

Made in Australia



Är Arbetsförmedlingens nya rating-system embryot till en kostnadseffektiv arbetsförmedling med fokus på resultat?

Rapporten har tagits fram av Svenskt Näringslivs arbetsmarknadsenhet,
Edward Hamilton, i samarbete med Katrin Östman vid konsultföretaget Vostra.

Innehåll

Rapportens huvudbudskap.....	2
Argumentation.....	4
Arbetsförmedlingens rating av externa utförare i Stöd och Matchning.....	5
Resultatsystemets konstruktion och beräkning.....	5
Ratingen beräknas i fyra steg.....	6
Profilering av deltagarna.....	8
Beräkning av leverantörernas ratingvärde.....	8
Sammanfattning: Bedömning av rating-systemets tillförlitlighet.....	9
Internationella jämförelser.....	10
Jämförelse med den australiensiska ratingens konstruktion.....	10
Jämförelse med den brittiska ratingens konstruktion.....	11
Hur Australien använder ratingen för att påverka det nationella resultatet.....	12
Hur använder Storbritannien ratingen för att påverka de nationella resultaten?.....	13
Sammanfattning länder-exempel.....	14
Vad säger leverantörerna i upphandlingen Stöd & Matchning?.....	18
Utvecklingsområden mot bakgrund av modellanalys och leverantörernas erfarenheter.....	23
Slutsatser.....	26
Källor.....	27

Rapportens huvudbudskap

Utredningen om framtidens arbetsförmedling ska presentera en delrapport till regeringen i oktober 2017 som särskilt innefattar användandet och utformningen av upphandling av externa förmedlings- och utbildningsleverantörer. Översynen leds av Cecilia Fahlberg.

Svenskt Näringsliv har länge efterlyst tydligare resultatstyrning i såväl de programinsatser som upphandlas, som för myndighetens egna verksamheter. Fokus för en modern arbetsförmedling bör vara resultat, transparenta jämförelser och tydlig uppföljning. Kostnadseffektivitet måste få en större genomslagskraft i systemet. Uppdraget till myndigheten måste därtill tydliggöras och renodlas. AF måste göras om i grunden.

Svenskt Näringsliv och våra medlemsorganisationers principprogram i den referensgrupp som är knuten till utredningen är tydligt: ett fungerande *rating-system* är centralt. Programinsatser som inte levererar goda resultat ska avvecklas. Aktörer som presterar bra ska få fler kunder. Leverantörer som inte levererar goda resultat ska gallras ut. AF:s egna enheter och interna insatser ska mätas och bedömas på lika villkor.

Vi har sedan flera år lyft fram länder-exempel som Australien och Storbritannien som föregångsländer i dessa avseenden. Forskning styrker tydligt Australiens *rating-system* påverkan på resultat och kostnadseffektivitet. Detta är ett grundläggande element som måste fungera för en större förändring och förbättring av hela AF-systemet.

I dagsläget upphandlar Arbetsförmedlingen (AF) förmedlingstjänster för ett värde om cirka 1,5 miljarder kronor per år. Under föregående år uppgick utgifterna till uppemot 7 miljarder kronor inberäknat för bland annat utbildningsinsatser.

I sammanhanget är upphandlingen som ska belysas i denna rapport ganska blygsam i omfattning: kostnaderna för *Stöd och Matchning* uppgår till 240 miljoner kronor för år 2016. Det motsvarar knappt 3 promille av den årliga omsättningen i hela AF-systemet. I genomsnitt återfinns 20 000 deltagare per månad i tjänsten. Det intressanta är inte storleken på insatsen, utan dess utformning.

Sedan 2014 har AF i LOV-upphandlingen *Stöd & Matchning* en modell som har tydliga anglosaxiska ingredienser. Individer matchas in i fyra separata spår beroende av den arbetslöses behov av insatser. Upphandlade aktörer mäts efter resultat. En banbrytande utveckling är att leverantörernas prestation från och med nu redovisas i ett trestjärnigt *rating-system*. Resultaten går nu att följa ner på regional nivå.

Nu är tiden inne för att göra en första granskning av tillförlitligheten i *rating-systemet*. Dessutom är det viktigt att fortsätta arbeta med utvecklingsområden och justeringar som kan förbättra systemet ytterligare. Rapporten syftar därför till att besvara tre frågor:

Är *rating-systemet* tillförlitligt? Dels ska frågan besvaras utifrån ett system-perspektiv, dels ur aktörernas perspektiv. Som grund för analysen har vi även jämfört AF:s *rating-system* med de modeller som finns i Australien och i Storbritannien.

Vilka utvecklingsområden är betydelsefulla för att ytterligare förstärka resultaten i upphandlingen och *rating-systemets* funktionssätt?

Den sista frågan är mer principiell: kan utformningen av *Stöd & Matchning* och dess resultatsystem vara vägen framåt för flera tjänster och för AF-systemet i stort?

Argumentation

Rapportförfattarna slår fast att AF:s nya rating-system är tillräckligt tillförlitligt. Den metodik som myndigheten använder sig av är tydligt inspirerad från det framgångsrika australiensiska exemplet. Det är en välkommen utveckling att resultatuppföljningar genomförs och redovisas. Det är slående att leverantörerna känner igen resultaten till och med för enskilda kontor.

Det finns emellertid uppenbara förbättringsbehov som på kort sikt behöver åtgärdas: För det första, informationen måste tillgängliggöras och digitaliseras på ett helt annat sätt än idag. Myndighetens försök att sprida informationen har inte varit tillräckligt tydlig. Det påverkar de arbetslösas kunskap om aktörernas prestation, men även företagens behov av *benchmark*-inlärning och kunskap om konkurrenssituationen.

För det andra, för att systemet ska vara rättssäkert, och ytterligare träffsäkert, måste AF:s bedömningsstöd standardiseras och implementeras på ett helt annat sätt än idag. Det finns en risk att deltagare inte får den typ av stöd de är förtjänta av. En missriktad målgrupp kan därutöver ge missvisande resultat i *rating-systemet*.

Ett genomarbetat *profiling*-system måste utvecklas i takt med att resultatsystemet utvidgas och finkalibreras. Det är särskilt angeläget i ett läge där vi ser att allt fler deltagare i insatserna kommer direkt från etableringsuppdraget och från jobb- och utvecklingsgarantins Fas 3.

För det tredje, de resultatvärden som redovisas måste kopplas tydligare till resultat i närtid. En *just in time* princip bör fastslås. Ett kvartals fördröjning i enlighet med det australiensiska exemplet bör vara ledstjärna.

För det fjärde, givet att bedömningsstödet kan säkerställa en tydligare objektivitet i kartläggningsförfarandet bör leverantörskontrakt påverkas tydligare vad gäller volymer, samtidigt som kundmakten får ett tydligare avgörande i marknads-hänseende.

Slutligen, givet att metodiken finns och de facto används bör uppmärksamhet riktas till andra programinsatser där ett rating-system kan implementeras. Även förmedlingens egna verksamheter bör omfattas. Det skulle med all sannolikhet leda till bättre resursoptimering, tydligare uppföljning och en gradvis förbättring av hela AF-systemet. Sammantaget kan det i förlängningen leda till större kostnadseffektivitet och bättre resultat.

Arbetsförmedlingens rating av externa utförare i Stöd och Matchning

Behovet av professionella upphandlingssystem vid AF har varit uppenbart under lång tid. Tidigare upphandlingar kantades av initiala implementeringsproblem. I media uppmärksammades frågan framförallt under 2011 och 2012. I efterdyningarna av jobbcoach-debatten har AF tagit flera steg mot att skapa fungerande KA-tjänster bland annat genom införandet av enhetliga rutiner och ett ratingsystem för externa leverantörer av tjänsten *Stöd och matchning*.

Ett leverantörsråd har bildats. Dessutom har AF:s ledning besökt länder som Australien och Storbritannien för att söka inspiration för förnyelse och förbättring. Allt detta är en välkommen utveckling.

