

Collin m fl har inte underlag för sina slutsatser

TOBIAS FRIDHOLM

I *Ekonomisk Debatt* 4/2021 presenterar Collin, Sandström och Wennberg en analys av närings- och innovationspolitiska utvärderingar 2009–19. Studien har fått stor uppmärksamhet genom sin huvudsakliga slutsats: att utvärderingar utförda av konsulter avsevärt oftare än andra innehåller positiva omdömen. Min bedömning är att de saknar underlag för den slutsatsen, eftersom de inte har kontrollerat för vilken typ av utvärdering som genomförts. Jag påvisar här en mycket tydlig arbetsdelning mellan olika utförare och utvärderingstyper och hävdar att utvärderingstyper som primärt utförs av konsulter har högre sannolikhet att generera positiva omdömen oavsett vem som utför utvärderingen. Min slutsats är att studien troligen framför allt mäter korrelationen mellan utvärderingstyp och positiva omdömen. Det skulle innebära att Collin m fl:s policyrekommendationer är felaktiga.

Min huvudsakliga invändning mot Collin m fl är att de inte tar tillräcklig hänsyn till mångfalden inom utvärderingsfältet. Begreppet ”utvärdering” används för alltifrån kvantitativa kontrafaktiska analyser till insatser där utvärderaren främst fungerar som bollplank under genomförandet av ett projekt. I Europeiska kommissionens guide för utvärdering av strukturfonderna (European Commission 2014) kallas den förstnämnda typen för *impact evaluation* och den senare för *implementation evaluation*, inte olik *development evaluation* (OECD 2010). Vilken måttstock som bedömningen i utvärderingen görs emot varierar också. Oftast görs den mot interventionens syfte och

mål. Effekttutvärderingar där bedömningen görs genom empiriska jämförelser, exempelvis med aktörer som inte påverkats av interventionen, är mindre vanliga. Det förekommer också utvärderingar där måttstocken är experters bedömningar av nivån på prestationerna. Det är avsevärt svårare att erhålla positiva omdömen i utvärderingar baserade på empiriska jämförelser. Därför är det problematiskt att Collin m fl inte har kontrollerat sitt underlag utifrån vilken typ av utvärdering som gjorts och vilken måttstock som använts.

Upplägget av en utvärderingsprocess är i stort sett detsamma som för en forskningsstudie, med ett tillägg: En utvärdering inbegriper per definition en normativ bedömning, exempelvis om en intervention är ”lyckad”. En generell struktur på utvärderingsprocessen är således:

- 1) Val av utvärderingens syfte och frågeställningar
- 2) Val av måttstock (mot vilken/vilka kriterier görs den normativa bedömningen?)
- 3) Val av metod och material
- 4) Genomförande
- 5) Resultat

Precis som i forskning är resultaten i en väl genomförd utvärdering i stort sett en funktion av stegen 1–4. Collin m fl har dock valt att bortse från steg 1 och 2 (och 4, men det är förståeligt) och är helt inriktade på steg 3 och 5. Möjligen styrdes upplägget av praktiska omständigheter, men precis som på forskningsseminarier kan diskussionen och slutsatserna landa snett om man inte också tar hänsyn till syfte och frågeställning och i det här fallet även vilken måttstock som använts.

Jag noterar också att Collin m fl inte har genomfört någon bortfallsanalys. Deras metod att samla in utvärderingar från myndigheters hemsidor innebär att sammansättningen av deras material

REPLIK

Tobias Fridholm är postdok i innovationsstudier vid Internationella handelshögskolan i Jönköping och utvärderingskonsult på Sweco. tobias.fridholm@ju.se

är direkt kopplad till myndigheternas benägenhet att publicera utvärderingar där. Det beteendet kan inte antas vara slumpmässigt utan är troligen systematiskt och kan variera över tid.

