering
Kvalitetsbedømmelse af forskning, når vilkårene
for vidensproduktion forandres***

Finn Hansson

WP 7/2000

June 2000

REMAP

www.remap.dk

*Research Management
Processes under rapid
change*

Institut for Ledelse, Politik og Filosofi
Analyseinstitut for Forskning
Haldor Topsøe A/S
NKT Research Center A/S
TeleDanmark A/S

Risø, Afdelingen for Systemanalyse
Danisco Cultor Innovationen
Institute of Drug Analysis A/S
Reson A/S

MPP Working Paper No. 7/2000 ©

June 2000

ISBN: 87-90403-75-4

ISSN: 1396-2817

Dette working paper er udgivet som en del af **REMAP**-projektet:

Research Management Processes under rapid change

Yderligere oplysninger: www.remap.dk

Department of Management, Politics and Philosophy

Copenhagen Business School

Blaagaardsgade 23B

DK-2200 Copenhagen N

Denmark

Phone: +45 38 15 36 30

Fax: +45 38 15 36 35

E-mail: as.lpf@cbs.dk

<http://www.cbs.dk/departments/mpp>

Peer reviews i forskningsevaluering

Kvalitetsbedømmelse af forskning, når vilkårene for vidensproduktion forandres

Finn Hansson

Copenhagen Business School
Department of Management, Politics and Philosophy
Blaagaardsgade 23 B
DK-2000 Copenhagen N.
E-mail: fh.lpf@cbs.dk

Indledning

Det videnskabelige peer review har været anset som en helt central institution i den moderne videnskabs udvikling til vurdering af videnskabelig kvalitet i de sidste 100 år - fra udvælgelse af artikler over vurdering af projekter til ansættelse af forskere. Peer review'ets betydning er hidtil ikke blevet påvirket af de seneste års samfundsmæssige forandringer i videnskabens og forskningens organisation og funktion. Den moderne videnskabens fremvækst og institutionalisering skete i en historisk periode, hvor det fra videnskabens side handlede om at frigøre sig fra tidligere tiders politiske og religiøse forsøg på indgriben og kontrol gennem at udvikle sine egne, interne regler og normer. Dette kan ikke siges at være tilfældet i dag, hvor videnskaben i stigende grad integreres i samfundet - ikke mindst på baggrund af videnskabens eller den videnskabelige videns omfattende betydning for samfundets økonomiske udvikling. Dette betyder også, at de systemer (normer, regler) der oprindeligt fremmede videnskabens autonomi i samfundet ikke længere kan antages at have den samme betydning og vægt i samfundet som tidligere.

Artiklen giver en kort redegørelse for det videnskabelige peer reviews fremkomst og dernæst et overblik over litteraturen om peer reviews med særlig vægt på de metodeproblemer, som oftest diskuteres i litteraturen. På baggrund af et eksempel fra forskningsevaluering på danske

universiteter inddrages de forskningspolitiske styringsforsøg, som indenfor en overordnet New Public Management forståelse har søgt at konkretisere indsigt med og styring af den offentlige forskning. Især diskuteres konsekvenserne for den klassiske peer review model af, at peer reviews nu anvendes til at bedømme forskningen på et mere overordnet og organisatorisk plan og på mere summariske måder end i den klassiske model. I forlængelse heraf peges på de tilløb til en måske uønsket standardisering og ensliggørelse af vidensproduktionen, som dette let kan føre med sig. Desuden rejses spørgsmålet om hvilken rolle peer review-bedømmelser har i forhold til de tilløb til nye former for forskningsorganisering og vidensproduktion, (som eksempelvis formuleres med begreber som om Mode 1 og mode 2 videnskab og Triple Helix), og som lægger vægt på at forskningen betragtes som en kompleks proces, hvor samarbejde og anvendelse fremhæves overfor den klassiske videnskabelige tekst.

I slutningen stilles spørgsmålet om de historiske betingelser for den peer review - baserede kvalitetsbedømmelse af forskningen nu er så forandrede, at en fortsat fastholdelse af denne evaluering- eller bedømmelsesmetode kan være direkte kontraproduktiv. Betyder eksempelvis fremvæksten af nye organisationsformer i forskning og vidensproduktion, at forskningens organisering, dvs. teamet eller samarbejdsrelationerne, i langt højere grad skal have en fremtrædende plads i kvalitetsbedømmelsen end den skrevne tekst? Hvorledes skal forskningens anvendelse vægtes i forhold til bedømmelsen af traditionelle forskningsresultater?

1. Det videnskabelige peer reviews historie

Begrebet peer reviews har i dag i store dele af den offentlige debat vundet indpas som en særlig for form videnskabelig kvalitetskontrol, baseret på en model der ideelt foreskriver, at 'ligemænd måler og vejer' og derigennem vurderer kvaliteten af andres videnskabelige arbejder. Denne forståelse svarer også meget godt til en kombination af de to oversættelser vi har på dansk af det engelske ord 'peer'; *a peer* ligemand (oprindeligt adelig) og *to peer*, at undersøge, at se nøjere efter.

I forbindelse med videnskabens institutionelle udvikling og organisering, der for alvor slog igennem i slutningen af 1800-tallet, opstod behovet for en videnskabelig kvalitetsvurdering,

der ikke var helt afhængig af enkelte personers luner og holdninger. Ligemandsbedømmelse havde man allerede set i funktion i mange år, bl.a. i England, i retssystemet i form af nævningeinstitutionen, der går tilbage til den tidlige middelalders forestillinger om, at retfærdighed bedst kan sikres ved at lade spørgsmål om skyld og straf afgøre af en gruppe af den anklages ligemænd (nævninge)¹. Der er nok ingen tvivl om, at den hurtige udbredelse og almene accept af denne videnskabelige ligemandsbedømmelse i et ikke ringe omfang skyldes, at der i to af de vigtigste lande mht. videnskabens institutionelle udvikling i forrige århundrede, USA og England, allerede eksisterede en flere hundrede år gammel tradition for denne form for vurdering indenfor retssystemet i form af nævningeinstitutionen. Brugen af den retslige legitimering kan også genfindes i de forskellige former for ret-lignende systemer, der i forskellige sammenhænge, i professionelle organisationer, i statslige institutioner, er blevet nedsat til at afgøre videnskabelige konflikter, herunder spørgsmål om efterligning eller om svindel, som det almindelig peer review system og den åbne videnskabelige diskussionen ikke har kunnet håndtere².

Spørgsmålet om udvælgelse af den videnskabelige viden af bedste kvalitet bliver et samfundsmæssigt problem af stor betydning under den industrielle kapitalismes eksplosive vækst i begyndelsen af det 18. århundrede. Her opstod der en samfundsmæssig efterspørgsel efter såvel naturvidenskabelig (kemi, fysik) som samfundsvidenskabelig (statistik, politisk økonomi) viden. I denne komplicerede proces hvor videnskaben også for alvor udviklede sig med egne institutioner og en selvregulering. I den forbindelse blev spørgsmålet om at finde metoder, der kunne vurdere kvalitet og relevans af ny videnskabelig viden med en rimelig grad af gyldighed og pålidelighed, et presserende problem både i forhold til anvendelsen og dermed legitimiteten af den udvalgte viden og internt i videnskaben selv. Problemet var og er

¹ “Scholars disagree on the time and place of the trial jury's birth. Some suggest that King Alfred the Great of England initiated the institution in the 9th century. Others trace it to the Norman Conquest of England (1066). The petit jury emerged as a distinct form when the Articles of Visitation in England (1194) separated accusatory and trial juries—the grand and petit juries of today.” (Opslag på ordet “jury” i Encyclopedia Britannica Online 2000). Det er måske værd at bemærke, at nævningeinstitutionen først blev indført i Danmark ved retsplejerreformen i 1916 – efter et langvarigt opgør med enevældets retstraditioner. Jf. Ditlev Tamm, *Dansk Retshistorie*, s.307.

² I en analyse af de samfundsmæssige betingelser for videnskabelig autoritet skriver Gaskins (1993): “Although lacking the institutional hierarchy of legal systems, scientific procedures are identified with distinct communities of investigators, through which judgments of scientific validity are rendered.” (s. 143)

først og fremmest, at spørgsmålet om hvad der er videnskabelig viden, hvad er 'ny viden' og 'viden af høj kvalitet' er et yderst komplekst og sammensat spørgsmål³, som det indtil i dag har været anset for umuligt at vurdere i sin helhed ved brug af entydige kvantitative målemetoder og faste standarder. I den situation viste ligemandsinstitutionen sig at være et columbusæg. Princippet om at lade udvalgte ligemænd, dvs. de højtspecialiserede fagfolk foretage en samlet vurdering på samme måde som proceduren ved nævningeafgørelser, dvs. vurdere det fremlagte materiale, og træffe en afgørelse eller dom, viste sig at kunne tilgodese både en ekstern og en intern legitimitet af beslutningen.

Det videnskabelige peer review, hvis historiske oprindelse vi finder i udvælgelse af manuskripter til de først videnskabelige tidsskrifter, viste sig således at kunne erstatte den viden om forskningens kvalitet som tidligere var baseret på tillid gennem de personlige relationer i de enkelte fagområder, som eksisterede i perioden før videnskabens vækst og institutionalisering⁴ på universiteterne og i virksomhederne. Vurderingen af forskningens kvalitet baseret på tillid og personligt kendskab erstattes nu af et kompleks af sociale normer og forventninger mellem deltagerne i videnskabssystemet. Disse kan etablere en ny form for tillid, som er baseret på en objektiviseret, ikke personorienteret vurdering foretaget af andre anerkendte forskere. Den klassiske beskrivelse af disse normer for anerkendelse og kritik findes tydeligst beskrevet som CUDOS-normerne hos Robert Merton⁵. Denne ligemandsbedømmelse implicerer en homogenitet og konsensus med hensyn til hvad der anerkendes som viden af kvalitet, gyldig viden. Dette er uhyre vigtigt for videnskabens udvikling som institution, dvs. som en samfundsinstitution, der kan fastholde social

³ Det sammen gør sig gældende for kunst. I Danmark blev peer review princippet blev indskrevet i lov om Statens Kunstfond fra 1962 som det overordnede princip for udvælgelse af den støtteberettigede kunst og kunstnere. Her findes en meget konsekvente udgave af peer review systemet, idet et udvalg bestående af 3 kunstnere indenfor et område, udpeget som sagkyndige af faglige organisationer og minister, i 3 år suverænt udvælger støtteberettiget kunst/kunstnere - og afgørelserne kan ikke ændres af andre myndigheder.

