

Sosiologisk årbok blir 'tidsskrift med referee'

En diskusjon inklusive råd til våre fagkonsulenter

Per Otnes

Bruken av *referees*, heretter fagkonsulenter, har til oppgave å kvalitetsbedømme innsendte vitenskapelige artikler for fagtidsskrifter. *Sosiologisk årbok* (SÅ) har i 15 årganger ikke benyttet slike systematisk og regelmessig. Vi har rådspurt sakkyndige konsulenter når vi har vært i tvil, men ellers som regel nøydt oss med bedømmelser fra medlemmer av redaksjonen eller av årbokas råd, alltid minst to. Vi planlegger nå en overgang til systematisk bruk av eksterne konsulenter for alle originalartikler vi mottar. Vi ønsker altså å bli 'tidsskrift med referee' for framtida, fordi det i vide kretser anses som et viktig kjennetegn på faglig standard, og tilsvarende, 'tidsskrift uten referee' blir sett på som tegn på usikker eller sviktende standard. Vi er et stykke på vei enig i denne vurdering, dog med noen nyanseringer nevnt under.

Ordningen har i mange år vært svært mye diskutert, forsvart og kritisert, noe den som søker på internett under *peer review system* lett kan forvise seg om. Vår anvendelse bygger mest på arbeidene til professor Gerhard Frölich i Linz, særlig artikkelen "Anonym kritikk"

publisert i denne årboka, også trykt i Frölich 2002, som imidlertid refererer til et meget stort antall andre arbeider om emnet. Frölichs artikkel skiller seg ut i nett-litteraturen, ikke bare ved sin grundighet, men ikke mindre ved at den diskuterer ordningen på et prinsipielt, vitenskapsteoretisk grunnlag, ikke bare som en type praktisk ordning med *ad hoc* fordeler og ulemper. Dette er ikke anledningen for bredere diskusjon, bare noen hovedpunkter for å begrunne SÅs nyordning.

I litt gammelmodig fagterminologi kan man si at referee- eller konsulentordningens manifeste eu-funksjon er å sikre høy faglig standard, et mål alle kan slutte seg til. Hvorvidt dette er eneste eller viktigste funksjon kan diskuteres, likeledes om det ikke fins dysfunksjoner.

Når det gjelder ordningens – eller ordningenes, for det fins varianter – praktiske gjennomføring prøver mange tidsskrifter å gi inntrykk av at det dreier seg nærmest om en slags *kvantitativ måling* av faglig kvalitet, delt opp i en sjekklister med en rekke enkeltdimensjoner, som f.eks. teoretisk fordyping, metodisk adekvans, bredde og kvalitet på lesning av teori og tidligere forskning, problemstillingens klarhet, logisk gang i diskusjonen fram mot en konklusjon, som igjen skal vurderes etter klarhet, fruktbarhet, faglig relevans osv (se f.eks. Smith 1990). Alt forsøksvis skalert på tre- eller firepunkts skalaer av typen god– brukbar–mindre brukbar–svak. Og alt dette skal lede fram mot en samlende konklusjon, igjen skalert, av typen trykkes uten endringer– trykkes med mindre endringer–kan trykkes etter omfattende endring– bør ikke trykkes. Konsulenten bes ofte også om å avslutte med en kortere, kvalitativ karakteristikk av artikkelen.

Øvrige vanlige råd til konsulenter er f.eks. å tilstrebe nøytralitet i forhold til ulike skoleretninger, herunder egen personlig tilhørighet, i forhold til (gjetninger om) hvem forfatter er eller hvilket miljø han eller hun tilhører osv. – alt sammen godt, velment og rimelig nok.

For vår del vil vi se ordningen som en i bunn og grunn *kvalitativ vurdering*, med unntak bare for konklusjonens tre- eller firepunkts

'ordinalskala', som krever rangering. Konsulenter som ønsker det, er velkomne til å følge en sjekkliste som nevnt, men fortrinnsvis med hovedvekt på en kvalitativ, språklig karakteristik av en artikkels viktigste sterke og svakere sider.

