

Recension

Liv Langfeldt

Decision-making in expert panels evaluating research: Constraints, processes and bias
(Oslo 2002) 267 pp.

Peer review är en viktig mekanism för fördelning av forskningsresurser inom akademiska forskningsfält. I Sverige har denna form för kollegial bedömning blivit dominerande i takt med att stora delar av sektorsforskningen har rådifierats. Det inflytande som tidigare tillkom myndigheternas handläggare har överförts till forskare och till kollegialt organiserade processer. I detta avseende har det under det senaste decenniet skett en akademisering av FoU-politiken.

Metoden att utnyttja kolleger med expertkompetens som grindvakter för vetenskapliga nyheter har gamla anor och användes från början för att bedöma manuskript till artiklar för vetenskapliga tidskrifter. Tanken var att söka upp den eller de kolleger som var mest kompetenta på området och låta dem ge sina (anonyma och ej offentliga) synpunkter för att hjälpa tidskriftens redaktör. Med vetenskapens växt fordrades arbetskonomi även för redaktörerna. Förutsatt att dessa var tillräckligt informerade om forskningsläget och om bedömarens kompetens kunde hävdas att manuskripten granskades (rättvist) innan de publicerades. Granskningsarbetet kan också sägas vara en viktig del av själva forskningsprocessen.

Idén om kollegial granskning kom också väl till pass när forskningsråden började organiseras i olika länder. Ställda inför risken att förlora en del av sin autonomi var det många forskare som var misstänksamma mot råden och önskade att nya resurser fortsatt skulle fördelas via fakultetskollegiet. Genom att forskningsråden etablerade bedömningskommittéer och lät fakulteterna välja representanter med uppgift att utse ledamöter till råden hölls den kollegiala idén levande och forskarsamhället accepterade modellen med forskningsråd. Tanken var att forskningsråden skulle premiera de bästa idéerna och ge forskarna bättre villkor för sin verksamhet.

Peer review i forskningsråden gäller ansökningar där enskilda forskare begär anslag för en specifik projektidé. Uppgiften att bedöma en ansökan är väl avgränsad och bedömningen görs med hjälp av ett antal frågeställningar om genomförbarhet, forskarens (forskargruppens) kompetens och grad av angelägenhet med hänsyn till det aktuella forskningsfältets strategiska läge. Många forskningsråd använder metoden att skicka ut ansökningar på remiss till de kolleger som kan antas vara bäst skickade att bedöma ansökan, men av olika (bl.a. praktiska) skäl har man också tillämpat metoden att tillsätta en mer permanent kommitté som under en bestämd tidsperiod bedömer alla ansökningar. Fördelen med den senare metoden är inte minst att det blir möjligt att göra en rankning av ansökningarna. Remissmetoden innebär att bedömarena bara ser en eller några få ansökningar och därför inte kan gradera. Istället fordrar det att det hos finansören finns starka *programme managers* som kan väga samman remisserna på ett konstruktivt sätt.

Redan kommittésystemet gör således att det sker en viss avvikelse från idealet om att bedöma ansökningarna med hjälp av den bästa expertisen på det specifika området. Remissmetoden (och tidskriftssystemet) bygger nämligen på den paradoxala idén att den lilla gruppen av experter inom en specialisering skall kunna ge varandra konstruktiv kritik. Detta antagande är inte självklart: Antingen är man varandras konkurrenter i en knivskarp kamp om pengarna och då är det troligt att omdömet kan svikta (negativ bias) eller är man vapenbröder och oförmögna att ha sådan distans att kritiska synpunkter skulle komma fram (positiv bias). Empiriska studier av hur forskningsråd i USA arbetar har visat att det är stor risk att det blir fel på ett eller annat sätt. Risken att få fel bedömare, dvs. att få någon som har svårt att ge en förutsättningslös granskning, drabbar ungefär en tredjedel av ansökningarna. Med ett annat val av remissperson är det högst sannolikt att många ansökningar skulle ha rönt ett annat öde.

Så är det i forskarsamfundets lilla liv av familjära konflikter om fördelning av resurser. Med det lilla livet menas att striderna handlar om enstaka projekt och enstaka artiklar. Ibland kan familjen drabbas av någon som inte förstår eller har kompetens att begripa och då blir det fel, men för det mesta fungerar det som det var tänkt. Svårare är att tänka sig att det skulle vara på samma sätt i forskningens storsamhälle. Där handlar det om att många klaners knog och stretanden sätts under lupp och bedöms av några få upplysta, s.k. experter som dock inte kan vara experter på allt som försiggår i storsamhället.

Storsamhällets peer review handlar om uppföljning och utvärdering av redan utförd eller pågående forskning. Det lilla livets legitimitet brukas då för processer där politiker och beslutsfattare har behov av underlag om hur det stora livet skall administreras. Expertmetoden expanderar då från mikronivån till att nyttjas för makro-relaterade problemställningar. I de senare fallen handlar det om att bedöma hela forskningsområden, discipliner, stora och långvariga forskningsprogram, institutioner, universitet och forskningsinstitut. Ofrånkomligen kommer forskare inom ett antal specialiseringar att drabbas av skevheter i så måtto att de utsätts för bedömningar av experter som inte har någon som helst, eller endast liten, kompetens att bedöma just deras område.