Ett resultatsystem i form av *star-rating* har nu sätts på allvar. Under 2016 publicerades resultaten framförallt för tjänsten inom *spår 1*. I slutet av året utvidgades resultaten även till *spår 2, 3 och 4*. Detta är av vikt, eftersom dessa även inkluderar utrikes födda som står för en allt större del av de inskrivna vid AF.

Svenskt Näringsliv har länge argumenterat för att Arbetsförmedlingen ska använda *rating*-instrument av alla sina tjänster oavsett om de upphandlas eller utförs i egen regi. Vid en RFI (*Request for information*) år 2014 delgavs Svenskt Näringsliv att inkomma med synpunkter för utformningen av *Stöd & Matchning*. I detta remissförfarande framförde vi tydligt att ett *rating-system* är en viktig beståndsdel för en framgångsrik upphandling.¹

Resultatsystemets konstruktion och beräkning

Australien gjorde om sin offentliga arbetsförmedling i grunden i slutet av nittioalet. Ett femgradigt *Rating-system* där de aktörer som presterar väl belönas med mer uppdrag är en central grundbult i systemet. Leverantörer som presterar dåligt sorteras ut. Kundval återfinns i systemet, även om deltagarnas användande av den möjlighet är relativt låg.

Här hemma i Sverige har AF valt en tregradig "Star Rating" som utgår från resultatmättet "andel som är i arbete eller studier 90 dagar efter avslutad insats". Som arbete/studier räknas resultat som har registrerats och godkänts i Arbetsförmedlingens så kallade KA-system.² Arbete kan vara heltidsarbete eller deltidsarbete och med studier avses sådana studier som är godkända enligt förfrågningsunderlaget.

Resultatuppgifterna hämtas från myndighetens KA-system vilket innebär att ratingen baseras på uppgifter som har registrerats och godkänts av AF efter det att leverantörerna har skickat in intyg från varje enskild deltagares arbetsgivare eller det att deltagaren har fullföljt en studietid.

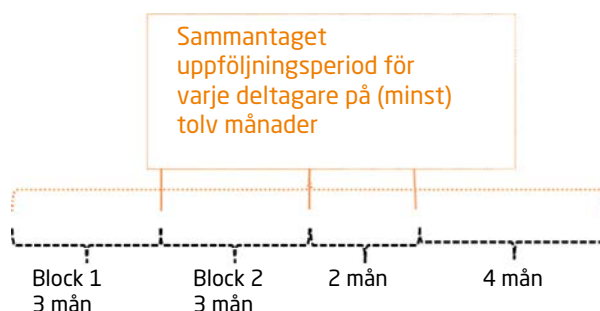
¹ RFI 2014 Svenskt Näringsliv

² KA står för kompletterande aktörer

Den sammantagna uppföljningsperioden för varje deltagare uppgår därför till minst tolv månader vilket beror både på tjänstens längd och på de krav som ska vara uppfyllda för att en ersättning ska kunna utgå. Det registrerade arbetet skall ha varat i minst fyra månader och som godkända utfall räknas även personer som fick jobb två månader efter det att block 2 avslutades, vilket kan ses i bilden nedan. Dessa behöver kunna följas upp i minst fyra månader.

Trots den långa uppföljningsperioden uppger AF att den är något snävt vald eftersom det i praktiken kan ta längre tid att få sin ansökan om utbetalning registrerad och godkänd. Den första publiceringen av ratingen publicerades i juni 2016 med uppgifter som samlats in under perioden mellan december 2014 till maj 2015. En andra publicering av ratingen som omfattar samtliga fyra spår publicerades i december 2016. Resultaten publiceras på en pdf-fil som går att återfinna på AF:s hemsida.³

Uppföljningsperiod för deltagare i STOM, spår 1



Arbetsförmedlingen använder en standardiserad metod för datainsamling vilket gör det möjligt att analysera uppgifterna individ för individ och att också koppla ihop uppgifter om individens förutsättningar att få arbete, med vilken leverantör som har utfört insatsen och resultatuppgifter om individen kommit i arbete eller studier.

Den standardiserade datainsamlingen och kodningen innebär att AF jämför lika för lika i alla steg i ratingsystemet. De ratingvärden som tas fram är därmed jämförbara mellan olika verksamheter trots att de har olika deltagarsammansättning. Detta innebär också att AF har löst frågan om datainsamlingen vilket annars brukar vara en större utmaning i utveckling av tillförlitliga resultatjämförelser snarare än konstruktionen av relevanta utfallsmått.

Ratingen beräknas i fyra steg

En konsekvens av den individbaserade datainsamlingen är att Arbetsförmedlingen kan tillämpa beräkningssätt som används sedan länge inom interventionsforskning och i befolkningsstudier. I detta fall sker detta genom att ratingvärdet beräknas i fyra steg:

1. Deltagarnas statistiska sannolikhet att gå till arbete/studier kartläggs i en "profilering".
2. Gick deltagarna till arbete/studier?
3. Sannolikheten för en person att gå till arbete/studier jämförs med hur det gick i verkligheten.
4. Leverantörer jämförs med varandra.

³ Rating-uppgifter, Stöd och matchning, (dec -16) <https://www.arbetsformedlingen.se/download/18.546b84d6158f5ee077677e60/1483088621936/foretag-stod-och-matchning.pdf>

Forskning om rating-system

Australien var först ut med *star rating*-system. I Australien-exemplet visar Dan Finn, professor och författare bakom studier från bland annat EU-kommissionen, att den enskilda aspekt som har haft störst betydelse för utvecklingen av Job Network var införandet av *Star Rating* av utförare som kom 1999.

Två år efter införandet av *Star Rating* så hade utfallet för arbetssökande i arbete efter 13 veckor förbättrats från 15 procent, till 35 procent.⁴¹

OECD genom författaren David Grubb menar därutöver i en annan rapport att "utvärderingar tyder på att *starratingen* – och dess koppling till urvalet av utförare – har haft en mycket stor betydelse för den ökade effektiviteten, så anges till exempel effektivitetsökningen i arbetsförmedlingsverksamheten till följd av den andra upphandlingen år 2000 till cirka 25 procent".⁵

Även Storbritannien har genomfört en omfattande reformering av till exempel programmen för långtidsarbetslösa, något som uppmärksammas i ny forskning från 2015. Privata och ideella aktörer står för utförandet. Ett fåtal aktörer har kontrakterats av det brittiska arbetsmarknadsdepartementet.

Dessa har i sin tur underleverantörer. Resultatbaserad ersättning och uppföljning har medfört en tydligare styrning i Storbritannien-modellen. Det brittiska exemplet påvisar tydlig kostnadseffektivitet. Resultat: kostnaden för insatserna har halverats på fyra års tid samtidigt som effektiviteten i programmen är ungefär på samma nivå.⁶

Även i Sverige går det att skönja positiva effekter av just marknadsanpassning och kostnadseffektivitet i samband med tydliga resultatsystem. Ett tidigare *rating-system* från 2013 har utvärderats av AF. Trots att det endast fanns till i ungefär fem månader visar utvärderingen följande:

"I rapporten visas att det finns tydliga indikationer på att publiceringen av aktörernas tidigare prestationer i form av ratingvärden har betydelse för hur arbetssökande väljer. Publiceringen av ratingvärden har stärkt kopplingen mellan utförarnas "kvalitet" (mätt i förmågan att förmedla deltagare till jobb) och antalet nya deltagare som dessa har kunnat knyta till sig".

"Utifrån ett mer teoretiskt perspektiv innebär resultaten att ett väsentligt villkor för potentiella samhällsekonomiska effektivitetsvinster är uppfyllt, nämligen att de leverantörer som presterar bättre ökar sina marknadsandelar. Det skapar incitament för alla leverantörer att sträva efter att förbättra kvaliteten på sina tjänster så att den genomsnittliga kvaliteten ökar över tid".⁷

⁴ EU, Finn, Subcontracting employment services, 2011 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/214527/rrep752.pdf

⁵ Citerad i Jansson & Kilsved (2009) s. 24, i sin tur tagen från Grubb (2006) s. 5.

⁶ Läs mer i "Storbritanniens jobbmirakel", Hamilton, med färskas källor (2015).

⁷ Gerdes (2015) AF working paper.