Metod

Jag har försökt att återskapa Collin m fl:s underlag genom att precis som dem samla in alla utvärderingar på Energimyndighetens, Tillväxtanalys, Tillväxtverkets och Vinnovas hemsidor för åren 2009–19. I likhet med Collin m fl utslöt jag ex ante-utvärderingar (utvärderingar som görs före en intervention, exempelvis för att uppskatta förväntade utfall och utmaningar). Jag fann 209 utvärderingsrapporter jämfört med 241 i ursprungsstudien (av vilka 110 slumpades fram och ingick i analysen). I den masteruppsats (Collin 2020) som deras studie baseras på indikeras att de har funnit fler utvärderingar än mig på Energimyndighetens hemsida och möjligen färre på Tillväxtverkets. Givet antalet utvärderingar och tydligheten i de resultat jag visar nedan ligger diskrepanserna dock troligen inom acceptabla gränser. Jag kodade aktörstyp på samma sätt som jag förutsätter att Collin m fl har gjort men tydliggjorde kategorin ”forskare/expert” för att inkludera de fall där utvärderaren uppenbart anlitas för att med sin individuella expertis genomföra *peer review*-liknande utvärderingar, vilket i samtliga fall gällde Vinnovaprogram. I vissa fall förekommer att två aktörstyper har delat på genomförandet. Kategoriseringen är då gjord utifrån huvudsaklig utförare.

Jag operationaliserade steg 1, Syfte och frågeställning, genom att dela in utvärderingarna i de gängse kategorier som förekommer inom utvärderingsfältet (jämför European Commission 2014; Sandberg och Faugert 2020; Vedung 1998). Inom utvärderingen används ofta termerna *summativ* och *formativ* för att beskriva syfte och karaktär. Sum-

mativ avser en kontrollerande ansats medan *formativ* handlar om att bidra till lärande. Jag kategoriserade utvärderingarna enligt följande:

- 1) Effekttvärdering. Genomförs vanligen en längre tid efter att interventionen har avslutats. Vanligen framför allt *summativ* men kan ha betydande *formativa* inslag.
- 2) Metautvärdering. Sammanvägning och analys av ett flertal utvärderingar, ibland kompletterat med viss ny empiri. Balansen mellan *summativ* och *formativt* kan variera avsevärt.
- 3) Ex post-utvärdering. Genomförs ganska kort efter att interventionen har avslutats. Vanligen både *summativ* och *formativ*.
- 4) Etapputvärdering eller motsvarande. Genomförs under pågående intervention. Vanligen både *summativ* och *formativ*, med tyngdpunkt på det senare som underlag till eventuella anpassningar under återstående period. Ibland görs den inför beslut om fortsatt finansiering och är då mer *summativ*.
- 5) Löpande utvärdering (även kallad följeforskning, på engelska *on-going evaluation* och snarlik det som kallas *development evaluation*). Särskilt vanligt i strukturfonderna, där det genomförs både på program- och projektnivå. Utvärderaren följer interventionen och gör både utvärderande och processtödande insatser. Balansen mellan de två rollerna varierar betydligt mellan olika utvärderingar, men de *formativa* inslagen är nästan alltid stora, särskilt på projektnivå.
- 6) Uppföljning. Genomförs under pågående intervention och innebär vanligen insamling av empiri för att kontrollera om interventionen kan förväntas bidra som önskat. Syftet är oftast *formativt*.

Steg 2, måttstock, operationaliserade jag enligt följande:

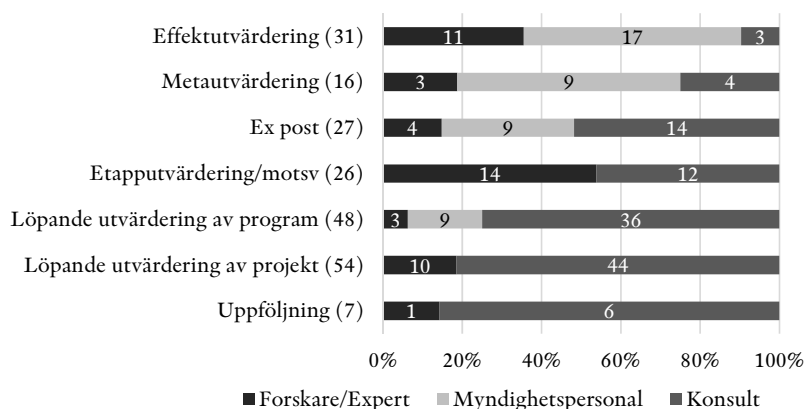
- 1) Måluppfyllelse: Uppfyllande av syfte, mål och/eller effektlogik.¹
- 2) Expertutlåtande: Jämförelse med andra insatser och aktörer utifrån sakkunniga individers uppfattningar, dvs liknande *peer review*.
- 3) Empirisk jämförelse: Jämförelse med andra insatser och aktörer utifrån empiriska analyser och skattningar. Kan inkludera kontrafaktisk analys.