Vi vil også minne om at ordningen er viktig, men likevel bare én blant flere vurderingsformer av vitenskapelig produksjon på ulike nivå. Den skiller seg ikke *prinsipielt* fra vurderinger av typen eksamensprestasjoner som kurs- eller hovedoppgaver, fra doktoravhandlinger eller – forelesninger, fra vurdering av søkeres produksjon ved stillingssøknader, eller fra kollegadiskusjoner av forhåndsutvalgte tekster presentert på nasjonale eller internasjonale konferanser. Hvis den skiller seg ut, er det kanskje mer i det kvantitative: Ledende fagtidsskrifter kan trykke opptil 60-70 originalartikler pr. år, og refusjonsraten vites å være høy for flere, over 90 % skal forekomme f.eks. innen psykologi. Dette gir vink om at tallet på vurderte i vår sammenheng sannsynligvis langt overstiger de andre nevne, med mulig unntak bare for eksamensprestasjonenes nødvendigvis mer inkluderende begynnernivå.

En særlig omstridt side ved ordningen er *graden av anonymitet*; man skiller mellom *single*, *double* og endog *triple blind*: Konsulentens identitet holdes skjult for forfatteren, men ikke omvendt, begges identitet er skjult for den andre, og endelig forfatters og konsulents identitet(er) holdes skjult også for det behandelende redaksjonsmedlem. Inspirasjonen er åpenbar, fra medisinsk og samfunnsvitenskapelig eksperimentell praksis, hvor det kan være grunn til å frykte at deltakernes ubevisste holdninger vil prege resultatet. Med andre ord en ny bestrebelse på å gi ordningen et preg av det rent kvantitative, nesten naturvitenskapelige og dermed per formodning udiskutable.

Det bør minnes om at det fins berømte unntak fra tanken om at ekstern konsulentbruk er nødvendig for å sikre faglig kvalitet – Webers *Archiv*, Durkheims *Année*, Bourdieus *Actes*, kan ikke ha brukt ordnin-

Sosiologisk årbok blir 'tidsskrift med referee'

gen, iallfall ikke konsekvent. Liketil *Tidsskrift for samfunnsforskning* innførte den først etter flere tiårs utgivelser.

Hovedsvakheten ved ordningen er at *den trekker faglig diskusjon ut av den bredere fagoffentlighet*, sterkere jo større grad av personhemmelighet – 'blinds' – som praktiseres. Hensikten er god for så vidt som det gjelder å beskytte kolleger som får negativ vurdering – 'slakt' – fra å bli 'hengt ut'. Hensikten er imidlertid utpreget dårlig for så vidt som den hindrer bredere meningsutveksling om faglig omstridte emner. Her skiller den seg også markant fra andre vurderingsordninger: Ved stillingssøknader og eksamener fins det klageadgang og bedømmernes identitet er kjent for de bedømte. For doktordisputaser, den eldste og vel mest ansette vurderingsformen, er åpenheten nærmest ekstrem idet ikke bare doktoranden, men også opponentene må presentere sine standpunkt utførlig, i full åpenhet overfor en nærværende, oftest velkvalifisert forsamling som har full anledning til å delta, 'opponere *ex audiorio*'. At ikke bare forfattere, men også konsulenter kan trenge beskyttelse mot det nag som kan oppstå etter 'slakt', er derfor mindre opplagt: Hvorfor bare for publisering, og ikke for eksamener, disputaser, fagkonferanser, og annen faglig meningsutveksling?

En annen svakhet er at ordningen synes å bygge på systematisk *mistro* kolleger imellom. Mistro enten til at kolleger flest holder høyt nok faglig nivå eller er i stand til å innse sine egne svakheter. Eller endog mistro til kollegers gode hensikter, i den grad det antas at større andeler innsendte artikler er produsert kynisk av kolleger som søker faglig gevinst, anerkjennelse, gjennom å få på trykk bidrag de må vite eller ane er underlødige. Eller, med et eksempel fra fag som har idéen om en 'least publishable unit': en forskningsprosess presenteres ikke som samlet monografi, men oppdelt på det størst mulige antall små enkeltartikler det anses mulig å få fagpressen til å publisere hver for seg. Man publiserer mer for å øke sin egen publiseringsportefølje enn for å bidra til fagenes framgang.