Den här förändringen står i fokus för NIFU-forskaren Liv Langfelts intressanta avhandling *Decison-making in expert panels evaluating research*. Hon frågar sig vad som händer med expertbedömningar när de används för utvärderingar på större och vidare forskningsområden. I mångt och mycket är detta frågan om en pionjärinsats. Trots att peer review sedan tre decennier varit en modell för utvärdering av resursfördelning i forskningssystemet har det knappast genomförts några djuplodande och systematiska studier av hur denna mekanism fungerar. Merparten av litteraturen på området har framkommit till följd av att enskilda forskare stött på problem och att dessa utifrån sina egna personliga erfarenheter genomfört fallstudier för att belysa de klagomål som man haft att komma med.

Är man det minsta bekant med forskarsamhället vet man att det finns missnöje med bedömningssystemet. En vanlig uppfattning bland forskare är att peer review är bra, men att det tillämpas på fel sätt av forskningsråden. De som blivit utsatta för utvärderare anser inte sällan att dessa ger utrymme för tillfällighetsbetonade och ogrundade uppfattningar. Misstanken att utvärderare värnar om sina egna skyddslingar, skriver upp sitt eget område och att de ofta skriver allmänt och svårtolkat om sådant de inte har egen kunskap om är vitt spridd i litteraturen. Mot denna bakgrund är det välgörande att Langfeldt i avhandlingen visar att ståndpunkter som hävdar att det finns olika former av kognitiv eller annan bias förutsätter uppfattningar om forskningskvalitet som är ganska rudimentära och naiva till sin karaktär.

Här gör Langfeldt en viktig insats som bör få konsekvenser för den fortsatta forskningen på området.

Langfeldt väljer en forskningsmetod som bygger på omfattande datainsamling via dokument och intervjuer av sex stycken viktiga norska utvärderingar (ex post), två av dessa handlar om forskningsfält, två om större forskningsprogram och två om forskningsinstitut. Hon intervjuar i princip alla som varit inblandade som utvärderare och sekreterare i utvärderingarna. De konflikter och beslutssituationer som varit aktuella i de olika utvärderingarna återskapas mycket noga och detta ger läsaren en förståelse av hur arbetet i utvärderingar ofta är strukturerat till att resultera i rätt så mediokra resultat. På ett belysande sätt gör hon en ingående beskrivning av hur utvärderarna valts ut, vilka områden de täcker, hur utvärderingsarbetet organiserats, vilka kriterier olika utvärderare anslutit sig till, och vilka strategier som använts.

Hennes resultat är kontroversiella eftersom de är ganska så nedslående. De mest framträdande kännetecknen är att utvärderingsrapporterna innehåller vaga och svårtolkade värderingar som tvingar läsaren att läsa mellan raderna, att de sällan eller aldrig använder explicita kriterier för bedömningarna och att utvärderarna tenderar att försöka komma till slutsatser och rekommendationer som understödjer de forskare man granskar. Till stor följer detta av att utvärderingarna arbetat med en strikt arbetsfördelning, sällan har överlappande kompetenser och i alltför liten utsträckning ger sig tid eller har kraft att diskutera kriterier och krav. Avhandlingen visar att utvärderare inte förmår förklara hur de använt olika kvalitetskriterier. Möjligen skulle kunna hävdas att det i takt med vetenskapens specialisering blivit allt svårare att hitta experter som är idealiska för uppgiften att bedöma hela forskningsfält, forskningsprogram eller institutioner.

Mot denna bakgrund är det inte svårt att förstå att det sällan blir annat än konsensusprocesser i utvärderingar. Även tidspressen gör att man inriktar sig på att konflikter skall elimineras eller tonas ned. Konflikter förekommer nämligen, men konsensus skapas under processens gång varefter panelerna bestämmer sig för att hävda det som är minsta gemensamma nämnare, dvs. att i alla fall grundforskningen bör få mer stöd. Detta skulle tala för att forskare har svårt att kritiskt värdera varandras insatser i sammanhang som är offentliga och ej anonymiserade. Skrået värnar om sina medlemmar. Langfeldt finner att de flesta utvärderare ser som sin roll att hjälpa den forskning som man granskar. Lojaliteten till olika områden är verksam och hon framhåller att det där finns ett potentiellt skevhetsproblem. Om sådan lojalitet är nödvändig för att kunna delta i en utvärdering finns inte heller anledning förvänta annat än stödjande och mildt kritiserande utvärderingar. Om så är fallet har beslutsfattare utanför det akademiska systemet kanske inte mycket att hämta från utvärderingsmaskinen.

Ulf Sandström
Docent och forskningsledare

Institutet för studier av utbildning och forskning (SISTER)
Stockholm