Särskilt värt att notera är att den första typen av måttstock, måluppfyllelse, utgår – något hårdraget – från interventionen själv. Expertutlåtanden och empiriska jämförelser baseras på företeelser som de ansvariga för interventionen knappast har något inflytande över. Klassificeringen är inte knivskarp, en intervention kan naturligtvis ha som explicit mål att ge större effekter för deltagarna än för de som inte deltar, men det är nästan alltid bara ett av flera mål.

Vad gäller bortfallsanalysen har jag nöjt mig med att koda vilket år respektive utvärdering har publicerats. En stor variation över tid torde indikera systematiska skevheter i materialet eftersom det totala antalet genomförda utvärderingar sannolikt inte har förändrats så mycket från år till år.

Resultat

Figur 1 visar fördelningen mellan aktörstyper per typ av utvärdering. Jag har ordnat dem efter hur jag uppfattar balansen mellan summativa och formativa inslag i respektive utvärderingstyp. Som figuren visar råder det en långtgående arbetsdelning mellan aktörstyperna beroende på vilken typ av utvärdering som avses. Forskare, experter och myndighetspersonal dominerar i utvärderingar med mer summativa inslag och i utvärderingar som genomförs en period efter att en intervention avslutats. Utvärderingar som brukar ha betydande formativa inslag och genomförs under interventionen eller precis efter att den avslutats utförs däremot huvudsakligen

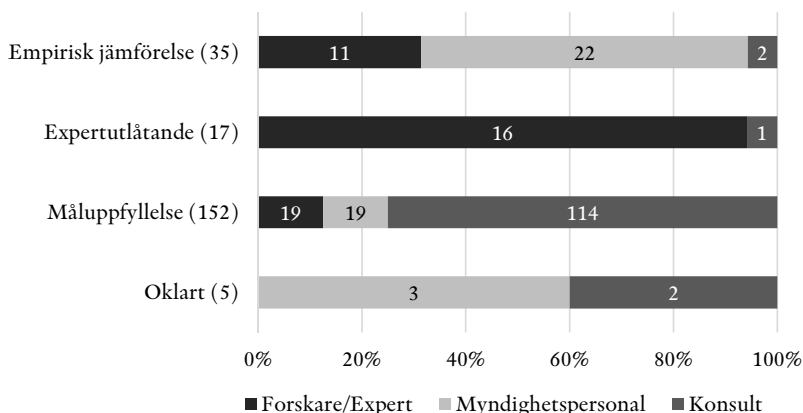


Figur 1
Aktörstyp per typ av utvärdering

Källa: Egna beräkningar.

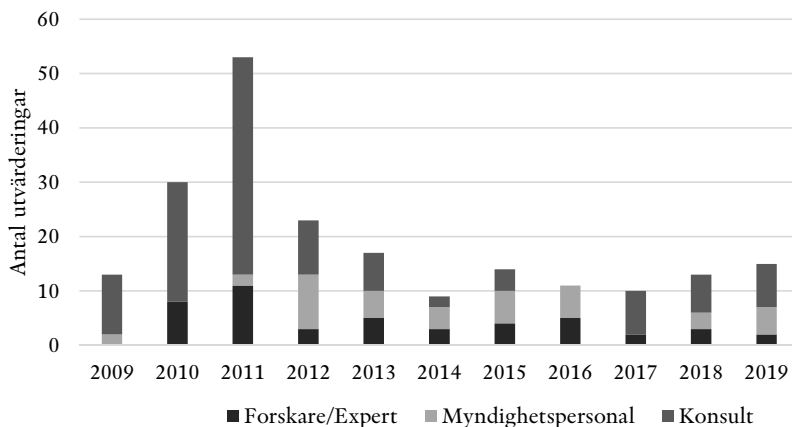
¹ Effektlogik, liksom de väsentligen synonyma termerna programlogik, projektlogik, interventionslogik och förändringsteori, är en slags karta över vilka avtryck en intervention bör göra för att uppfylla sitt syfte. På kort sikt ska resultat och utfall uppkomma, vilka på kortare eller längre sikt – men ofta bortom interventionens genomförandeperiod – ska bidra till effekter, vilka i sin tur ska bidra till syftets uppfyllande.

Figur 2
Aktörstyp per typ av måttstock



Källa: Egna beräkningar.

Figur 3
Antal utvärderingar per aktörstyp och år



Källa: Egna beräkningar.

av konsulter. Det bör särskilt uppmärksammas att hela 80 av de 119 utvärderingar (67 procent) som genomförts av konsulter har handlat om löpande utvärdering.