Profilering av deltagarna

Alla arbetssökande som deltar i insatsen *Stöd och Matchning, spår 1* har bedömts stå förhållandevis nära arbetsmarknaden och behöva stöd för att få en anställning. Trots detta finns betydande variationer inom gruppen som deltar i insatsen. Även för de individer som ingår i *spår 2, 3 och 4* varierar förutsättningarna.

Här återfinns numera personer som tidigare var i jobb- och utvecklingsgarantins Fas 3 och även nyanlända flyktingar i etableringsreformen. Därmed blir en kartläggning av individernas specifika behov och träffsäkerhet i förhållande till de ersättningar som erbjuds i de separata spåren mycket viktig.

AF försöker hantera detta genom att resultatmålet tar hänsyn till både deltagarnas statistiska sannolikhet för att få en anställning. Myndigheten tar även i beaktande regionala skillnader. Både AF och Australiens arbetsmarknadsdepartement använder begreppet ”profilering” för att beskriva detta.

Profileringen tar i det här fallet hänsyn till ålder, kön, utbildning, arbetslivserfarenhet, yrkesfärdighet, födelseland, arbetslösheten i kommunen, hemmahörande i storstadsområde, tidigare arbetslöshetsperioder och när personen skrevs in på Arbetsförmedlingen. Särskilt hänsyn tas till att deltagarna befinner sig i olika delar av landet och att detta innebär olika förutsättningar att hitta arbete/gå till studier.

AF gör vissa anpassningar av de variabler som hämtas från beslutsstödet. Bland annat används separata modeller för män och kvinnor för att möjliggöra en ökad flexibilitet i eventuella framtida analyser av deltagande och resultat. Ett exempel är att fler män jobbar inom den privata sektorn vilket gör att både arbetsmarknadens utveckling och enskilda variabler som utbildning och tidigare arbetslöshetshistorik kan påverka män och kvinnor på olika sätt.

Därefter validerar Arbetsförmedlingen den viktning som görs för profileringen med en logistisk regressionsmodell med variabler från Arbetsförmedlingens beslutsstöd och den nationella sannolikheten för att bli långtidsarbetslös i sex månader. Regressionsmodellen används för att ta fram ett förväntat värde för varje deltagares förutsättning för att gå vidare till arbete eller studier.

Beräkning av leverantörernas ratingvärde

Beräkningar görs enbart för de leverantörer som på respektive kontor haft fler än 12 deltagare under uppföljningsperioden. Uppgifterna om deltagarna var i arbete eller studier 90 dagar efter avslutad insats rapporteras in av leverantörerna själva och innebär att de hämtar ett intyg från deltagarnas arbetsgivare som skickas till AF och registreras i det KA-systemet.

Rent praktiskt innebär kodningen att resultatuppgifterna kopplas till varje arbetssökande som har en viss leverantör under en period. Varje enskild deltagares sannolikhet att gå till arbete eller studier jämförs därefter med det faktiska resultatet för varje enskild deltagare. I sammanfattning beräknas detta som:

- En leverantör har fyra deltagare, som utifrån beräkningsmodellen bedöms ha sannolikhet att få arbete/studier lika med 0.1, 0.2, 0.2 respektive 0.3.
- När dessa fyra värden adderas får vi ett aggregerat *förväntat* värde på 0,8 ($=0.1+0.2+0.2+0.3$).

- Det *verkliga* resultatet blev att en av dessa fyra personer fick arbete/studier så att det samlade värdet på det faktiska utfallet blir lika med 1 ($=1+0+0+0$).
- Skillnaden mellan det verkliga och förväntade värdet blir 0.2 ($=1-0.8$).
- Därefter divideras differensen med antal deltagare, vilket ger resultatet 0.05 ($=0.2/4$), det vill säga ett slags medelvärde över hur det gick för deltagarna hos denna leverantör givet förutsättningarna att kunna lyckas. Utförarna påverka alltså sin egen rating genom att prestera över det förväntade värdet.

Efter att skillnaden mellan leverantörens förväntade värde och faktiska utfall har fastställts jämförs leverantörerna i varje region med varandra och indelas i percentiler och att den lägsta gruppen får en stjärna, medelgruppen får två stjärnor och den högsta gruppen får tre stjärnor. I första releasen innebär detta att ungefär 60 procent av de leverantörer som blev ratade tilldelades två stjärnor vilket i stort följer normalkurvan.

Sammanfattning: Bedömning av rating-systemets tillförlitlighet

Genomgången visar att AF har konstruerat den typ av *ratingssystem* som Svenskt Näringsliv efterfrågar. *Ratingen* uppfyller också de krav på tillförlitlighet och relevans som kan ställas på en första publicering. AF har använt vedertagna vetenskapliga metoder och ratingen har presenterats på vetenskapliga konferenser och för OECD.⁸

De uppgifterna som används för profileringen är samordnade med AF:s beslutsstöd. Även om IFAU har beskrivit att urvalet av AF:s beslutsstöd och metoder sällan skett på vetenskaplig grund innebär regressionsmodellen i sig en kvalitetssäkring av den viktning som görs.

Detta är en anledning till att statistiskt framtagna bedömningsinstrument anses mer tillförlitliga än kliniska om det finns ett stabilt utfallsmått som de kan valideras emot. Den konstruktion för datainsamling och validering av ratingen som AF har valt innebär också att de variabler som ligger till grund för ratingen kan kvalitetssäkras efter hand med hjälp av de statistiska beräkningarna.

De beräkningssätt som AF har valt innebär att de variabler som ligger till grund för ratingen kan kvalitetssäkras (valideras) med hjälp av statistiska beräkningar vilket gör dem tillförlitliga. Den standardiserade datainsamlingen innebär också att det blir möjligt att analysera hur matchningsprocesserna fungerar trots att resultaten påverkas av flera olika aktörer. Därmed blir det också möjligt att förbättra arbetsmarknadspolitiken tröfsäkerhet och utveckla en förbättrad processtyrning av det slag som Svenskt Näringsliv länge har argumenterat för.

⁸ Arbetsförmedlingen 2014, Gerdes, C, 2015.

Internationella jämförelser

Jämförelse med den australiensiska ratingens konstruktion

I arbetet med rapporten har vi studerat och jämfört metodiken och modellerna som återfinns i länder som Australien och Storbritannien. Vi har försökt göra en komparativ analys mellan det svenska resultat-systemet och de anglosaxiska modellerna.

I likhet med AF använder Australiens arbetsmarknadsdepartement en *star rating* med en profilering av de individuella deltagarna som tar hänsyn till deras statistiska förutsättningar att få en anställning.

Australien använder också samma beräkningssätt med en regressionsmodell för att fastställa den viktning som ligger till grund för profileringen och för att ta hänsyn till regionala skillnader på arbetsmarknaden.⁹ OECD har beskrivit den metod som Australien använder för att fastställa ratingvärdet på som robustare och mer tillförlitligt än den mål- och resultatstyrning som används i andra OECD länder.¹⁰

Ytterligare en likhet mellan Australien och Sverige är att båda använder ett beslutsstöd¹¹ för att inplacera de arbetssökande i fyra olika spår beroende på hur långt från arbetsmarknaden de har bedömts stå. Till skillnad från i Sverige är det beslutsstöd som Australien använder evidensbaserat men det har också använts och vidareutvecklats sedan 1999 medan Sverige började implementera beslutsstödet först år 2012 vilket innebär att det ännu inte har kunnat utvärderas.¹²

För beräkningen av ratingvärdet finns en skillnad mellan Australien och Sverige som handlar om att de utfallsmått som används i Australien är något annorlunda uppbyggda jämfört med i Sverige där AF endast tar hänsyn till om den arbetssökande är i arbete eller studier 90 dagar efter avslutad insats. Den australiensiska ratingen tar också hänsyn till variabler om hur lång tid det tar tills den arbetssökande kommer i arbete eller studier samt i hur hög grad den arbetssökande har deltagit aktivt i de insatser som beviljats.