⁴ Robert Merton og Harriet Zuckermann har i artiklen *Institutionalized Patterns of Evaluation in Science*, *Minerva*, 9, 1, 1971, optrykt i Merton 1973, kap.21, s.460-496 en kort men glimrende gennemgang af dette historiske forløb, der oprindeligt kun var rettet mod udvælgelse af artikler og manuskripter til udgivelse.

⁵ Disse normer er: ”**Communism** (common ownership of goods (knowledge)), **Universalism** (truth claims are to be subjected to preestablished impersonal criteria), **Disinterestedness** (a distinctive pattern of institutional control of a wide range of motives which characterizes the behavior of scientists), **Organized Scepticism**”, s.606-614 i Merton 1968a.

anerkendelse fra det omgivende samfund og dermed denne institutions autonomi og selvregulering.

I den nu nærmest historisk artikel af Merton og Zuckermann (1971) diskuteres en række af de problemer, der er opstået og/eller er udviklet i takt med at peer review-metoden i tiltagende omfang institutionaliseres som metode til kvalitetsbedømmelse eller evaluering af forskning. Det er først og fremmest spørgsmålet om peer reviews af artikler til tidsskrifter giver en pålidelig udvælgelse, således at kvalitetsartikler med temmelig stor sandsynlighed ikke vurderes negativt. Merton og Zuckermann's konklusion på dels en gennemgang af forskellige undersøgelser af forkastelsesgrader (dvs. den % manuskripter som afvises ud af den tilsendte mængde) samt deres egen undersøgelse af peer review bedømmelsen i tidsskrifter *Physical Review* sammenfatter de i følgende citat:

“Errors of judgement, of course, occur. But the system of monitoring scientific work before it enters into the archives of science means that much of the time scientists can build upon the work of others with a degree of warranted confidence. It is in this sense that the structure of authority in science, in which the referee system occupies a central place, provides an institutional basis for the comparative reliability and cumulation of knowledge.” (s.495)

Merton og Zuckermann finder således empirisk grundlag for at placere peer review - systemet i videnskabelig publicering som et helt afgørende forhold i sikringen af videnskabelig kvalitet og derfor også som den centrale evalueringsmetode, når vi har med videnskabelig kvalitet at gøre. Det ‘community of science’ som Merton adskillige steder taler om som den institution, der sikrer gyldighed og kvalitet i videnskabelig viden gennem den aldrig endelige videnskabelige kritik og skepticisme, har således fundet sin operationelle og institutionelle form i peer review modellen. Denne beskrivelse af idealnormerne for videnskabelig adfærd fastholdes af Merton som en nærmest ‘idealtypisk’ sammenfatning af vilkårene for videnskaben og dens frembringere i det moderne samfund, trods de skævheder i videnskabens belønnings- og kommunikationssystemer som han selv har undersøgt, først og fremmest Matthæus-effekten⁶. Men han gør sig ingen mere præcise forestillinger om de mere historisk specifikke betingelser for disse normers eksistens og betydning og i hvilket omfang de er

⁶ Se f. eks. fremstillingen i Merton 1968b.

afhængige af ændringer i det omgivende samfund. Tværtimod forekommer det som om Merton implicit antager, at videnskaben har udviklet sig til et relativt selvstændigt og lukket system i forhold til det omgivende samfund og at dette system hviler på disse normer og kan friholdes fra politiske interesser og konflikter, i forhold til det omgivende samfund⁷.

Som metode til at foretage videnskabelige kvalitetsvurderinger blev peer reviews hurtigt den absolut foretrukne metode for udvælgelse af artikler og bøger til publicering, af ansøgere til stillinger, til forskningsbevillinger og i en række andre situationer, hvor udvælgelse efter den bedste videnskabelige kvalitet var afgørende. Det skal dog fremhæves, at peer reviews ikke kan beskrives som en entydig og veldefineret metode, men snarere som et overordnet og anerkendt princip for videnskabelig kvalitetssikring. Der kan nogle gange foreligge en formel beskrivelse af proceduren for hvordan evalueringen eller bedømmelsen skal gennemføres og hvordan resultatet skal præsenteres, og der kan fastlægges principper for hvad der skal eller ikke skal medtages, men grundlaget for evalueringen er først og fremmest 'den videnskabelige tekst'. Selve bedømmelsen af indholdet, vurderingen af videnskabelig kvalitet, står uantastet som en kvalitativ proces, der stort set er fritaget for eksternt formulerede formelle procedurer og metodikker og som hviler på 'de videnskabelige ligemænds' mere eller mindre udtalte kvalitetskrav.

2. Er peer reviews den rette model til forskningsevaluering?

⁷ Et eksempel giver Simon Locke 1999 som skriver om Mertons normer, at "[...] in the West, the critical norms of the scientific community fortunately coincide with and mutually reinforce those of democratic politics." (s. 28)

Der kan være en række grunde til at stille dette spørgsmål. Den mest oplagte er måske, at de Merton'ske normer om universalisme og uinteressethed/uhildethed (disinterestness) forudsætter en form for nærmest idealtypisk tilstand, hvor viden kan eksistere uafhængig af en samfundsmæssig sammenhæng, dvs. uafhængig af værdier og holdninger hos såvel den enkelte forsker som i den øvrige videnskabelige verden. Med tanke på den enorme vækst som videnskaben har gennemgået i det sidste århundrede, og ikke mindre den hurtigt voksende direkte statslige politik overfor videnskabens udvikling siden 2. verdenskrig, er det meget relevant at se nærmere på, om der er sket afgørende forandringer i de samfundsmæssige forudsætninger for peer review-metoden. I den Merton'ske idealtypiske beskrivelse af forudsætningerne for peer reviews findes en antagelse om et relativt uafhængigt 'community of science'. Er dette fortsat en holdbar forudsætning? Ikke mindst med tanke på, at denne vurderingsmetode i stigende omfang anvendes i mere tekniske og administrative og ikke mindst forskningspolitiske sammenhænge er det vigtigt at få rejst denne diskussion. Er metoden stadig at foretrække som den mest kvalificerede når en forsker eller et forskningsprodukt skal bedømmes mht. kvalitet?⁸ Kan den erstattes eller udvides med andre vurderingsmetoder? Gibbons og Georghiou (1987) har i en større undersøgelse af forskningsevaluering i OECD set nærmere på peer review og skriver:

“[...] peer review is the central method on which most evaluations are based. It operates at all levels of evaluation, over all types of research activity in every national research system. Equally, there is little doubt that peer review is believed to work reasonably well, though there are justifiable criticisms about its appropriateness in particular instances. [...] We simply wish in these concluding remarks, to underline the fact that many countries are experimenting with peer review; [...]”. (s.70)

Godt 80 år tilbage i sociologiens historie finder vi et af de første og meget markante forsøg på at problematisere forholdet mellem videnskab og samfund. En afpersonificering af videnskabelig viden blev allerede diskuteret af Max Weber i det berømte essay *Videnskab*

⁸ Hvad kvalitet i forskning er er i sig selv en meget omfattende diskussion, der ikke tages op her. Søren Wenneberg gennemfører en længere diskussion af begrebet bl.a. ved brug af Hemlins empiriske forsøg på at opstille kvalitetsparametre. I sammenhæng med forskningsevaluering er det pt. tilstrækkeligt at fastholde, at peer review metoden implicit fastlægger en kvalitetsdefinition gennem de udtalte kriterier, de deltagende bedømmere anvender. Man kan måske gå så langt som at hævde, at peer review evalueringer er baseret på at kvalitet er 'tacit knowledge' mellem bedømmerne. Se Søren Barlebo Wenneberg 1999.

som kaldt eller profession (Weber (1919) 1992/1999). Her argumenterer Weber for, at videnskaben i dag (1920'erne) skal betragtes som et fagmæssigt udført arbejde, som tjener selvbesindelse og sagkundskab og ikke som en nådegave hos profeter, eller en kamp for sandhed. I den affortryllede verden, som er Webers analytiske begreb for det moderne samfund, betyder den øgede intellektualisering og rationalisering en bevidsthed om, at det principielt er muligt at beherske og kontrollere verden (den affortryllede). Videnskab bliver en karriere i en sammenhæng, hvor værdier og personlige anskuelser bør skilles fra den videnskabelige indsigt, forstået på den måde, at 'ingen videnskab er absolut forudsætningsfri'. For Weber handler det ideelt om at skelne mellem at erkende og fremstille kendsgerninger og konsekvenser og heroverfor at tage stilling til værdier og handlinger i samfundet. Men dette er en uløselig konflikt, hvor værdier og viden ustandselig vil blive sammenflettet, da videnskaben frembringes og anvendes i en samfundsmæssige sammenhæng, hvor værdierne i høj grad gør sig gældende.