Er slike systematiske former for mistro egentlig nødvendige? Bør vi ikke heller som hovedregel heller tro vel om våre kollegers kvaliteter og dømmekraft? Dette har vært, og er SÅ-redaksjonens standpunkt. Og kan ikke mistroen hvis den blir systematisk gjennomført liketil *framkalle* nettopp den kynisme som den tar sikte på å fjerne (*'self-fulfilling prophesy'* med et nytt litt gammelmodig faguttrykk)? Og hvis mistroen blir mer allmen, hvorfor skal tidsskriftredaksjonene selv og deres utvalgte konsulenter være systematisk unntatt fra den? Det er velkjent at skjulte og mindre edle hensikter forekommer også her. Ingen fører kontroll med at 'tidsskrifter med referee' faktisk følger ordningen til punkt og prikke; tar hensyn til konsulenters ytringer, informerer forfattere om utfall, og konsulenter om andre konsulenters synspunkter ved 'sprik'. Og alle vet at et 'skjønnsomt' valg av konsulent nærmest kan bestemme sluttvurderingen – 'trykkes/ikke trykkes'.

Nå har lignende kritikk i alle år vært reist mot bedømmelsen f.eks. ved stillingsansettelser, hvor både betenkningens form og valget av bedømmelseskomité kan begrense slik fri og fair faglig konkurranse som sikrer at den beste vinner¹. Når man likevel beholder ordningen, er det fordi den vel må anses som i stort sett og i det lange løp *forholdsvis* fair, og bedre enn mulige erstatninger (som imidlertid i liten om noen grad utredes). Tilsvarende for konsulentordningen – den kan misbrukes, men trolig ikke som oftest. Og forutsatt grundig og velvillig innsats gir den ikke sjelden bidrag til å heve artikkelers faglige nivå, helt i samsvar med dens manifeste funksjon.

Frölich forøvrig gode artikkel er inne på tanken om at konsulentordningen kan fungere nærmest som en type sensur. Det mener vi er en tendesiell, og kanskje tendensiøs overdrivelse. Da er det mer nærliggende å minne om Janteloven som mulig metafor for utarting, f.eks 'Du skal ikke tro du er noe mer enn Oss'. Fagmiljøene er ofte

¹ Her er det i seg selv noe betenkelig ved at flere og flere av professoratene f.eks. ved vårt institutt i Oslo besettes ved opprykk etter individuell søknad og lukket søknadsbehandling, ikke etter åpen konkurranse mellom flere søkere.

Sosiologisk årbok blir 'tidsskrift med referee'

ikke ubetinget enhetlige, men preget av blandet konsens og dissens, ofte sentrert om persongrupper eller enkeltinstitusjoner, 'skoleretninger', noen åpne for andres synspunkter, andre mer lukkede og avvisende – endog sekterisme forekommer. Jante var en liten by, og det kan finnes svært mange slike, selv om det ikke er hovedregelen, også metaforisk brukt om fagmiljøer. Tankegangen minner om at tidsskriftene gjerne kan være mange og ulike, særlig innenfor et raskt voksende fagområde som sosiologien.

I tillegg bør nevnes noen praktiske ulemper med ordningen: Den er tid- og kostnadskreven, noe som vil tyngre mer for små og fattige tidsskrifter enn for de store, rike og veletablerte.