I Australien presenteras ratingen kvartalsvis, dels i form av en pdf-fil på arbetsmarknadsdepartementets hemsida och digitalt på portalen www.jobactive.gov.au som samlar tjänster för både arbetssökande och arbetsgivare. De jobbar också aktivt med att informera deltagarna om leverantörernas resultat inför val, till exempel genom obligatoriska informationsträffar för arbetssökande.

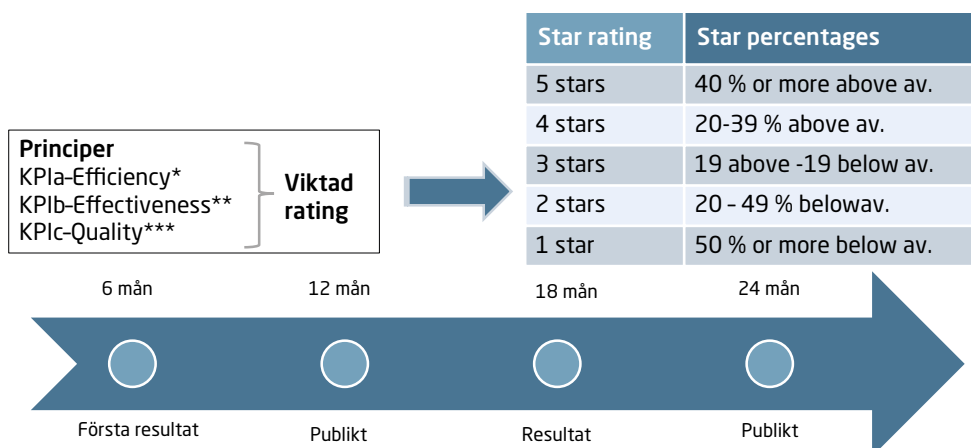
Australiens star rating-system är viktad enligt tre KPI:er, se nedan. Poängsättningen är därefter strukturerad utifrån genomsnittsresultat för aktörerna. Resultaten redovisas på regional nivå.

⁹ Metoden finns beskriven på webbsidan <https://employment.gov.au/jobactive>

¹⁰ OECD, 2012.

¹¹ I beslutsstöd används enhetliga begrepp och skalor för att säkerställa att alla handläggare gör samma bedömning.

¹² IFAU 2011.



Jämförelse med den brittiska ratingens konstruktion

Även Storbritannien använder en *star rating* – metodik. Den har dock en annan uppbyggnad och konstruktion än den som används i Sverige och i Australien. En avgörande skillnad handlar om hur datainsamlingen görs. I UK samlas data och utfallsmått in på verksamhetsnivå, som medelvärde eller med poängsättning. Ett flertal olika variabler ligger till grund för ratingvärdet och beräkningen sker på olika sätt för olika variabler.

Utfallsmåtten om andel i arbete beräknas till exempel genom att verksamheternas medelvärden rangordnas i en relativ poängsättning. För variabler om insatsernas kvalitet används ett formulär för självutvärdering som fylls i av utföraren och som resulterar i en i förväg definierad poängsättning. Även variabler om avtalsefterlevnad och finansiell stabilitet ligger till grund för ratingen. De variablerna följs däremot upp, och poängsätts, av personal från den upphandlande myndigheten.¹³

Sättet att samla in data på innebär att ratingen inte kan ta hänsyn till regionala skillnader och till att olika företag kan ha deltagare med olika förutsättningar, det vill säga det som Sverige och Australien använder profileringen till. Detta innebär också att ratingens tillförlitlighet inte kan testas statistiskt eftersom olika verksamheter kan ha deltagare med olika förutsättningar.

Resultatuppgifterna är därför inte jämförbara mellan olika verksamheter, men resultatvariablerna ges heller inte samma genomslag för ratingen som i Australien och i Sverige. UK har istället lagt större tyngd på de variabler som mäter avtalsefterlevnad och följsamhet till nationell policy. Tanken är att variablerna ska korrelera med resultaten.

Ratingen presenteras kvartalsvis som en pdf-fil på *Department of Work and Pensions* hemsida. Ratingen är svår att hitta men de brittiska deltagarna har heller inte valfrihet på samma sätt som i Sverige och Australien.

¹³ Metoden finns beskriven på webbsidan <http://webarchive.nationalarchives.gov.uk/20100515192520/dwp.gov.uk/supplying-dwp/what-we-buy/welfare-to-work-services/star-rating-system/>

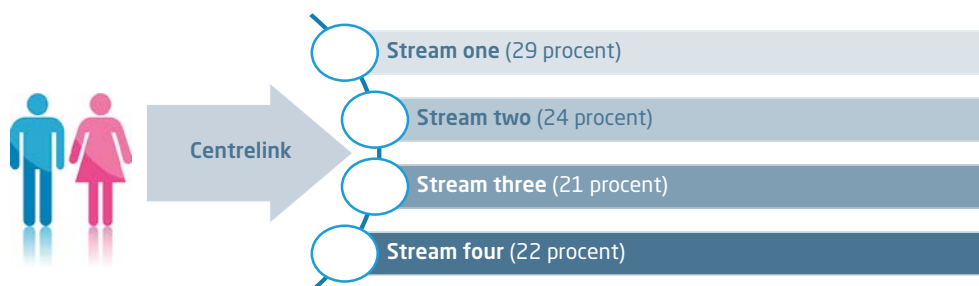
Hur Australien använder ratingen för att påverka det nationella resultatet

Den typ av rating som AF har valt att använda har stora likheter med den modell som används i Australien och som Svenskt Näringsliv har lyft fram som ett gott exempel som Sverige bör lära av. Den skillnad som finns mellan länderna handlar främst om hur ratingen används för att påverka de nationella resultaten.

Australien drog tidigt slutsatsen att resultatbaserade ersättningar och kundval i sig inte var tillräckliga åtgärder för att leda till en höjning av de nationella resultaten. De valde istället en strategi som går ut på att styra volymerna av deltagare till leverantörer med bra resultat genom att avsluta kontrakten med de leverantörer som har sämst resultat. Därigenom har de också skapat en hävstångseffekt i systemet som ger leverantörer med höga resultat incitament att förbättra sig ytterligare.¹⁴

Incitamenten uppstår eftersom medianvärdet och det förväntade värde som leverantörerna behöver överträffa för att få högsta rating höjs när volymerna ökar hos leverantörer med höga resultat.

Australien började också tidigt att utveckla en utvärderingsstrategi som gör det möjligt att sprida kunskap om framgångsrika arbetssätt till fler utförare. De har både utnyttjat de möjligheter som profileringsinstrumentet ger genom att analysera om olika arbetssätt ger olika resultat för olika målgrupper och för att tillvarata kunskap från interventionsforskning om effektiva arbetssätt.



Detta sker till exempel genom att de australiensiska leverantörerna ges lättillgänglig information om hur de själva presterar i jämförelse med andra leverantörer, vad det egna förväntade värdet är och information om vad som är "best practice" i tillgänglig forskning.¹⁵

De genomsnittliga resultaten i de flesta aktiva arbetsmarknadspolitiska program är låga, men ett sätt att höja resultaten är att oftare ge rätt insats till rätt person.¹⁶ Australien använder profileringen till detta i kombination med en dialog med leverantörerna om hur insatserna kan anpassas efter deltagarnas behov.

Detta har också lett till att många mindre leverantörer har specialiserat sig på särskilda målgrupper. De australiensiska myndigheterna har på så sätt kunnat tillföra kunskap som leverantörerna inte själva hade kunnat upptäcka enbart genom att följa upp sina egna processer. Därmed har de också åstadkommit ett lärande som omfattar hela systemet och som inte enbart sker inom respektive organisation.¹⁷

¹⁴ OECD, 2012.

¹⁵ Finn, D., 2007.

¹⁶ IFAU, 2016.

¹⁷ OECD 2013, Finn, D., 2011.