Den adskillelse mellem videnskab og værdier som Weber fremhæver, ligger til grund for Mertons formulering af CUDOS-normerne. Men i modsætning til Weber fastholder Merton ikke den uløselige konflikt mellem værdier og videnskabelig viden, som for Weber er grundvilkår for videnskaben i det moderne samfund. Webers andet hovedpunkt er at den beherskelse af verden, som han ser under udvikling og sammenfatter i begrebet 'affortryllelse', indebærer en stigende anvendelse af den videnskabelige viden i alle områder - og hermed har vi så at sige 'genindført' sammenkoblingspunktet mellem videnskab og politik. Endvidere peger Weber på det forhold at videnskab udvikler sig som en profession, et fag, et fagligt arbejde, lønarbejde. Det betyder, at forskere bliver lønarbejdere, med karriere, ansættelse i en forskningsorganisatorisk sammenhæng.

På to afgørende punkter kan vi allerede hos Weber finde en problematisering af den meget lukkede og funktionelle fremstilling af videnskaben som et selvstændigt og uafhængigt system, kernen i Mertons analyse af de videnskabelige normer. Efterfølgende samfundsforskere, ikke mindst fra kritisk teori, har udviklet forskellige ansatser til en mere omfattende problematisering af antagelser om, som hos Merton, at videnskaben kan betragtes som et område eller en sfære med en relativ høj grad af selvstændighed og uafhængighed i forhold til det moderne kapitalistiske samfunds økonomiske og politiske udvikling. Jürgen

Habermas (1968) har i en række sammenhænge kritiseret det moderne senkapitalistiske samfunds legitimerende brug af videnskab. Der er for ham en klar sammenhæng mellem det senkapitalistiske samfunds tekniske og instrumentale interesser for kontrol og beherskelse og anvendelse af viden frembragt gennem metodologiske og regelbaserede rammer for vidensproduktionen. Dette slår igennem i både de empirisk-analytiske og normativ-analytiske samfundsvidenskabers vidensproduktion, der objektivistisk og naivt korrelerer teoretiske udsagn med facts. Alfred Sohn-Rethel (1970/1091) har forsøgt udvikle en teori om forholdet mellem ånds- og håndsarbejde, grundlaget for den kapitalistiske arbejdsdeling og klassestruktur og dermed også for den samfundsmæssige reproduktion af særlige videns- og bevidsthedsformer, der tilsyneladende spontant vokser frem hos og blandt aktørerne på basis af selve varebyttehandlingens logik som særlige erkendelses- og tankeformer.

En anden grund til at stille spørgsmål ved den traditionelle tilgang til forskningsevaluering gennem peer reviews har mere direkte at gøre med de normer og øvrige forudsætninger for denne videnskabelige kvalitetsbedømmelse, som findes så klart formuleret hos Merton. Antagelsen om at peer reviews er den bedste metode til at vurdere forskningskvalitet, baserer sig på en udtalt antagelse om, at videnskaben er et relativt selvstændigt område i samfundet som har og kan fastholde nogle særlig normer, der er så udbredte og anerkendte, at de kan udgøre et vurderingsgrundlag, som fremstår som gyldigt og uantastet ikke blot indenfor videnskaben men også med stor autoritet og legitimitet udenfor. Allerede med Thomas Kuhns (1970) banebrydende undersøgelse af paradigmeskift og videnskabelige revolutioner som viste, at uenigheder og uklarheder om teori og metodologi kendetegner forskningens 'frontlinie', blev der rejst spørgsmål ved forestillingen om videnskabelig enighed. Det rammer naturligvis også Mertons antagelser om den nødvendige faglige enighed som forudsætning for peer review-bedømmelsen - i hvert fald når der ikke er tale om hvad Kuhn betegner 'normalvidenskab'. En række senere forsøg på at reformulere og udvikle forskellige beskrivelser af et fælles normsæt for videnskaben, f. eks som hos Hagstrom (1965), har ikke fjernet sig væsentligt fra de præmisser, som ligger bag Mertons CUDOS-normer og kan derfor ikke bruges til at løse dette problem. Den væsentligste kritik af Merton (f.eks hos Fuller 1997) drejer sig om hvilke konsekvenser det må antage at have for CUDOS-normerne, og dermed også for peer review som kvalitetskontrol, at videnskabens samfundsmæssige betydning og relationerne til de øvrige dele af samfundet har udviklet sig så hastigt de seneste år.

Whitley (1984) har med en organisationsanalytisk tilgang forsøgt at vise, at der eksisterer meget store indbyrdes forskelle mellem forskellige videnskabsområder med hensyn til måden at organisere vidensproduktionen på. Disse forskelle afspejler i høj grad forskelle imellem forskellige videnskabelige områders forhistorie og ikke mindst i tilknytning og anvendelse i forskellige dele af samfundet. Whitleys påvisning af den store betydning de organisatoriske og institutionelle forhold har i forskellige videnskabsområder med hensyn til videnskabsområdernes interne kontrol og centralisering af viden og anerkendelse har flere konsekvenser⁹. Disse forhold gør det vanskeligt at tale om videnskab i en mere generel forstand disse forskelle taget i betragtning og derfor også om fælles normer og kriterier. Desuden lægger vægningen af de organisatoriske og institutionelle forhold omkring videnskaben direkte op til at medtænke betydningen af efterkrigstidens voldsomme udvikling i anvendelsen af videnskaben i den samfundsøkonomiske produktion og statslige politikområder, som ikke mindst er kommet til udtryk i den stadig voksende interesse for forskningspolitik. Når staten gennem forskningspolitikken i stigende omfang forsøger at styre forskningen, når videns/forskningsanvendelse i virksomheder vokser med stigende hastighed, så peger dette på, at forholdet mellem videnskab og samfund må genovervejes. Det klassiske 'område' for statens regulering og udvikling af videnskab er universiteterne. Her har Borum og Foss Hansen (1999) vist, at den model der har udviklet sig i de senere år som 'modellen' i dansk forskningsevaluering på universiteterne, 'den modificerede 'peer-review' model', modsat den traditionelle anvendelse af peer reviews retter sig mod forskningens organisatoriske rammer og samlede produktion og ikke direkte mod det enkelte produkt (teksten) eller person. Med baggrund i nogle af de samme erfaringer peger Engwall (1998) på, moderne evalueringer på universiteterne skal rette sig mod strukturer og ikke individer.

De seneste års udvikling i såvel videnskabssociologien som den mere generelle samfundsvidenskabelig teori rummer en række bidrag, som har den samfundsmæssige anvendelse af videnskabelig viden og teknologi som et helt centralt tema. Det drejer sig både om omfattende ændringer i videnskabens og forskningens organisering og institutionalisering

⁹ I et nyt indledningskapitel til en ny udgave af bogen *The Intellectual and Social Organization of the Sciences*, udtrykker Whitley dette således: "It emphasises the considerable variety of ways in how research is organised and controlled across the sciences, and shows how these variations are related to different patterns of intellectual organisation. The explanation of such differences and their continued reproduction is linked to variations in reward systems and in the structure of legitimate audiences for intellectual contributions, as well as in the conditions governing access to research funds and other key resources" Whitley 2000, s.2.

og ikke mindst om videnskabens voksende betydning og synliggørelse i alle dele af samfundet (risikosamfundet jvf. Beck 1994, 1999, Giddens 1994). Det overordnede spørgsmål der rejses her er, om der er sket så omfattende samfundsmæssige ændringer i vilkårene for videnskab og vidensproduktion i de moderne samfund, at den klassiske model for opfattelsen af videnskaben som en aktivitet med en relativ autonomi i forholdet til samfundets politiske og økonomiske strukturer, fortsat er holdbar? Den kritiske teoretiske påpejning af videnskabens tætte sammenfletning med ikke mindst de økonomiske strukturer i et kapitalistisk samfund kunne i dag se ud til at have fået en mere empirisk eller konkret betydning.

Videnskabens voksende afhængighed og sammenfletning med det øvrige samfund tematiseres også på forskellige måder i den nyere videnskabssociologi, mest fremtrædende som forskellen mellem Mode 1 og Mode 2 videnskab eller i begrebet eller modellen Triple Helix.¹⁰ Begge disse tilbage er baseret på en række case-studier fra forskellige fag- eller videnskabsområder, hvor man har kunnet lokalisere væsentlige forandringer i organisering af forskning og produktion af viden. Konsekvensen af en fortsat udvikling af disse nye tendenser i forskningens og videnskabens samfundsmæssige relationer bliver bl.a., at den klassiske model for kvalitetsvurdering også må sættes under fornyet diskussion. Meget kort formuleret handler det i denne sammenhæng om hvilken betydning for kvalitetsproblematikken i videnskaben og vurderingsmetoderne som de forskellige tilløb til en ny eller fornyet forståelse af videnskabelig videns kontekstafhængighed indebærer.

3. Peer reviews - et kort overblik over litteraturen

Lige siden Merton gennemførte sin nu klassiske analyse af peer reviews i udvælgelsen af videnskabelige artikler, har denne metode til videnskabelig kvalitetssikring og -udvælgelse stået i centrum i videnskabssociologiske undersøgelser af videnskabelig kvalitet og kvalitetssikring. I den senere år har også informationsvidenskaben i stigende omfang beskæftiget sig med problemområdet, omend fokus oftest har været rettet mod mere

¹⁰ Begreberne mode 1 og mode 2 stammer fra Michael Gibbons 1994. Triple Helix-teorien findes hos i en række arbejder af Henry Etzkowitz og Loet Leydesdorff, se f.eks. 1997.

kvantitative metoder såsom citationsanalyser, impact-analyser og andre bibliometriske metoder.