Ingen i litteraturen om refereeordningene synes å ha vært inne på tanken om at den i noen trekk kan ligne en domstol – det søkes fakta og felles dommer om 'tiltalte', dvs. fagtekster og deres kvalitet. En referee er slik sett en slags domsmann eller jurymedlem, eller kanskje snarere en innkalt sakkyndig, eller et oppnevnt medlem av en voldgiftskommisjon. Det ville nok vekke en viss oppsikt dersom noen av disse verv kunne utøves anonymt – det ville bryte med grunnleggende rettsprinsipper: En tiltalt eller part skal vite hva anklage eller påstand gjelder, og bevis som ikke legges åpent fram under saksbehandlingen, betraktes som ikke-eksisterende.

Den konklusjon vi trekker for egen del legger hovedvekten på dette med at faglig meningsutveksling i størst mulig grad bør foregå i den videst mulige offentlighet (for å si det nesten med Bentham, for en gangs skyld). Med andre ord, vi innfører eksternt konsulentvurdering for alle faglige originalbidrag. For slike bidrag som har vært gjenstand for åpen og kvalifisert vurdering av andre, like- eller høyere stilte faglige fora slik f.eks. doktorforelesninger, er kravet om eksternt konsulentbehandling mindre absolutt og avhengig av at det kan fremmes eksplisitt tvil om kvaliteten på arbeidet. Kommentarer, bokanmeldelser o.l. er som vanlig som regel unntatt.

Av hensyn til den åpne diskusjon holder vi som hovedregel også på åpen – '*no blind no way*' – konsulentbehandling. Faglitteraturen viser at f.eks. det ansette BMJ, *British medical journal*, forholdsvis nylig gikk over fra '*blind*' til åpen vurdering. Vi gjør unntak for forfattere eller konsulenter som *kan begrunne* særlig behov for anonymitet. Vi mener, igjen som hovedregel, at en faglig vurdering som syns den må skjule sitt opphav taper i troverdighet hvis den ikke kan begrunne det motsatte. En negativ konklusjon er ikke aleine noen tilstrekkelig grunn. Hoch Habermas!

Vi har nevnt tanken på refereeordningen som en slags nesten sensur, og motsatt som entydig kvalitetsstempel. Iallfall både angripes og forsvares ordningen med stor heftighet. Heftigheten i forsvaret bunner kanskje i frykt for å bli 'holdt utafor' vitenskapelig publisering, på svakt eller ukjent faglig grunnlag. Heftigheten i forsvaret derimot skyldes nok behovet for et pålitelig vitenskapelig kvalitetsmål som er lett forståelig også utafor fagmiljøene, vel særlig blant forskningsentreprenører og blant anvendelsesinteresser i næringsliv og offentlig administrasjon. Vi tror dette gjenspeiler en grunnleggende tvetydighet i holdningen til vitenskap og forskning: På den ene side er dette en eminent produktivkraft, slik bl.a. den unge Habermas påpekte, en som kan gi avkastning utafor alle proporsjoner. På den andre kan den jo ikke sjelden gi overraskelser, f.eks. forskningsfunn som vil medføre store tap, kanskje desimering av hele næringsgreiner – tenk på tobakksindustrien og kreftforskningen eller industriutslipp og miljøødeleggelser. Eller bare støtende eller 'upassende' resultater – liketil Darwin skal jo ha vært redd for å støte samtidens troende. Leibniz' 'beste av alle mulige verdener' sett fra entreprenørnivåenes ståsted ville i denne sammenheng bestå i å ha et robust, akseptert, 'objektivt' og lettforståelig mål som tillater dem å styrke det første, og forkaste det siste, men med all mulig diskresjon.

Litteratur

Frölich, Gerhard (2002): "Anonyme Kritik. Peer Review auf dem Prüfstand der empirisch-theoretischen Wissenschaftsforschung", i Pripp, E.: *Drehscheibe E-Mitteuropa. Information: Produzenten, Vermittler, Nutzer. Die gemeinsame Zukunft*, Wien, Phoibos Verlag, Biblos-Schriften 173. Også tilgjengelig fra forfatteren direkte, gerhard.froelich@jku.at

Smith, Alan Jay (1990): "The task of the referee", <http://www.computer.org/tpami/taskofreferee.html>