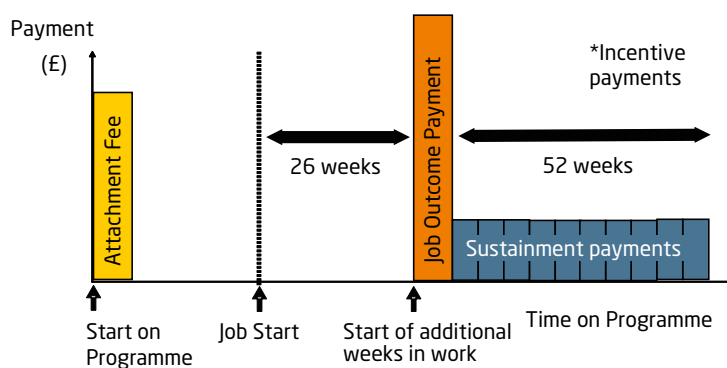
Hur använder Storbritannien ratingen för att påverka de nationella resultaten?

Motiven till Storbritanniens vägval finns dokumenterade i en underlagsrapport till parlamentet. Där angavs framförallt att de ville hitta en modell som kunde motverka de implementeringsproblem hade funnits i tidigare program.

UK valde därför istället att inkludera frågor i ratingen om huruvida leverantörerna själva har rutiner för en systematisk kvalitetsuppföljning inom organisationen och att varje deltagare får insatser anpassade efter de egna förutsättningarna. Den australiensiska datainsamlingen beskrevs också som alltför komplex för att kunna införas så snabbt som politikerna efterfrågade.¹⁸

De förändringar som Storbritannien valde att genomföra med Australien som förebild handlade istället om att upphandla matchningstjänster av privata leverantörer och att använda resultatbaserade ersättningar för att ge leverantörerna incitament att göra sitt bästa för alla få alla deltagare i arbete.

The Payment by Results Model



Trots att Storbritannien har valt att använda en annan typ av ratingsystem har deras strategi för att förbättra de nationella resultaten likheter med hur AF arbetar. Båda har eftersträvat att undvika detaljstyrning men är däremot noggranna med att följa upp avtalens skall-krav som ett sätt att säkerställa att de får en tjänst som motsvarar förväntad kvalitet. I båda länderna baseras också eventuella beslut om att avsluta avtal på om leverantörerna lever upp till avtalens skall-krav eller inte.

Varken AF eller UK använder alltså ratingresultat för att avsluta avtal med leverantörer. Detta är också det vanliga i andra OECD länder eftersom resultaten inte är jämförbara mellan olika verksamheter utan profilering av deltagarna.¹⁹

Eftersom skall-krav och andra benchmarkingmått sällan korrelerar med resultaten innebär detta också att de flesta länder, inklusive Storbritannien och Sverige har svårt att påverka sina nationella resultat enbart genom att avsluta avtal. Svårigheten i att lösa detta hänger ihop med vad OECD brukar beskriva som heterogenitet och variation i sina granskningar.²⁰

¹⁸ British Parliament, 2007.

¹⁹ OECD 2013, OECD 2014.

²⁰ OECD 2013.

Heterogenitet och variation i verksamheterna är en anledning till att många länder använder en så kallad "black-box" i kombination med resultatstyrning. Tanken är att myndigheter och politiker ska undvika att styra med regler, resurstilldelning och skall-krav just eftersom de inte korrelerar med resultaten varken på verksamhetsnivå eller för de deltagare som avviker.

Att enbart formulera resultatmål som baseras på föregående års nationella resultat är dock visat sig vara otillräckligt eftersom det bara ger incitament till de verksamheter som presterar under genomsnittet och samtidigt har en ambition om att konkurrera med bra kvalitet. Det är också vanligt att verksamheter jämför sig med det nationella genomsnittet trots att resultatet dras ned av de aktörer som underpresterar.

I samband med de reformer som genomfördes år 2011 valde Storbritannien att hantera dilemmat med heterogenitet och variation genom att fortsättningsvis använda enbart resultatbaserad ersättning. Tanken var dels att ge leverantörerna incitament att prioritera samtliga deltagare och att samtidigt göra marknaden oattraktiv för oseriösa leverantörer.²¹

Samtidigt införde Storbritannien också ett standardiserat bedömningsinstrument kallat *Work Capability Assessment*. Instrumentet som används för att bedöma om individen ska delta i en matchningstjänst eller är berättigad till ersättning via något annat socialförsäkringssystem.²²

Det används däremot inte för profilering av deltagarna trots att reformerna syftade till att förbättra kostnadseffektiviteten och lättare kunna utvärdera de upphandlade tjänsterna. Anledningen till detta finns inte motiverad. I intervjuerna med AF framkommer däremot att de helt enkelt inte hade insett att de nationella resultaten också påverkas av hur volymerna av deltagare fördelas mellan olika verksamheter och att verksamheter med hög rating får nya incitament om de avslutar avtalen med leverantörer med lågt resultat.

Sammanfattning länder-exempel

Vid en komparativ jämförelse står det klart att det finns tydliga likheter, men också avgörande skillnader. Ett rättssäkert och fungerande system förutsätter att även bedömningsstödet är väl utformat och anpassat för att bedöma den arbetslöses behov av stöd. Det är även viktigt att det resultat som beräknas utgår från rätt målgrupp.

Dessutom står det klart att en tydlig styrning av volymer till aktörer som presterar bra måste komplettera de effekter som kommer av kundval. I övrigt bör noteras tre viktiga intryck:

a. Lärdom från Australien: Styr volymerna av deltagare till leverantörer med höga ratingvärden

Den typ av resultatstyrning som Australien utvecklat efter införandet av ratingen och som AF bör lära sig av skiljer sig från den resultatstyrning som används i de flesta andra OECD-länder. De flesta OECD-länder använder ratingresultat som baseras på genomsnittliga verksamhetsdata.

²¹ Cumming, L. M., 2011, Department for Work and Pensions, 2011.

²² Svenskt Näringsliv, 2016.

De kan därför inte använda ratingen för att avsluta avtal eller för att analysera resultatvariation, varken mellan olika verksamheter eller mellan olika grupper av deltagare. De flesta OECD-länder inklusive Sverige och England använder istället skall-krav som grund för att avsluta avtal trots att dessa sällan korrelerar med resultaten. De flesta formulerar också sina nationella resultatmål baserat på föregående års genomsnittliga resultat trots att det nationella resultatmålet då bara kan ge incitament till de verksamheter som presterar under det nationella genomsnittet.²³

I Australien drogs däremot tidigt slutsatsen att resultatbaserade ersättningar och kundval i sig inte var tillräckliga åtgärder för att leda till en höjning av de nationella resultaten. De valde istället en strategi som går ut på att styra volymerna av deltagare till leverantörer med bra resultat genom att helt enkelt avsluta avtalen med de leverantörer som har sämst resultat.²⁴

Förutom att Australien höjer de nationella resultaten ”matematiskt” genom att utesluta aktörer som underpresterar har de därigenom också kunna skapa en hävstångseffekt i hela systemet som även ger leverantörer med höga resultat incitament att förbättra sig ytterligare.

Hävstångseffekten uppstår eftersom medianvärdet och det förväntade värde som leverantörerna behöver överträffa för att få högsta rating höjs när volymerna minskar hos leverantörer med låga resultat. Det som ofta förbises vid jämförelser av olika verksamheters resultat är att antalet deltagare kan vara olika i olika verksamheter och att detta påverkar de nationella resultaten.

Tyvärr gäller också motsatt samband. Ett exempel är att om många deltagare väljer ett fåtal företag som inte eftersträvar resultatersättningen så kommer både det förväntade värdet och det nationella resultatet att sjunka. Detta kan vara svårt att upptäcka eftersom det kan ske utan att antalet verksamheter med dåligt resultat blir fler.

Eftersom detta också innebär att gränsen för att få högsta rating sänks kan de andra leverantörerna få intryck av att de fortfarande presterar bra trots att den egna verksamheten egentligen försämras. Som en del i resultatstyrningen behöver Arbetsförmedlingen därför också analysera hur volymerna fördelar sig mellan olika verksamheter eftersom det påverkar de nationella resultaten.

b. Lärdom från Storbritannien: Deltagarnas val är inte tillräckligt för att skapa kvalitetskonkurrens

Trots att Storbritannien har valt att använda en annan typ av ratingsystem har AF:s strategi för att förbättra de nationella resultaten större likheter med Storbritannien än med Australien. Båda har eftersträvat att undvika detaljstyrning men är däremot noggranna med att följa upp avtalens skall-krav som ett sätt att säkerställa att de får en tjänst som motsvarar förväntad kvalitet. I båda länderna baseras också eventuella beslut om att avsluta avtal på om leverantörerna lever upp till avtalens skall-krav eller inte.