Det er ikke hensigten her at give en større systematisk præsentation af litteraturen om peer reviews, men at give et billede af de væsentligste for- og bagdele ved denne kvalitative forskningsevalueringsmetode. Et af de mest omfattende nyere forsøg på at sammenfatte forskningen om peer reviews er Cicchetti's (1991) undersøgelse af peer reviews i forbindelse med bedømmelse af manuskripter og ansøgninger om forskningsprojekter. Ikke blot sammenfatter Cicchetti et stort antal peer review-undersøgelser fra en række forskellige fagområder, men hans artikel rummer også en lang sektion, hvor 34 fremtrædende forskere, der alle har beskæftiget sig med peer reviews i forskellige discipliner, diskuterer Cicchettis konklusioner. Cicchettis hovedkonklusion er, at der på tværs af discipliner ser ud til at være en forbløffende mangel på pålidelighed ved brug af peer reviews, især hvad angår den såkaldte positive vurdering, dvs. udvælgelsen af hvad der er videnskabelig værdifuldt. Den stigende konkurrence om begrænsede forskningsmidler og publiceringsmuligheder gør, at ringe pålidelighed i udvælgelsen af den værdifulde forskning rammer megen god og relevant forskning hårdt - omend der er stor forskel på konsekvenserne i forskellige fagområder. Dette får Cicchetti til at stille spørgsmålet: hvad vil en højere grad af pålidelighed betyde og kan den overhovedet opnås? Dette spørgsmål rejses også af Hemlin (1996), der har sammenfattet en række empiriske studier af peer reviews i en oversigtsartikel over litteraturen om forskningsevaluering fra 1996. Hemlin ser løsningen på problemer med pålideligheden i peer reviews i brugen af kompetente eksperter (Hemlin 1996, s.244) og ikke i de mange forslag om at forbedre bedømmelserne gennem oplæring af bedømmere, idet "[...] these attempts were successful only in the training case." (Hemlin 1996, s.234)

Ronald Kostoff (1995/99)¹¹ fra den amerikanske flådes forskningscenter, har foretaget en endnu mere omfattende sammenfatning af litteraturen om peer reviews som evalueringemetode. Dette som et led i en gennemgang af en række forskellige

¹¹ Ronald Kostoff 1995. Kostoff er tilknyttet Office of Naval Research, Arlington, Va., USA. Teksten findes også med i en nyere og opdateret udgave (1999) som er forsynet med en række bilag på internettet på adressen: <http://www.dtic.mil/dtic/kostoff/> Citaterne i denne artikel er alle hentet fra internetudgaven, men med sidehenvisninger til den trykte 1995-udgave. Dels af hensyn til tilgængelighed og genfindingsmuligheder og dels fordi Kostoff mig bekendt er den nyeste og mest omfattende præsentation af forskningsevalueringslitteraturen har jeg valgt at bringe lidt større og sammenhængende citater end ellers.

forskningsevalueringsmetoder, som en række offentlige amerikanske forskningsinstitutioner anvender ved fordeling af forskningsmidler (NSF, DOE, NIST, NAS). Herved kan Kostoff inddrage nogle problemstillinger, som fremkommer når forskningsbedømmelserne, her peer reviews, ses i deres institutionelle sammenhænge. Dette er sært nok et forhold som sjældent fastholdes i analyser af peer reviews. Med denne indgang til analysen kan Kostoff ikke undgå at inddrage de politiske og samfundsmæssige forhold som karakteriserer videnskabsudviklingen i dag. Det medfører også, at Kostoff kan inddrage dels organisatoriske og dels et videnssociologiske problemer i sin analyse.

Kostoffs sammenfatning af den mere snævret baserede litteratur om peer review-baserede forskningsevalueringer, går i samme retning som hos Cicchetti og Hemlin. Alle anerkender, at det er vanskeligt at forestille sig andre metoder til forskningsevaluering, der har samme fleksibilitet, mulighed for at gå i dybden, og som kan håndtere det vanskeligt definerbare område eller begreb: videnskabelig videns kvalitet. Desuden er peer reviews en nærmest uløselig del af den videnskabelige tradition og indgår på en kompleks måde i fastlæggelsen af, hvad der er videnskabelig viden. Men de slår samtidig ned på de mange problemer forbundet med denne metode, hvor den empirisk påviste lave pålidelighed mellem bedømmere fremhæves som det alvorligste problem. Kostoffs sammenfatning af litteraturen om problemer peer review metoden følger her, da den er den mest omfattende og grundige præsentation af problemstillingerne forbundet med peer review metoden og medtager de punkter som både Cicchetti og Hemlin opstiller. På baggrund af en undersøgelse foretaget af Chubin, sammenfatter Kostoff de *positive sider* ved metoden således:

- “1. an effective resource allocation mechanism;
2. an efficient resource allocator;
3. a promoter of science accountability;
4. a mechanism for policymakers to direct scientific effort;
5. a rational process;
6. a fair process;
7. a valid and reliable measure of scientific performance.”

hvorimod *problemerne* hvis vi ser bort fra diskussionen om pålidelighed først og fremmest er:

- “1. Partiality of peers to impact the outcome for non-technical reasons;

2. an 'Old Boy' network to protect established fields;
3. a 'Halo' effect for higher likelihood of funding for more visible scientists/ departments/institutions;
4. reviewers differ in criteria to assess and interpret;
5. the peer review process assumes agreement about what good research is, and what are promising opportunities.” (Kostoff 1995/99, s.40-42)

Denne oversigt, over problemerne ved denne evalueringsmetode, er der (som nævnt) stort set enighed om i størstedelen af litteraturen om peer reviews. De mange empiriske undersøgelser af peer reviews har ikke udviklet sig nævneværdigt metodisk og kan i det store og hele beskrives som en nærmest direkte fortsættelse af Mertons klassiske, men meget empiriske undersøgelse af Psysical Reviews artikeludvælgelse. De kritiske indvendinger mod peer review evalueringer, der igen og igen dukker op i disse studier, beror i stort omfang ifølge Kostoff på, at vurderingsmetoden har stor udbredelse og nyder almen anerkendelse som en troværdig bedømmelsesmetode. Dette samtidig med, at der i store dele af litteraturen herom findes og gentages nogle misforståelser af denne metodens og akkuratesse som måleinstrument, hvad der viser sig i en nogle gange helt overvældende fokusering på pålidelighedsproblemer (Cicchetti). På baggrund af litteraturen sammenfatter Kostoff de krav der stilles for at kunne udføre *peer review evalueringer af kvalitet*:

- “1. The method, organization and criteria for an evaluation should be chosen and adjusted to the particular evaluation situation;
2. Different levels of evaluation require different evaluation methods;
3. Program and project goals are important considerations when an evaluation study is carried out;
4. The basic motive behind an evaluation and the relationships between an evaluation and decision making should be openly communicated to all the parties involved;
5. The aims of an evaluation should be explicitly formulated;
6. The credibility of an evaluation should always be carefully established;
7. The prerequisites for the effective utilization of evaluation results should be taken into consideration in evaluation design.”

Kostoff går videre og konkluderer, at:

“Assuming these considerations have been taken into account, three of the most important intangible factors for a successful peer review are:

Motivation, Competence, and Independence. The review leader's motivation to conduct a technically credible review is the cornerstone of a successful review. The leader selects the reviewers, summarizes their comments, guides the questions and discussions in a panel review, and makes recommendations about whether the proposal should be funded. The quality of a review will never go beyond the competence of the reviewers.”

(Kostoff 1995/99, s.40)

Det helt afgørende for gode peer reviews er således “Motivation, Competence, and Independence” hos bedømmerne, fremhæver Kostoff. Men han går videre og inddrager nogle af de vigtigste organisatoriske problemer ved brug af peer reviews som bedømmelsesmetode. Det er først og fremmest de *omkostningsproblemer*, der uløselig er forbundet med peer reviews - både i form af forbrug af tid (og penge) som selve peer review-processen kræver, og de tidsomkostninger for både organisationen og den enkelte forsker, i form af ventetid på at få frembragt peer reviews. Kostoff bygger denne del af sin analyse på undersøgelser af peer review metoden som den er anvendt i store, ofte teknisk orienterede, forskningsorganisationer i USA, hvor peer review-metoden indgår som et vigtigt led i ‘research impact assessment’ og sammenfatter de organisatoriske problemer således:

“Another problem with peer review is cost. The true total costs of peer review can be considerable but tend to be ignored or understated in most reported cases. For serious panel-type peer reviews, where sufficient expertise is represented on the panels, total real costs will dominate direct costs by as much as an order of magnitude or more. The major contributor to total costs for either type of review is the time of all the players involved in executing the review. With high quality performers and reviewers, time costs are high, and the total review costs can be a non-negligible fraction of total program costs, especially for programs that are people intensive rather than hardware intensive.” (Kostoff 1999)

4. Organisatoriske og videnssociologiske perspektiver på peer reviews

Men Kostoff begrænser sig ikke kun til de organisatoriske problemer ved brug af denne metode. Det hyppigt omtalte pålidelighedsproblem ser han som værende af mindre betydning, idet han tilføjer diskussionen om peer review metodens 'bagside', *et central spørgsmål om metodens omkostninger i et forskningsorganisatorisk og videnssociologisk perspektiv*. Dette indebærer at omkostninger kan betragtes i forhold til hvorledes nye, uprøvede og kontroversielle ideer lever og overlever i en forskningsorganisation, hvor peer reviews hyppigt forekommer. Hvad angår disse omkostninger for organisationen skriver Kostoff:

“[...] another potentially negative aspect of peer review (is) not addressed by the literature; namely, whether the knowledge of periodically scheduled reviews would stifle the pursuit and presentation of very innovative but far-out ideas. Would performers be reluctant to present these ideas in a public forum, where the credibility of the performers could be challenged for these ideas? In other words, does the practice of peer review, and especially panel-based program peer review, effectively result in self-censorship of radical ideas? This is an area where research is needed to ascertain whether ideas have been suppressed in periodically reviewed programs, and then to determine how this problem could be surmounted if it exists.” (Kostoff 1999)

Kostoff åbner her diskussionen af nogle af de mere alvorlige problemer, der er forbundet med brug af peer review evalueringer i forbindelse med store programmer og forskningsorganisationer og som synes uundgåelig i forbindelse med den tiltagende forskningspolitiske styring og regulering. To centrale problemstillinger bliver meget synlige, når peer reviews skal bruges i en større forskningsorganisation og ikke, som i traditionelle peer reviews, næsten udelukkende med fokus på den enkelte forsker eller det enkelte forskningsprodukt (artikel). Omkostningerne ved at gennemføre høj-kvalitets evalueringer er så store, at de udgør en synlig del af et programs samlede omkostninger og måske endnu vigtigere, er der en ukendt men sandsynlig risiko for at periodiske peer reviews i en forskningsorganisation kan begrænse forskningens originalitet og lede til selvcensur.