Figur 2 visar motsvarande fördelning mellan aktörstyp och vilken typ av måttstock som utvärderingen baserats på. I det här fallet är arbetsdelningen mellan olika aktörstyper ännu mer genomdriven. Empiriska jämförelser görs nästan uteslutande av myndighetspersonal eller forskare/experten, medan konsulter i 96 procent av fallen baserar utvärderingen på i vilken grad inter-

ventionen uppfyller (eller kan förväntas uppfylla) sitt syfte, mål och/eller effektivitet. Expertutlåtanden görs av naturliga skäl av forskare/experten. Figuren visar också att 73 procent av utvärderingarna görs utifrån måluppfyllelsen och att endast 17 procent baseras på empiriska jämförelser. (Det ska dock noteras att kategoriseringen är gjord utifrån huvudsaklig måttstock; i stora utvärderingar av måluppfyllelsekaraktär förekommer det inte sällan inslag av empiriska jämförelser och expertutlåtanden.)

Figur 3 visar antal utvärderingar per år och aktörstyp. Figuren visar att

myndigheterna publicerat betydligt fler utvärderingar på sina hemsidor under periodens tidiga år än senare. I praktiken förklaras detta av att Tillväxtverket har publicerat ett mycket stort antal rapporter från löpande utvärderingar på projekt- och programnivå från struktur-fondsperioden 2007–13.

Slutsatser

Resultaten visar att Collin m fl har baserat sin analys på ett mycket skevt underlag. De borde ha kontrollerat för vilken typ av utvärdering som genomförts och vilken måttstock som respektive utvärdering har använt. De borde även ha reflekterat över bortfallet. Som framgår av min undersökning är det oklart om de mäter korrelationen mellan utförare och omdöme, mellan typ av utvärdering och omdöme, eller mellan typ av måttstock och omdöme. Således saknar de av allt att döma underlag för sin huvudsakliga slutsats. Det är extra beklagligt eftersom de har slagit upp den stort i media och den potentiellt har stora policyimplikationer.

Min hypotes är att Collin m fl:s utfall framför allt korrelerar med vilken måttstock som använts. Det är utan tvekan avsevärt enklare att få ett positivt omdöme om måttstocken är interventionens syfte och mål än om det handlar om empiriska jämförelser med aktörer som inte berörts av interventionen. Det gäller i synnerhet om utvärderingen görs under tiden som interventionen pågår eller om den precis har avslutats; de ur ett samhällsligt perspektiv mest meningsfulla eller mätbara avtrycken (nya produkter på marknaden, väsentligt ökad omsättning etc) behöver som regel några år för att uppkomma. Bedömningen blir då ofta huruvida interventionen genererat aktiviteter och resultat som pekar i rätt riktning. Det är också betydelsefullt hur de ansvariga för interventionen har formulerat syfte, mål och effektlogik – vilket förvisso inte innebär att låga am-

bitioner per automatik leder till positiva omdömen, men den diskussionen rymms inte här. Mot den bakgrunden är det särskilt noterbart att hela två tredjedelar av konsultrapporterna rör löpande utvärdering, den mest aparta av de sex former av utvärdering jag kategoriserat materialet i.

Min skepsis mot Collin m fl grundas också i mina konsulterfarenheter. Jag har svårt att föreställa mig att forskare eller myndighetspersonal så särskilt ofta skulle komma till annorlunda slutsatser i sak än konsulterna ifall de ställdes inför samma uppgift. Däremot kan det nog finnas en viss systematisk variation i uttrycksätten mellan olika aktörstyper. Det vore knappast heller en affärsmässigt framgångsrik strategi som konsult att bete sig så oprofessionellt, kundstyrt eller rentav inkompetent som resultaten och resonemangen i Collin m fl implicerar.

Avslutande reflektioner

Diskussionen knyter an till en större debatt om vilka effekter innovations- och näringslivspolitiska insatser har. Som förhoppningsvis har framgått påstår jag varken det ena eller andra vad gäller nytan av dessa insatser. Däremot hoppas jag att inlägget bidrar till en mer konstruktiv debatt, eftersom jag visar att merparten av de utvärderingar som görs inte kan förväntas hålla ”vetenskaplig kvalitet” såsom det begreppet används av Collin m fl (men de kan förvisso göra det i andra avseenden, som genom kvalitetsäkrade underlag och ändamålsenliga metodval).