²³ OECD 2013.

²⁴ OECD 2012.

OECD har dock beskrivit att skall-krav och andra benchmarkingmått sällan korrelerar med resultaten och att detta är en anledning till att de flesta länder har svårt att påverka sina nationella resultat enbart genom att avsluta avtal.²⁵ Eftersom AF:s rating är både tillförlitlig och jämförbar bör de därför övergå till att istället använda *rating-resultaten* för att avsluta avtal.

För att detta skall kunna ske på ett rättssäkert sätt krävs dock AF först säkerställer att beslutsstödet tillämpas korrekt samt att de förbättrar transparensen om deltagarnas profilering, till exempel genom att kommunicera verksamheternas aggregerade förväntade värde. AF:s upphandlingar medger däremot redan att de använder ratingen för att utesluta leverantörer.

Rating är emellertid endast ett av flera styrinstrument som staten kan använda för att utveckla kvaliteten i upphandlad verksamhet. Några av de andra förändringar som Storbritannien genomförde när de inte fick tillräcklig effekt i sina program har också haft goda resultat som AF bör lära sig av.

Storbritannien har till exempel valt att enbart använda resultatbaserad ersättning som ett sätt ge leverantörerna incitament att prioritera samtliga deltagare och att samtidigt göra marknaden oattraktiv för oseriösa leverantörer. Storbritannien införde också ett standardiserat bedömningsinstrument som används för att bedöma om individen ska delta i en matchningstjänst eller är berättigad till ersättning via något annat socialförsäkringssystem.

Deltagande i matchningstjänsten har därför en tydligare karaktär av myndighetsutövning än i Sverige och Australien och är istället ett krav för att få arbetslöshetsersättning. Dessutom kan de arbetssökande endast byta leverantör om det finns särskilda skäl vilket ger leverantörerna bättre möjligheter att ställa krav på deltagarna.

För de arbetssökande som är motiverade och vill kunna välja den leverantör som har störst sannolikhet att kunna hjälpa dem vidare till arbete eller studier ger AF:s rating ett förbättrat underlag. Det går dock inte att bortse ifrån att många arbetssökande saknar egen motivation samtidigt som arbetsgivarna har ett stort rekryteringsbehov.

c. Koppla ihop uppgifter om förutsättningar, insatser och resultat till individen

Tanken med den resultatstyrning som används i de flesta OECD-länder är att myndigheter och politiker ska undvika att toppstyra med regler, resurstilldelning och skall-krav eftersom dessa sällan korrelerar med de genomsnittliga resultaten. Den resultatstyrning som används i de flesta OECD-länder har istället utvecklats som ett sätt att hantera den heterogeniteten som finns inbyggd i välfärds-verksamheter.

När resultatstyrningen började spridas under 1990-talet hade offentlig sektor fortfarande inte digitaliserats vilket gjorde att man då saknade metoder för att kunna samla in resultatuppgifter på individnivå vilket är en förutsättning för att kunna analysera heterogenitet och variation i ett system.

²⁵ OECD 2013.

De förvaltningspolitiska styridéerna om den ”svarta lådan” har däremot aldrig varit ett ideal. De var snarare sätt att hantera att staten inte kunde följa upp vilka insatser som gav resultat eftersom detta varierar beroende på individens behov. När staten använder resultatanalyser som baseras på genomsnittliga verksamhetsresultat döljs nämligen kvalitetsbrister som drabbar enskilda arbetssökande eller grupper av arbets-sökande.

Australiens sätt att använda datainsamling och profileringsinstrumentet innebär däremot att de kopplar uppgifter om förutsättningar, insatser och resultat till individen vilket gör det möjligt att analysera om rätt insats ges till rätt individ. Den osäkerhet som finns i mängddata som omfattar en heterogen population kan därmed reduceras. Anledningen är att det går att analysera hur resurserna används och att rätt insats ges till rätt individ.

En avgörande skillnad mellan den resultatstyrning som Australien har utvecklat jämfört med traditionell styrning av offentligfinansierad verksamhet är därför att kvalitetsförbättring och effektivisering i det sammanhanget också handlar om att kunna rikta om till exempel politiska beslut, upphandlingskrav och resurstilldelning till nya åtgärder som bedöms ha större påverkan på de nationella resultatmålen.

Detta innebär också att kvalitetsförbättringar kopplas till förbättrad träffsäkerhet och blir ett ansvar för alla aktörer i systemet, inte bara för utföraren för sig eller staten för sig. Staten gör alltså inte enbart kontroller i slutet av en så kallad ”svart låda” när det ofta är för sent. Arbetet med att förbättra resultaten är istället integrerat i hela processen och i hela styrsystemet.

Brittiska tjänstemän bedömde till exempel att de inte skulle hinna införa en individbaserad datainsamling av australiensisk modell innan ”the work programme” implementerades.²⁶

Svårigheterna märks också i IFAU:s utvärdering av AF:s införande av beslutsstödet där det framkommer att handläggarna inte har förstått att uppgifterna också används i andra syften än att styra deras beslut.²⁷ För att säkerställa att uppgifterna registreras korrekt krävs därför ett större implementeringsarbete. Storbritanniens sätt att out-sourca bedömningarna kan vara en förebild för implementeringen.

²⁶ British Parliament, 2007.

²⁷ IFAU 2011.

Vad säger leverantörerna i upphandlingen Stöd & Matchning?

a. Röster från verksamheterna

Det är slående att leverantörerna känner igen sig i sina egna resultat. De är generellt positiva till ratingen men förbättringsområden finns. I intervjuerna har det bland annat framkommit att beräkningarna och profileringen är svåra att förstå och att tiden fram till publicering är alltför lång för att ratingen ska vara rättvisande, men också möjliga felkällor i datainsamlingen.

Leverantörerna framför att det är bra att AF har valt ett resultatmått som är relevant och återspeglar syftet med tjänsten och att de använder beräkningar som gör ratingen jämförbar mellan olika verksamheter. Alla uppger att de känner igen sig i sina egna resultat oavsett om det är högt, medelhögt eller lågt.

De som har olika resultat i olika regioner anser själva att kontoren med sämre rating också är sämre och ingen har menat att konkurrenter har fått högre rating än vad de förtjänar. De har få synpunkter på resultatmättet. De flesta framför istället att det viktiga är att det hålls jämförbart över tid och att AF använder ett robust effektmått som inte bygger på värderingar av vad som är kvalitet.

Till synpunkterna hör att det kan finnas ytterligare frågeställningar som också avspeglar om leverantörerna har gjort ett bra jobb. Det kan till exempel handla om att mäta deltagarnöjdhet och arbetsgivarnöjdhet för att följa upp att de matchningar som görs också är långsiktigt hållbara så att deltagarna inte kommer tillbaka i arbetslöshet. Det har också framkommit att resultaten ibland slås ihop för flera kontor och presenteras som ett genomsnitt.

De intervjuade leverantörerna har också varit mycket tydliga med att ratingen inte är en tillräcklig åtgärd för att skapa kvalitetskonkurrens eller för att de ska kunna hjälpa fler deltagare att komma i arbete. De framför att de resultatbaserade ersättningarna har större betydelse för kvalitetskonkurrensen. För att kunna hjälpa fler deltagare i arbete efterfrågar de både motivationshöjande åtgärder och möjligheter att bättre kunna tillgodose arbetsgivarnas efterfrågan på arbetskraft, till exempel genom att få tillgång till alla arbetssökande direkt vid inskrivningen och inte som nu när de redan förlorat motivationen.

b. Transparensen behöver förbättras

De intervjuade har svårt att bedöma om AF:s beräkningar är korrekta. Även om de uppger att de litar på att Arbetsförmedlingen faktiskt har gjort rätt kan också framföras att de till exempel tycker att det blir konstigt att de kan få samma rating för två olika kontor trots att resultaten är olika.