I den nyere videnskabssociologi findes der enkelte forsøg på at overskride nogle af de begrænsninger, der findes i størstedelen af litteraturen om peer reviews, hvor diskussionen af problemet om pålidelighed i målingen stort set har overskygget alle andre problemer ved denne metode. To af kommentatorerne til Cicchetti's artikel, videnskabssociologerne Stephen

Cole og Steve Fuller, afviser ud fra en videnskabssociologisk tilgang til analysen af peer review, at det er tilstrækkeligt at begrænse diskussionen af denne evalueringsmetode til alene til produktet, artiklen, forskeren eller ansøgningen. Stephen Cole, der har beskæftiget sig med peer reviews og videnskabsudvikling, peger på, at Cicchetti's konklusioner i retning af forskellige tiltag der kan forøge pålideligheden bygger på en implicit antagelse om faglig konsensus i de enkelte videnskabsområder - en forudsætning som ifølge Cole er ganske uholdbar. Cole's eget nu næsten klassiske studie af vurdering af ansøgninger til forskningsrådene viser, at "because of the inherent lack of consensus on research frontier science, it is *inevitable* that many worthwhile proposals will be rejected and some proposals of little value will be funded. This was the major finding of my peer review study." (1981) Steve Fuller følger denne kritik af Cichetti op og peger på at Cicchettis antagelse om, at bedre pålidelighed i vurderingen af den enkelte forsker eller det enkelte forskningsresultat/produkt skulle betyde forbedret videnskabelig kvalitet, viser en manglende forståelse for de mange forskellige processer eller selektionsmekanismer der er i spil.

Hverken Cole eller Fuller afviser at peer review er en vigtig del af den videnskabelige bedømmelses- og udvælgelsesproces, tager de omvendt klart afstand fra de konklusioner på de mange samstemmende undersøgelser, der fremhæver problemerne med lav eller ringe pålidelighed ved peer reviews, som Cicchetti og mange andre lægger frem. Cicchetti og andre peger alle på forskellige tiltag som kan forbedre pålideligheden, såsom flere bedømmere, træning af disse og brug af andre mere kvantitative metoder til at supplere bedømmelsen. Tiltagene har det fælles, at de overfokuserer på spørgsmålet om pålidelighed, dvs. ensartethed i målingen og mulighed for forudsigelse, fremfor spørgsmålet om gyldighed, hvad der er forskning af god kvalitet. Netop dette spørgsmål er så omstridt på forskningsfronten i alle videnskabsområder, ikke mindst siden Kuhn (1970), at en overfokusering på pålidelighedsproblemet uhyre let vil kunne komme til at ske på bekostning af gyldighed. Den løsning på problemet om usikkerheden i peer reviews som Cole (1998) bla. peger på, er at fremme en dialog mellem bedømmere og bedømte, at forbedre begrundelser samt at skaffe flere publiceringsmuligheder. Selvom hverken Cole eller Fuller trækker nogle klare konsekvenser af deres videnskabssociologiske kritik af den traditionelle tilgang til peer review evalueringer, så peger deres kritik tydeligt nok på, at peer review som evalueringsredskab alene kan forstås som et led i en vidensproduktion og derved uløseligt forbundet med først og fremmest de organisatoriske og politisk sammenhænge, der er etableret omkring denne proces. Den implicite antagelse der ligger bag analyserne af peer review fra Merton til

Cicchetti er - forenklet sagt - at de omgivende vilkår for vidensproduktionen antages at være uforandrede. Cole og Fuller har påvist at denne status quo betragtning er videnssociologisk ganske uholdbar. Kostoff understreger dette med al ønskelig tydelighed, når han med baggrund i undersøgelser af peer review evalueringer i en række forskningspolitiske organisationer når frem til, at de økonomiske omkostninger ved denne metode og dens implicitte 'disciplinering' af forskningen er forhold, der er blevet meget synlige i disse organisatoriske sammenhænge og som derfor med tiden vil gøre sig gældende i alle de sammenhænge, hvor peer review evalueringer anvendes.

Merton har som nævnt tidligere været den mest fremtrædende videnskabssociolog, der har forsøgt at give en sociologisk forklaring på og begrundelse for, at videnskaben skulle forstås som et relativt lukket og selvregulerende system, med et internt system af normer, der via deres udbredelse og anerkendelse kunne regulere udvælgelsen af den bedste forskningskvalitet, fordeling af anerkendelse og andre ressourcer, på en måde der havde en udbredt accept internt såvel som eksternt. Peer review-systemet har her haft en afgørende betydning. Styrken i Mertons analyse har klart været, at denne har kunnet fastholde en fokusering på nogle vigtige strukturelle og organisatoriske sider af videnskaben set som et socialt system. Men antagelsen om den relative isolation i forhold til det omgivende samfund og ikke mindst de samfundspolitiske reguleringer blev der allerede sat spørgsmålstegn ved længe før Mertons analyse af Max Weber, som påpegede, at videnskaben i stigende omfang udgør et centralt element i det moderne samfunds udvikling, ikke mindst igennem en den fremvoksende rationalitet, der sigter mod kontrol og beherskelse af verden. Habermas har senere fulgt denne problematik op med sin diskussion af videnskabens dobbelte funktion i senkapitalismen dels som legitimerende dels som frigørende. I moderne samfundsteori hos Giddens og Beck er forestillingen om videnskaben som et isoleret system nærmest vendt på hovedet - denne opfattelse ses som udtryk for en mangelfuld forståelse af dynamikken og kompleksiteten i modernitetens udvikling. I begrebet 'risikosamfundet' kommer dette klart til udtryk, idet den samfundsmæssige anvendelse af videnskab og teknologi her indgår i en kompleks sammenhæng. Her er anvendelsen af videnskab og teknologi både en central del af forklaringen på problemer og konflikter i det moderne samfund, ofte i form af uventede følgevirkninger af anvendelsen, og en del af de mulige løsninger. Hvorledes dette har resulteret i ændringer i videnskaben internt og i forhold til politisk regulering har flere nyere

videnskabssociologiske studier søgt at belyse. Her kan nævnes begreberne Mode 1 og Mode 2 videnskab og Triple Helix¹².

5. Offentlig forskningsevaluering i Danmark - et eksempel på institutionelle forandringer i brug af peer reviews

Forestillingen om at videnskaben har en særlig samfundsmæssig placering eller status, der gør at den er nogenlunde modstandsdygtig overfor eksterne påvirkninger og politiske reguleringer kan der sættes en række forskellige men meget velbegrundede spørgsmålstejn ved. I dette afsnit er det hensigten at undersøge hvad der er sker med anvendelsen af den klassiske peer review model, når den anvendes i en moderne organisatorisk kompleks sammenhæng. Det sker ved at se nærmere på udviklingen i forskningsevaluering ved de danske universiteter i de sidste ca. 10 år.

¹² Mats Benner og Ulf Sandström har i artiklen *Institutionalizing the triple helix: research funding and norms in the academic system*, Research Policy 29 (2000) s. 291-301 undersøgt forandringer i det svenske system til finansiering af forskning og kan vise, at der udover den klassiske model, baseret på interne videnskabelige kvalitetsbedømmelser er fremkommet to nye modeller; en målrettet styring af forskning efter industriens interesser og en 'transinstitutional' model, der kombinerer disse.

I den hidtil mest omfattende analyse af udviklingen i evalueringen af den offentlige forskning i Danmark viser Foss Hansen and Borum (1999), hvorledes de særlige organisatoriske og politiske forhold omkring universitetsforskningen over en lang periode, betinger udviklingen af en særlig form for standardisering af forskningsevalueringen, den *modificerede 'peer-review' model*. Denne model vokser frem som et resultat af flere forhold. Danmark var meget sent ude med at tage forskningsevaluering i brug som styringsredskab og de politisk-organisatoriske forhold omkring universiteterne - især styrelseslovens decentrale magtstruktur - gjorde en politisk top-down styring kompliceret. En række initiativer fra ikke mindst OECD om udviklingen af en national forskningspolitik gjorde styring/kontrol og evaluering af dette område politisk aktuelt. Den modificerede 'peer-review' model som blev importeret fra Sverige, byggede på forskningsverdenens traditionelle standardmodel for kvalitetsbedømmelse i forbindelse med ansættelser, bevillinger, udgivelse af artikler og bøger mm., dvs. på 'peer reviews'. Denne traditionelle standard for at organisere en kvalitativ vurdering i forskningsverdenen, og indeholder ikke kriterier eller andre indholdsmæssige krav (Hansen & Borum s.309), disse er det indirekte overladt til de udvalgte bedømmere at opstille. Dette karakteriserer også den modificerede peer review model, men modsat den traditionelle anvendelse retter denne model sig ikke mod det enkelte produkt eller person, men mod forskningens organisatoriske rammer og samlede produktion.