Jag kan inte nog understryka vikten av att ta hänsyn till i vilket syfte en utvärdering genomförs. Som framgår är endast en mindre andel av utvärderingarna effektutvärderingar baserade på empiriska jämförelser. Dessa framställs ibland i debatten som definitionen av en legitim utvärdering. Jag anser att det är en orimlig utgångspunkt, och har i den

ståndpunkten stöd av såväl tunga policydokument (European Commission 2014; OECD 2010) som standardreferenser i den akademiska utvärderingslitteraturen (Patton 2008; Vedung 1998). Det går utmärkt att argumentera för att fler och bättre effektutvärderingar bör genomföras på området – det är en åsikt jag själv för övrigt delar. Men det är närmast oetiskt att argumentera för den ståndpunkten genom att implicit hävda att alla andra utvärderingar, dvs 80–90 procent av de utvärderingar som görs, skulle vara mer eller mindre illegitima bara för att de inte bidrar tillräckligt i det avseendet. Det går också bra att hävda vikten av att utvärderare ska kunna agera självständigt och ha uppdragsgivare som förstår betydelsen av att låta empirin styra slutsatserna (vilket samtliga professionella utvärderare gissningsvis instämmer i), men jag anser inte att Collin m fl förmår påvisa en systematisk variation mellan hur olika typer av utvärderare – eller snarare konstellationer av utvärderare och uppdragsgivare – agerar i det avseendet.

Det går på motsvarande vis att argumentera för att det görs alltför många utvärderingar som är av till synes mindre betydelse. Men den diskussionen behöver föras på ett nyanserat sätt, med utgångspunkt i vad utvärderingarna faktiskt är avsedda att bidra med. Exempelvis är det rimligt att en myndighet vill veta vad en bidragsfinansierande intervention bidrar till och förstå varför den gör det. Vad kommer ut av projekten och varför blev det så? Vilken roll spelar interventionen för att främja kunskapsutbyte mellan universitet och företag? Och så vidare. Svaren på den typen av frågor fordrar inte nödvändigtvis en kontrafaktisk analys. Med det sagt förväntar även jag mig att alla myndigheter känner ansvar att fokusera på insatser som ger samhällsligt mervärde och därmed ser behov av att förstå hur och när det uppstår.

En försvarande omständighet är att termen ”effekt” används på två väsensskilda sätt inom utvärderingsfältet. Det är en självklarhet för oss inom området men jag förstår att det kan ställa till det i den externa kommunikationen. I effektutvärderingar baserade på empiriska jämförelser syftar ordet effekt på interventionens additionalitet, alltså dess mervärde. I utvärderingar där syfte, mål och effektlogik är måttstocken, syftar det däremot på ett skede i effektlogiken. En projektfinsierande insats genererar resultat som i ett senare skede bidrar till att effekter uppstår. Exempelvis kanske ett projekt genererar en pusselbit till en teknisk lösning (ett resultat), som bidrar till att ett företag fyra år senare sätter en bättre produkt på marknaden som ökar dess omsättning (effekt). Detta säger inte att interventionen i sig har ett mervärde jämfört med vad som skulle ha skett utan den. Det är viktigt att vi utvärderare uttrycker oss klart och tydligt om vad vi syftar på när vi talar om effekter och jag tror att vi har förbättringspotential i det avseendet.

REFERENSER

- Collin, E (2020), *Vem utvärderar utvärderarna? En tvärsnittsstudie av innovationspolitiska utvärderingar*, masteruppsats, Statsvetenskapliga institutionen, Uppsala universitet.
- Collin, E, C Sandström och K Wennberg (2021), ”Utvärderingar av näringspolitik – en intressekonflikt mellan myndigheter, konsultföretag, politik och skattebetalare?”, *Ekonomisk Debatt*, årg 49, nr 4, s 30–41.
- European Commission (2014), ”Guidance Document on Monitoring and Evaluation: The Programming Period 2014-2020”, European Commission, Bryssel.
- OECD (2010), *Quality Standards for Development Evaluation*, OECD, Paris.
- Patton, M Q (2008), *Utilization-Focused Evaluation*, fjärde upplagan, SAGE, London.
- Sandberg, B och S Faugert (2020), *Perspektiv på utvärdering*, fjärde upplagan, Studentlitteratur, Lund.
- Vedung, E (1998), *Utvärdering i politik och förvaltning*, andra upplagan, Studentlitteratur, Lund.