Det framkommer också missförstånd i intervjuerna till exempel om att konkurrenter som har profilerat sig mot deltagare som befinner sig närmre arbetsmarknaden borde ha fått högre ratingen än det egna företaget som är inriktat på deltagare som har stått utanför arbetsmarknaden länge. Detta riskerar att dra ned leverantörernas tilltro till resultaten men också skapa oönskade beteenden.

Leverantörerna har också ifrågasatt hur inplaceringen av deltagarna görs vilket drar ned deras tilltro till ratingens tillförlitlighet. De har till exempel beskrivit att de har blivit ombedda att avlasta AF genom att själva genomföra intervjuerna med de arbetssökande för att bedöma inplaceringen. Andra har framfört misstankar om att utrikesfödda har fått välja att bli inplacerade i spår 1 för att på så vis kunna få en leverantör som talar deras eget språk.

Leverantörernas antaganden har stöd i en utvärdering som IFAU gjorde år 2015. Där beskrivs att handläggarna inte alltid tillämpar beslutsstödet som avsett och att det förekommer att de fyller i uppgifterna utan att ställa alla frågor eller att de fyller i uppgifterna så att inplaceringen samstämmer med deras egen bedömning.²⁸

Den typen av implementeringsproblem är vanliga i välfärdshandläggning men innebär att det inte är säkert att alla inmatade uppgifter är korrekta. Till de förbättringsmöjligheter som har framkommit när det gäller datainsamlingen hör därför implementeringen av beslutsstödet och förbättrad transparens om hur inplaceringarna och profileringen går till. Implementeringen av beslutsstödet är alltså ett fortsatt förbättringsområde för Arbetsförmedlingen liksom kvalitetssäkring av viktningen av utrikesfödda.

c. Insamlingstiden är alltför lång för att spegla aktuell kvalitet

Samtliga leverantörer har tagit upp att insamlingstiden är så lång att ratingen inte längre återspeglar aktuell kvalitet när den väl publiceras. AF förklarar den långa insamlingsperioden med flera saker. Dels handlar det om att resultatmättet innebär att ratingen kan beräknas tidigast 90 dagar efter att insatsen har avslutats.

Dels handlar det om att insamlingstiden har anpassats för att tillräckligt många deltagare ska ha hunnit avsluta insatsen för att beräkningen av ratingen ska vara tillräckligt tillförlitlig på regional nivå. I kommande releaser kommer insamlingsperioden att förlängas genom att slutdatum flyttas fram vilket innebär att fler leverantörer kommer att kunna få ett ratingvärde.

Alternativet hade varit att inkludera ratingvärden som inte är statistiskt säkerställda för att antingen kunna publicera snabbare eller för att fler leverantörer skulle kunna få ett ratingvärde.

En förutsättning för att få en rating är också att leverantören har haft minst tolv deltagare i spår 1 under uppföljningsperioden. Även fortsättningsvis kommer uppföljningstiden och behovet av ett minsta antal observationer att innebära begränsningar. AF har själva konstaterat att handläggarna i vissa regioner är snabbare att registrera uppgifterna än i andra. De har diskuterat att detta kan innebära att utförare kan drabbas negativt om det innebär att de inte kommer upp i tillräckligt många observationer för att få en rating.

De intervjuade har alla fått ett ratingvärde vilket innebär att frågan om konsekvenser för de som inte fått ett ratingvärde inte har berörts. Arbetsförmedlingen har däremot själva diskuterat om utförare kan drabbas negativt om de inte kommer upp i tillräckligt många observationer för att få en rating.

Vid samrådsmöten har leverantörer också framfört att de föredrar att få ett ratingvärde som inte är statistiskt säkerställt framför att som nu bli utan ratingvärde.

²⁸ IFAU 2015.

d. Ratingen motverkar inte kvalitetskonkurrens och är inte en tillräcklig åtgärd

År 2012 sjösattes en upphandling av externa aktörer i ett pilotprojekt med ett *rating-system*. Utvärderingen visade på att de företag som var bäst på att få arbetssökande i arbete också fick fler sökande.²⁹ Även om det fortfarande är för tidigt att bedöma om deltagarna inom Stöd och matchning kommer att agera på samma sätt är det ändå sannolikt att ratingen uppfattas som relevant av både arbetssökande och leverantörer.

Ratingen borde alltså kunna leda en bättre fungerande kvalitetskonkurrens, det vill säga att leverantörer med bra rating får fler deltagare och att leverantörer som får dålig rating tvingas lägga sin verksamhet.

I intervjuerna har dock leverantörerna haft många synpunkter på ratingens förutsättningar att leda till förbättrad kvalitetskonkurrens. De menar att ratingen inte motverkar kvalitetskonkurrens men att den är långt ifrån tillräcklig. För att kunna leda till kvalitetskonkurrens måste framförallt underlaget för deltagarnas val förbättras, målgruppsanpassas och tillgängliggöras.

Flera framför också att AF även bör utveckla sin muntliga information om ratingen eftersom de är tveksamma till om deltagarna själva engagerar sig i sitt val. Som det är i dagsläget upplever de inte att deltagarna väljer utifrån ratingen men de tror heller inte att de har förutsättningar att göra det. De framhåller också att andelen deltagare som väljer på andra grunder än sannolikheten att gå vidare till arbete är alltför stor för att leverantörer med dålig rating ska tvingas lägga ned sin verksamhet. De får deltagare trots att de har dålig rating.

För att underlätta kvalitetskonkurrens efterfrågar leverantörerna förändringar i upphandlingen av tjänsten främst avseende deras möjligheter att kunna informera arbetssökande om vad de kan erbjuda. Med dagens regler kan leverantörerna till exempel inte informera om vilka målgrupper de är bäst på att hjälpa eller om de har aktuell kontakt med arbetsgivare som behöver rekrytera en viss typ av kompetens.

De måste dessutom ha samma beskrivning för alla kontor vilket innebär att de inte kan informera om vilka språk de kan eller om personalens kompetens. Flera av de intervjuade vill istället kunna marknadsföra vad det egna företaget är bra på för att på så vis också få deltagare som man är bra på att hjälpa till arbete.

De intervjuade efterfrågar förändringar av upphandlingarnas skall-krav även om inte alla är överens om hur. Flertalet framför att skall-kraven sällan korrelerar med ratingsresultaten något som också bekräftas av Arbetsförmedlingen. Många lyfter ändå att Arbetsförmedlingen behöver se till att konkurrenterna lever upp till skall-kraven så alla konkurrerar på lika villkor.

De vill också att AF vara snabbare att avsluta avtalen med leverantörer som inte lever upp till skall-kraven för att på så vis förhindra att leverantörer med dålig kvalitet fortsätter att få deltagare. Ett företag föredrar istället att AF helt avskaffar skall-kraven för att istället använda ratingen för att bedöma om avtal skall avslutas. Ingen av övriga intervjuade har däremot föreslagit att Arbetsförmedlingen skall använda ratingsresultaten för att avsluta avtal.

²⁹ AF 2015.

e. Den resultatbaserade ersättningen kan ha större betydelse för konkurrenssituationen än ratingen

Leverantörerna menar att de resultatbaserade ersättningarna kan ha större betydelse för konkurrenssituationen än att deltagarna väljer andra leverantörer till följd av ratingen. Den främsta anledningen är att det är svårt för en seriös aktör att få lönsamhet utan resultatersättning och att det ger incitament att försöka få alla deltagare i arbete.

Detta förklarar också att de intervjuade vill att Arbetsförmedlingen ska avsluta avtalen med de leverantörer som inte lever upp till skall-kraven trots att de inte korrelerar med ratingsresultaten. De menar att en tillsyn mot skall-kraven är tillräcklig för att få bort oseriösa konkurrenter som inte eftersträvar att få så många deltagare som möjligt i arbete eller studier.