De første forskningsevalueringer blev gennemført ved politisk påbud, dvs. ud fra en top-down beslutning og i de første 7-8 år fra starten i 1985 blev der gennemført et fåtal evalueringer - måske overraskende set i forhold til den omfattende debat som disse evalueringer afstedkom. Men valget af den modificerede peer review model fremfor de meget mere performance-orienterede tilgange til forskningsevaluering som var fremme i denne periode, bl.a. via OECD, er formodentlig en vigtig årsag til at der i perioden efter 1992 er sket en kraftig vækst i disse evalueringer. Dog tager disse nu udgangspunkt i institutionerne egne beslutninger (Hansen & Borum 310) - dvs. man kan tale om at modellen nu er blevet institutionaliseret som en intern institutionel standard. Samtidig hermed er der på det politiske niveau sket en holdningsændring henimod ønsket om at bruge mere formaliserede, kvantificerbare og sammenlignelige forskningsevalueringer med hovedvægt på performance, i form af en efterligning af den engelske model og meget lig den udvikling som uddannelsesevalueringen havde gennemgået i Danmark. Diskussionen er stadig åben ikke mindst i lyset af den overordnede udvikling i den offentlige sektor i Danmark, hvor kvalitetsudvikling og styring ved brug af benchmarking og performance measurement mere og mere kommer på dagsordenen og i et eller andet omfang er på vej ind i universiteternes hverdag i form af de udviklingskontrakter, der er ved at blive implementeret på universiteterne nu. Det vil

formodentlig indebære en stigende fokusering på institutionernes forskningsaktiviteter som dels vil tage form af overordnede politikformuleringer i de indgåede aftaler og internt rejse krav om udvikling af interne instrumenter til styring og udvikling af forskningen. På baggrund af en omfattende undersøgelse af en række danske statsinstitutioner, som har arbejdet med mål- og rammestyret i en årrække, har Jan Mouritsen (1997) gjort opmærksom på, at sådanne overgange til ledelsesformer, der bygger på mere kvantificerede og formaliserede interne processer ikke blot betyder effektivisering eller modernisering af institutionen.

“Tælligheden indebærer ikke blot at der sættes tal på et eller andet organisatorisk fænomen. Den indebærer især, at tallet er gjort til den *centrale karakteristik* heraf. Hermed sættes statslige institutioners beslutningskriterier på en simpel formel: Institutionen skal gøre *selvstændig* og dermed til en potentielt udskiftelig del af den statslige koncern. Klienten skal gøres til *kunde*, hvorfor aktuel efterspørgsel er vigtigere end potentiel efterspørgsel. Administratoren/sagsbehandleren skal gøres til *leder*, således at faglige kriterier skal medieres - omend næppe helt afløses - af hensynet til resultatet.” (Mouritsen 1997, s.152)

Forskningsevaluering er et modsætningsfyldt begreb, som sammenfatter en tradition for kritisk vurdering af ny viden, som er immanent i videnskaben, hvorigennem ny videnskabelig viden vinder eller ikke vinder anerkendelse som videnskab. Prioritering af ressourcer i form af penge eller stillinger har i nyere tid været baseret på evaluering eller bedømmelse af forskning. Traditionelt var de dominerende metode peer reviews, dvs. andre fagfolks faglige bedømmelse. Selv i de sammenhænge, hvor der var tale om at evalueringen eller bedømmelsen ikke ‘blot’ vedrørte ny viden men ressourcer, var der i princippet tale om en dialog baseret på et afgrænset terrain af faglighed - og begrundelserne for trufne beslutninger kunne derfor diskuteres ud fra argumentationens faglighed. På en måde er der ikke så meget nyt i at anvende forskningsevaluering på institutioner - det nye ligger, som Borum og Hansen påpeger, i den kobling mellem den interne tradition for at bruge forskningsevaluering og det politiske-administrative systems ønsker om at bruge forskningsevaluering som redskab til at kunne gribe ind i de meget rodfæstede traditioner for forskningsautonomi og institutionelt selvstyre¹³. Det nye der nu kommer ind i billedet er således den eksplicitte kobling til den

¹³ Lars Niklasson (1998) diskuterer med eksempler fra svenske universiteter den betydning disse evalueringer har for eksterne politiske forsøg på at gennemtvinge væsentlige ændringer i organisationskulturen på disse institutioner i retning af en mindre individualiseret og mere kollektiv holdning.

politisk-administrative styring¹⁴. Denne kobling betyder at nye dimensioner i forskningsevalueringen trækkes frem, idet den politisk-administrative styring som, ikke mindst i en New Public management-tidsalder, ofte lægger op til, at beslutninger der skal træffes ofte baserer sig på standardisering og kvantificering af information og ikke på kvalitative ekspertvurderinger¹⁵.

Foss Hansen og Borums analyse illustrerer hvorledes peer review som evalueringsmetode udvikler sig historisk i en specifik politisk og organisatorisk sammenhæng mod en form, der nærmest fremstår som en ny standard med så store indholdsmæssige ændringer, at man måske kan tale om en ny metode. Dette sker igennem en årrække på baggrund af en række ændringer i anvendelsen og brugen af peer reviews til forskningsevaluering, som placerer peer reviews i større organisatoriske strukturer. Disse vil, i sin anvendelse af denne metode nærmest automatisk, forsøge at udvikle en vis standardisering eller forenkling og vil ofte have gennemførelsen af peer reviews sammenkoblet med en række andre organisatoriske mål. Dette er, som det skulle fremgå af den forudgående gennemgang af litteraturen om peer reviews, en nærmest naturlig konsekvens af, at denne metode er så udbredt og anvendt i så forskellige sammenhænge som tilfældet er. Den grundlæggende usikkerhed der er forbundet med resultaterne af peer reviews og de betragtelige omkostninger, som kvalitetsudgaver af peer reviews medfører, er hver for sig forhold, der kan give anledning til forsøg på at rationalisere eller modernisere processen gennem at udvikle standarder ved brug af denne metode. Alternativt kan en række supplerende og mere indirekte metoder anvendes. Gibbons og Georghiou (1987) har set nærmere på de mere indirekte evalueringsmetoder, først og fremmest bibliometriske såsom citations- og publikationsopgørelser, relationsindikatorer mm., men vurderer, at disse ikke ser ud til at have opnået nogen større udbredelse, bl.a. pga. af en række uløste metodiske problemer. De ser et betydeligt større perspektiv i anvendelsen

¹⁴ Se f.eks. Björn Rombach, og Kerstin Sahlin-Andersson (eds.) *Från sanningsökande till styrmedel. Moderna utvärderingar i offentlig sektor*, 1995. Stockholm: Nerenius & Santerus Förlag. Titlen på denne bog (og indholdet) siger meget præcist hvad det er for radikale forandringer der er undervejs i den offentlige sektors anvendelse af evalueringer.

¹⁵ Men kvantificeringen er, udover at være udtryk for en mangel på tillid fra beslutningstagere til dem der producerer viden, samt en strategi fra eksperter og vidensproducenters side til at (gen)vinde tilliden, også en socialteknologi, som ved at omforme individuelle forhold ved personer til kvantificerbare klassifikationer, grupperinger osv. muliggør en omfattende udøvelse af politisk kontrol og magt. Kvantificering gennemsyrrer således den politiske kultur i det moderne samfund i et sådant omfang, at en række også dagligdags fænomener optræder i en kvantificeret form, som fremstår som naturgiven og uproblematisk. De historiske kampe der resulterede i kvantificeringen af f. eks. standarder for mål, vægt og penge, der er afgørende for vareudvekslingen, er totalt glemt. (Hansson 98, Porter 95).

af samfundsvidenskabelige metoder som surveyundersøgelser og interviews, for herigennem at indhente evaluering information. Endvidere peger de på, at udviklingen af formelle og måske især mere uformelle informationsnetværk forskere imellem, i høj grad fremmes af elektroniske kommunikationssystemer. I det omfang disse netværk af meget forskellig karakter går ind i at fremskaffe evaluering informationer og måske aktivt fremlægger deres vurderinger kan følgende spørgsmål blive yderst relevant: "What would happen, then, if the respective evaluations did not agree? Could such a event overturn the decisions of the peer review process?" (Gibbons og Georghiou 1987, s.70)

Den her skitserede udvikling er sket på baggrund af meget omfattende samfundsmæssige forandringer, der ikke er gået sporeløst hen over videnskabssystemet. Både den uhyre vækst i midlerne til videnskab og forskning, som kan iagttages i de udviklede industrialiserede lande over de seneste 15-20 år¹⁶, og de strukturelle forandringer i videnskabens og forskningens samfundsmæssige rolle¹⁷, som diskuteres i dag, peger på at de samfundsmæssige forudsætninger for at fastholde den klassiske peer review metode, som *Metoden* til at vurdere forskningskvalitet hos den enkelte forsker og forskningsorganisation, ikke længere er tilstede.

6. Nye muligheder for forskningsevaluering?

Kostoffs gennemgang af peer review-litteraturen endte med at pege på to centrale problemer, som fremkommer når man undersøger hvorledes denne bedømmelsesmetode fungerer i en organisatorisk ramme. Det ene problem er de omfattende omkostningerne i både tid og penge ved at gennemføre kvalitets-evalueringer på denne måde i forskningsorganisationer. Det andet og på en måde måske det vigtigste, er spørgsmålet om en hyppig brug af peer reviews i en organisation vil betyde en indbygget selvdisciplinering eller selvcensur som effektivt kan undertrykke nye ideer. Foss Hansen og Borum har vist med eksemplet fra Danmark, at sammenfaldet mellem ønsker om politisk og organisatorisk kontrol med forskningen kan udvikle sig i forskellige former for standardiseringer og rationaliseringer af metoden.

¹⁶ Den nyeste belysning af omfanget af den offentlige forskning kan for Danmarks del ses i rapporten, *Analyse af den offentlige forskningsproduktion i Danmark i de seneste 10 år - sammenlignet med udvalgte lande. 1999*. DTU 1999. Se også OECD 1997.

¹⁷ Se f.eks. Ulrich Beck's og Anthony Giddens 'diskussion herom i Ulrich Beck, Scott Lash, og Anthony Giddens 1994 og i Ulrich Beck 1999.

Både Kostoff og Foss Hansen og Borum peger på en ret nyopdukket problemstilling i forbindelse med anvendelse af peer reviews i en større organisatorisk sammenhæng. En sådan anvendelse kan let betyde, at der indarbejdes forsøg på at kvantificere og objektivisere dele af grundlaget for denne, i sit udgangspunkt, kvalitativt orienterede metode, for derigennem at tilbyde forenklede og overskuelige tilgange til både frembringelsen og analysen af materialet. Herved er det muligt at bruge selve proceduren omkring peer review evalueringer og - ikke mindst vigtigt - den anseelse der er forbundet med den klassiske udgave af peer reviews. Men i stedet for at anvende den klassiske kvalitative vurdering anvendes mere summariske vurderingsmetoder baseret på anvendelse af forskellige kvantitative teknikker, eksempelvis diverse statistiske opgørelse, publikationstillinger, citationsanalyser mm. Men gennem anvendelse af sådanne 'modificerede' udgaver af peer reviews mister denne sin entydige faglige karakter af at være en kvalitativ kvalitetsvurdering. Især Kostoff har advaret mod de tilløb selvdisciplinering af vidensproduktionen, som disse nye former for anvendelse af 'modificerede' udgaver af peer reviews kan medføre.

Det tiltrækkende i at søge at kvantificere peer review-processen i forbindelse med forskningsevalueringer ligger i, at den således kvantificerede viden kan frigøres fra sine lokale bindinger og tolkningssammenhænge. Porter (1995)¹⁸ har vist hvorledes det er muligt at se kvantificering som en kommunikationsstrategi, der skaber en form for objektivitet, som gør det muligt at anvende den kvantificerede viden på afstand og uden direkte involvering. Kvantificering og den ofte tilknyttede standardisering kan, som Brunsson (1998, s.213) påpeger, have utilsigtede negative konsekvenser overfor faglige værdier og normer. Udviklingen af standarder indenfor faglige professioner, der over lang tid selv har udviklet interne værdisæt og normer for det faglige arbejde, kan være direkte nedbrydende og destruktivt for disse værdier og normer.

Men de forskellige former for 'modificerede' peer reviews rummer nogle indbyggede modsætninger. Det vil ofte være nødvendigt at anvende en standardisering og kvantificering af forskningsevalueringen, når denne ikke længere kan begrænses til enkeltpersoners produkter, men må inddrage forskningens organisatoriske sammenhænge. Herved erstattes tidligere eksisterende tillidsrelationer med mere formelle og distancerende relationer. I sin klassiske form er peer review metoden baseret på, at denne metode kan producere en

¹⁸ "Perhaps most crucially, reliance on numbers and quantitative manipulations minimizes the need for intimate knowledge and personal trust. Quantification is well suited for communications that goes beyond the boundaries of locality and community." Theodore Porter 1995, s. ix-x.

afpersonificerede viden, idet netop konsensus blandt bedømmere antages at være lig en godkendelse af viden uafhængig af sin oprindelse, uafhængig af personlige forhold. Denne forestilling om objektiv videnskabelig viden som Merton fremhæver igen og igen har vist sig at være utrolig sejlivet. Selv kritikere af Merton som Fuller har, så vidt jeg kan se, ikke rettet kritikken præcist mod denne helt centrale antagelse hos Merton, der siger at videnskabelig viden frembringes og eksisterer som afpersonificeret viden, fri for alle sammenhænge, som et produkt eller mere præcist i en kapitalistisk økonomi, som en vare (human capital).¹⁹ Men samtidig antager Merton implicit, at der eksisterer et bagvedliggende sæt af tillidsrelationer, nemlig de tillidsrelationer, der er baseret på den fælles accept af CUDOS-normerne, hvor det afgørende er, at der er tillid til at en vurdering alene sker på baggrund af forskningsresultatet, dvs. på et rent upersonligt plan.

Men i den 'modificerede' udgave af peer review'et forsøges denne udtalte form for tillid erstattet af kvantitative metoder og standardiserede procedurer. Det er der som tidligere anført mange årsager til. Undersøgelser af udviklingen i en række vidensområder og vidensanvendelse peger på så ændrede vilkår for produktionen af viden, at det personorienterede udgangspunkt for forskningsevalueringen, i form af peer reviews, ikke længere kan stå uantastet. Viden udvikles i dag i højere og højere grad i teams, ved samarbejde, og kan ikke længere alene analyseres som et individuelt projekt. Endvidere har tilliden til videnskab og teknologi på et mere samfundsmæssigt niveau i dag undergået en radikal ændring, således at den videnskabelige videns tidligere uantastede autoritet nu offentligt anfægtes og udfordres - hvad der blandt andet kommer til udtryk i teorien om risikosamfundet (Beck, Giddens).

Filosoffen John Hardwig (1991) har i en artikel om tillid og viden med bla. eksempler fra matematikken også vist, at viden i langt de fleste tilfælde udvikles i samarbejde, teams, og at forudsætningerne for succes i denne vidensudvikling er gensidig personlig tillid mellem de deltagende.²⁰ Hardwig gør også opmærksom på, at peer reviews tilsyneladende intersubjektivitet og upersonlighed ikke har hindret en række forsøg på videnskabelig svindel de senere år, hvor den klassiske naturvidenskabelige gentagelse af forsøg som kontrol ikke

¹⁹ Dette kommer eksempelvis meget tydeligt til udtryk i økonomernes begreb 'Human Capital', hvor viden betragtes som produkter på markedet - på en måde med god ret set i netop denne historiske udviklings lys. En fremtrædende eksponent for den tilgang til analysen af viden som produkt/vare er Gary Becker 1993.

²⁰ "I have claimed that trust in the testimony of others is necessary to ground much of our knowledge, and this trust involves trust in the character of the testifiers." s. 702

længere fungerer (omkostningerne er for store, det er ikke meritgivende at gentage forsøg, det er i flere og flere tilfælde umuligt pga. laboratoriers og forskningsorganisationens omfang og tekniske konstruktion og ikke mindst de særlige sociale konstruktioner som de deltagende forskere og teknikere producerer i selve forskningsprocessen: Harwig 1991, s.706). Tillid til forskeren i bred forstand, dvs. som en person i samfundet og i en konkret forskningssammenhæng, bliver således et væsentligt element i den videnskabelige troværdighed.

“Science, then, is not completely different from other cooperative enterprises; the reliability of scientific testimony, like the reliability of most other testimony, ultimately depends on the reliability of the testifier [...] An untrusting, suspicious attitude would impede the growth of knowledge, perhaps without even substantially reducing the risk of unreliable testimony. Trust in one’s epistemic colleagues is not, then, a necessary evil. It is a positive value for any community of finite minds, provided only that this trust is not too often abused.”(Hardwig, s.707)

Den videnskabelige vidensproduktion er vanskelig at forestille sig uden nogle mekanismer, som er rettet mod samarbejde og åbenhed, hvor gensidig tillid udgør meget vigtig faktor, hvis betydning længe har været undervurderet i videnskabssociologien. Udviklingen i den interne videnskabelige vurderingsmetode, peer review, til et mere og mere omfattende og standardiseret system der retter sig mod kontrol og politisk styring af forskningen, kan medvirke til at synliggøre hvor betydningsfuldt et forhold, den interne, gensidige tillid er, i en proces der frembringer viden.

7. Konklusion: behov for nye former for peer reviews?

Den forudgående undersøgelse af den klassiske eller idealiserede model for bedømmelse eller evaluering af forskning, det klassiske peer review, har på en række afgørende områder rejst så store spørgsmål ved denne model for forskningsevaluering, at den ikke længere kan stå alene som *den centrale model* for forskningsevaluering. Peer review modellen har som vist baggrund i nogle forestillinger om videnskab og dens udøvere som er meget historisk bestemte - grundlaget er efterkrigstidens vestlige landes politiske erfaringer med på en og samme tid voldsom udvikling i anvendelse af videnskab og teknologi og opgøret med videnskabens politiske anvendelse i totalitære systemer, Nazi-tyskland og Sovjetunionen, først og fremmest formuleret af Merton i CUDOS-normerne. Men vilkårene for videnskab og

forskning har ændret sig meget siden. Om man taler om Triple Helix (Etzkowitz, Leydesdorff), Mode 1 - 2 forskning (Gibbons) eller videnskabens funktion i risikosamfundet (Beck, Giddens) så er hovedpointen, at Mertons 'idealtipe' for videnskab som et selvstændigt område eller system, ikke længere har samfundsmæssig gyldighed.

En række studier af peer review-metoden har overbevisende vist, at det kan være forbundet med endog meget store vanskeligheder og omkostninger at gennemføre peer reviews baserede forskningsevalueringer, der kan give blot rimelige kvalitetsresultater. Kostoff fremhæver direkte problemet med at sikre, at nye forskningsområder og nye faglige tilgange og ideer kan komme til orde i denne i bund og grund klart konservative model for evaluering. Dertil kommer det hastigt stigende omfang af nye samarbejdsformer i forskningen. De seneste år har fremvist en voldsom vækst i nye organisationsformer, det være forskningsteams, grupper, virtuelle og lokale forskningsnetværk osv., alle former hvor de rent individuelle bidrag næppe eller kun i begrænset omfang lader sig registrere på traditionel vis og derved kunne udsættes for den individuelle forskningsevaluering. Endelig er der de høje omkostninger i både tid og penge, som forbundet med at gennemføre peer review evalueringer på et højt fagligt niveau (Kostoff) - og som i sig selv kan være årsag til forsøg på at rationalisere og standardisere forskningsevalueringer.

Vi har med et eksempel fra danske universiteter set, hvorledes disse forskellige problemer, der er forbundet med gennemførelsen af peer review forskningsevalueringer, kombineret med et stigende politisk pres for at gennemføre sådanne evalueringer, har resulteret i udviklingen af en særlig form for standardiseret forskningsevaluering, 'det modificerede peer review'. Foss Hansen og Borums undersøgelse af denne 'model' har vist, at udviklingen af denne evalueringsmodel ikke er et resultat af en rationel planlægning, men bedst kan beskrives ved hjælp af institutionel teori som "[...] an interplay between mimetic, normative and coercive logics, performed in organizational fields of multiple actors carrying different institutional traditions and interests." (s.324) På den anden side har Lars Engwall sammenfattet erfaringer fra danske og svenske institutions- og forskningsevalueringer og påpeget, at disse evalueringer har udviklet sig i retning af, at "evaluations should be carried out in dialogue form, and that they should apply a structural perspective rather than focus on individual actors." (Engwall 1998)

På universitets- og forskningsinstitutionsområdet, men også i forbindelse med evaluering af forskningsprogrammer, har udviklingen af tiltag til styring og regulering af forskning i de senere år taget form af forskellige revisioner af peer review modellen. Disse modificerede

eller reviderede udgaver skal tjene til at synliggøre institutionens forskningsaktiviteter overfor aftagere eller de politiske beslutningstagere, eller i en mere revisionsmæssig forstand begrunde anvendelsen af særlige forskningsmidler. Det er ikke så underligt, at peer review modellens ydre former er søgt bevaret. Den klassiske peer review model har lang tid anseelse i den politiske verden og i offentligheden for at være den bedste garant for vurdering af videnskabelig kvalitet. Men anvendelsen af disse 'modificerede' former for peer reviews kommer i en situation, hvor der sker så omfattende forandringer i forskningens organisering og samfundsmæssige funktion, let til at modarbejde de intentioner, som oprindeligt lå bag evalueringerne, nemlig at bidrage til udvikle forskningens faglige kvalitet. Disse evalueringer initieres hyppigt 'oppefra', fra institutionsledelsen eller højere oppe i det politiske hierarki og gennemføres ved brug af eksterne evaluatore (ofte fra udlandet med begrundelse at sikre uafhængighed og objektivitet). Som Foss Hansen og Borum påpeger kan kombinationen af institutionel modstand og politisk pres for øget kvalitetskontrol lede frem til udvikling en ny standardisering, evt. i et sæt af standarder for hvad forskningsevaluering er.

“The best preventive measure to guard against this risk would seem to be slowly, but surely to develop locally rooted systems that inspire confidence both internally and externally.” (Engwall 1998, s.22)

Er rationalisering og standardisering af forskningsevalueringer den bedste løsning, når forskning er en kreativ samarbejdsproces, præget af normer og traditioner for fri og uafhængig kritik mellem ligemænd, og at ledelse af kreative processer er en meget kompliceret og følsom sag? Eller handler det modsat om at forsøge at få igangsat en proces på dette evalueringsområde, der ser på evaluering som 'praktisk hermeneutik' (Schwandt 1997). Eller som en form for aktionsforskning, hvor de nære processer meget mere er i fokus? Spørgsmålet er om de former for forskningsevaluering, der har udviklet sig i Danmark som den modificerede peer review model, ikke er u hensigtsmæssige - eller endog måske direkte skadelige - set i forhold til det erklærede mål, en ledelse af forskning der fremmer kreativitet og dermed kvalitet. En central årsag hertil er den tætte - omend nogle gange ikke så klart udtalte - kobling mellem evaluering og kontrol, regulering.

Af hensyn til at sikre de bedst mulige vilkår for udvikling af ny viden er det vigtigt at få (gen)etableret klare adskillelser mellem vurdering af viden og vidensudvikling og den styring og ledelse af de organisatoriske og politiske omgivelser, denne viden fungerer i og ikke kan være foruden. Dette er en forudsætning for at forskningsevalueringen i form af peer reviews mister karakteren af at være en ekstern kontrolfunktion og bliver i langt højere grad et redskab

til den interne udvikling og styring af vidensproduktionen (Gibbons og Georghiou 1987, s.70).

I princippet burde det være en enkel sag at anvende en sådan tilgang til forskningsevaluering i forskermiljøer - ikke mindst fordi evaluering eller bedømmelse af aktiviteter sker løbende i den interne videnskabelige kontrolproces! Men måske kan nye perspektiver komme frem hvis man vender tilbage til begrebet peer reviews i dets oprindelige betydning som bedømmelse gennem ligemænd. Der er klart nogle perspektiver ved at tænke i deltager-involverende evalueringer (empowerment, aktionsforskning), hvor dialogen mellem de evaluerede og evaluator er i fokus, og hvor det er proces mere end produkt der sigtes efter. Nogle af de nye teorier om videnskabens ændrede rolle peger gennem deres fokusering på udvidelser i opfattelsen af videnskab indirekte på nogle af disse forhold. Gibbons betoner det transdisciplinære, samarbejdet på tværs og med andre ikke-forskere i vidensproduktionen, Etzkowitz og Leydesdorff i Triple Helix om vidensudvikling i netværk mellem forskellige organisationer og Ravaetz taler om 'post-normal science' og 'extended peer communities', hvor videnskaben overskrider sine traditionelle grænser. Et første skridt i denne retning kunne være at anvende peer reviews som en løbende, intern debat- og kommunikationsform i en forskningsorganisation, hvor kontrolaspekterne ved peer review evalueringer var fjernet og ført over i andre organisatoriske sammenhænge!

Litteraturliste

- Beck, Ulrich, Scott Lash, og Anthony Giddens 1994. *Reflexive modernization. Politics, tradition and aesthetics in the modern social order*. Cambridge: Polity Press.
- Beck, Ulrich 1999. *World Risk Society*. Cambridge: Polity Press.
- Brunsson, Nils (ed.) 1998. *Standardisering*. Stockholm: Nerenius & Santerus
- Cole, Stephen 1981. *Chance and Consensus in Peer Review*. Science Vol. 214, no. 881.
- Cole, Stephen 1998. *How does Peer Review Work and how can it be Improved*. Minerva no. 36.
- Engwall, Lars 1998. *Experiences From Evaluations*. in Tempelaar, Dirk T., Finn Wiedersheim-Paul, og Elving Gunnarsson (eds.) *Educational Innovation in Economics and Business II*. Dordrecht: Kluwer Academic Publishers.
- Foss Hansen, Hanne og Finn Borum 1999. *The construction and standardization of evaluation. The case of the Danish University Sector*. Evaluation Vol. 5, no. 3.
- Fuller, Steve 1997. *Science*. Buckingham: Open University Press.
- Gaskins, Richard H. 1993. *Burdens of proof in modern discourse*. New Haven: Yale University Press.
- Gibbons, Michael og Luke Georghiou 1987. *Evaluation of research. A selection of current practices*. Paris: OECD.
- 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 Publications.
- Habermas, Juergen 1968. *Erkenntnis und Interesse*. Frankfurt am Main: Suhrkamp.
- Hagstrom, Warren O. 1965. "Scientific community." New York: Macmillan Company.
- Hansson, Finn 1998. *Kvantitativ viden i anvendt samfundsvidenskab. Nogle videnssociologiske overvejelser*. København: Handelshøjskolens Bibliotek.
- Hardwig, John 1991. *The Role of Trust in Knowledge*. Journal of Philosophy Vol. 88.
- Hemlin, Sven 1996. *Research on Research Evaluation*. Social Epistemology Vol. 10, no. 2.
- Kostoff, Ronald N. 1995. *Handbook of research impact assessment*. Springfield, Va..
- Kostoff, Ronald N. 1999. *Handbook of Research Impact Assessment*. Office of Naval Research - <http://www.dtic.mil/dtic/kostoff/Handweb.html>.
- Kuhn, Thomas S. 1970. *The structure of scientific revolutions*. Chicago: University of Chicago Press
- Locke, Simon 1999. *Constructing "the beginning". Discourse of creation science*. Mahwah© N.J.: Lawrence Erlbaum Associates© Inc.

- Merton, Robert K. 1968a. *Social theory and social structure*. New York.
- Merton, Robert K. 1968b. *The Matthew Effect in Science. The reward and communication systems of science are considered*. *Science* Vol. 159, no. 3810.
- Merton, Robert K. 1973. *Sociology of science. Theoretical and empirical investigations*. Chicago, Ill..
- Mouritsen, Jan 1997. *Tællelighedens regime. Synlighed, ansvarlighed og økonomistyring gennem mål og rammer i statslige institutioner*. København: Jurist- og Økonomiforbundets Forlag.
- Niklasson, Lars 1998. *A Cultural Revolution in the Universities. The Possible Uses of Rational Choice Models*. *Evaluation* no. S. 278-293.
- OECD 1999. *OECD science, technology and industry scoreboard. Benchmarking knowledge-based economies*. Paris.
- Porter, Theodore M. 1995. *Trust in Numbers. The pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.
- Rombach, Björn og Kerstin Sahlin-Andersson (eds.) 1995. *Från sanningsökande till styrmedel.moderna utvärderingar i offentlig sektor*. Stockholm: Nerenius & Santerus Förlag.
- Weber, Max 1992. *Wissenschaft als Beruf 1917/1919. Politik als Beruf 1919*. Tübingen: Mohr.
- Weber, Max 1999. *Verdi og handling*. Oslo: Pax.
- Schwandt, Thomas A. 1997. *Evaluation as Practical Hermeneutics*. *Evaluation* Vol. 5, no. 1.
- Sohn-Rethel, Alfred 1970. *Geistige und körperliche Arbeit. Zur Theorie der gesellschaftlichen Synthesis*. Frankfurt am Main: Suhrkamp Verlag.
- Sohn-Rethel, Alfred 1971. *Warenform und Denkform*. Frankfurt am Main: Europäische Verlagsanstalt.
- Tamm, Ditlev 1996. *Dansk retshistorie*. København: Jurist- og Økonomiforbundets Forlag.
- Whitley, Richard 1984. *Intellectual and social organization of the sciences*. Oxford: Clarendon Press.
- Whitley, Richard 2000. *The intellectual and social organization of the sciences*. Oxford: Oxford University Press.