Resultatersättningens betydelse för de nationella resultaten förstärks dessutom av att högkonjunkturen gör att andelen motiverade deltagare har minskat. För att kunna få lika många i arbete som tidigare måste leverantörerna därför lyckas med en högre andel deltagare än tidigare som behöver motiveras, utbildas eller rustas på något annat sätt. Detta innebär nya incitament för leverantörerna men de tar samtidigt upp att AF borde ha anpassat hur de använder *Stöd och Matchning* till att det är högkonjunktur, till exempel genom att förlänga med ett andra block för fler deltagare.

Alla leverantörer tar också upp frågor om att de deltagare som går vidare till *Stöd och Matchning* ofta saknar egen motivation men menar samtidigt att det ingår i den upphandlade tjänsten att kunna hantera detta. Det finns en stor grupp deltagare som står så pass långt från arbetsmarknaden att STOM i praktiken är otillräcklig.

För den målgruppen har några framfört förslag om att ratingen bör kompletteras med resultatmått som mäter om de förflyttat deltagarna närmare arbetsmarknaden under den tid de deltagit i *Stöd och matchning*. Detta kan vara ett sätt att förhindra att leverantörer konkurrerar genom så kallad parkering och samtidigt uppmuntra att leverantörer utveckla lösningar som fungerar även för svåra målgrupper. De menar att deras insatser inte är bortkastade även om de inte leder till arbete inom utsatt tid.

f. Ratingen kan inte användas för jämförelser med konkurrenterna

En följd av att AF inte berättar för leverantörerna vilket aggregerat förväntat värde de har är att de också har svårt att använda uppgifterna för att förbättra sitt eget resultat. En förutsättning för att detta ska vara möjligt är att de kan bryta ned uppgifterna för att själva analysera hur de förbättringsåtgärder de vidtar påverkar ratingvärdet. I intervjuerna har en leverantör framfört att de skulle vilja få veta sitt eget förväntade värde för att kunna bedöma om de är på rätt väg.

Andra beskriver att de fortfarande försöker jämföra sig med konkurrenterna för att försöka bedöma om de är på rätt väg trots att de vet att de kan ha olika deltagarsammansättning. De flesta uppger bara att de inte har någon större användning av ratingen men att de förstås kan sätta upp egna mål baserat på sina tidigare prestationer.

De svårigheter i att jämföra sig med konkurrenterna som leverantörerna beskriver finns förstås också inom ratingssystem som baseras på genomsnittliga verksamhetsresultat. Det som har hänt är sannolikt att AF har synliggjort detta genom att använda en profilering av deltagarna för att göra ratingvärdet jämförbart mellan verksamheterna.

Därmed blir det också uppenbart att leverantörerna *kan* behöva få fler deltagare i arbete än konkurrenterna, de vet bara inte hur många fler. En konsekvens av profileringen är därför också att det blivit uppenbart att det inte självklart går att skapa kvalitetskonkurrens genom att ge leverantörerna incitament att försöka överträffa varandras kvalitetsjämförelser. Varken AF eller leverantörerna verkar däremot ha utvecklat någon ny strategi än för hur de ska använda ratingen för att förbättra sina egna och de nationella resultaten.

g. Samarbete med arbetsgivarna är avgörande för att få många deltagare i arbete

De som lyckas bra har förstås en egen strategi för vad som krävs för att få de egna deltagarna i arbete även om de inte kan bryta ned AF:s rating. Alla intervjuade beskriver att arbetsgivarrelationerna är avgörande för att kunna få många deltagare i arbete men tillägger också att detta förutsätter volym i verksamheten.

Volym är en förutsättning både för att kunna bygga arbetsgivarrelationerna och för att kunna ha ett flertal möjliga deltagare att matcha mot varje arbete. Av detta följer att leverantörerna också får tillgång till fler arbeten att matcha mot varje deltagare vilket ökar möjligheterna även för de deltagare som står längst från arbetsmarknaden. Eftersom det är högkonjunktur är det stor efterfrågan på all arbetskraft.

Eftersom volymen är så viktig för arbetsgivarrelationerna framför många intervjuade att AF anvisar för få deltagare och att det finns överetablering. Många beskriver också att den största utmaningen för att kunna få fler i arbete är att hitta arbetssökande som de kan förmedla mot de jobb som finns.

Fler av de intervjuade vill därför få tillgång till kompetensbeskrivningar för alla arbetssökande och inte bara för de som är inskrivna i *Stöd och matchning*. De menar att detta underlättar deras möjligheter att också kunna hitta jobb till de deltagare som inte kan matchas direkt mot ett ledigt jobb.

Samtidigt beskriver alla intervjuade att det kan ingå i uppgiften att ”rusta upp” den arbetssökande. Det handlar framförallt om att det motiverande arbetet är viktigt eftersom de flesta deltagare varit borta så länge från arbetsmarknaden att de har fått en föreställning om att de inte är anställningsbara. Det finns också deltagare som aldrig besökt en arbetsplats eller varit på anställningsintervju och därför behöver få grundläggande kunskaper om hur det fungerar på en arbetsplats.

Till de återkommande uppgifterna hör också att kunna informera om hur arbetslivet har förändrats de senaste åren eftersom många deltagare kan söka fler arbeten än vad de själva förstår. Andra leverantörer har mer handfasta utgångspunkter. Det kan handla om att samarbeta med arbetsgivare om korta yrkesutbildningar, till exempel i hantering av en maskin som används av ett specifikt företag eller om utbildningar om säkerhetsföreskrifter i byggbranschen och industrin.

Utvecklingsområden mot bakgrund av modellanalys och leverantörernas erfarenheter

a. Allmänna rekommendationer

Den typ av ratingsystem som AF har valt har stora likheter med den modell som används i Australien och som Svenskt Näringsliv har lyft fram som ett gott exempel. Den skillnad som finns mellan länderna handlar främst om hur ratingen används för att påverka de nationella resultaten.

AF utnyttjar dock ännu inte de möjligheter som ratingssystemet ger när det gäller att åstadkomma en förbättrad processtyrning som också innebär ett lärande i systemet.

AF bör därför dra lärdom av hur Australien har använt process- och resultatstyrning för att styra volymerna av deltagare mot leverantörer med bra resultat och därigenom skapa incitament för ytterligare förbättringar i systemet.

Lärdomar finns också när det gäller att eftersträva öppna och transparenta beslut eftersom detta är en förutsättning för att skapa ett lärande i systemet. AF har gjort andra vägval än vad Svenskt Näringsliv efterfrågar och undanhåller istället information om vilken profilering de arbetssökande har tilldelats.

Genom att använda ett profileringsinstrument har AF synliggjort de svårigheter som finns i att verksamheter försöker påverka sina resultat genom att jämföra sig med konkurrenternas ratingvärden.

Varken AF eller leverantörerna verkar däremot ha utvecklat någon ny strategi ännu för hur de ska använda ratingen eller andra kvalitetsjämförelser för att förbättra sina egna och de nationella resultaten. En förutsättning för att detta ska vara möjligt är att de kan bryta ned uppgifterna för att själva analysera hur de förbättringsåtgärder de vidtar påverkar ratingvärdet.

Eftersom AF inte berättar för leverantörerna vilket aggregerat förväntat värde de har kan leverantörerna heller inte använda de uppgifterna för att förbättra sitt eget resultat. AF bör därför utveckla en strategi om hur ratingen och profileringsinstrumentet kan användas för att ge leverantörerna incitament och förutsättningar för att kunna prestera så bra som möjligt. Några av de möjligheter som ratingsystemet ger beskrivs nedan.

Ratingen motverkar inte kvalitetskonkurrens men är inte en tillräcklig åtgärd. Arbete för att stärka och utveckling *profiling*-systemet är helt avgörande för att komma framåt.

b. Använd ratingen för att renodla Arbetsförmedlingens tjänsteutbud

En konsekvens av den rating som AF har valt att använda är att de i likhet med Australien kommer att kunna börja öppna den "svarta lådan". För att AF ska få full utväxling av det ratingsystem som de infört bör de därför också vidareutveckla hur de använder datainsamlingen som underlag för egna beslut och för att renodla sitt samlade tjänsteutbud till sådana insatser som faktiskt leder till arbete.

AF har dock valt att undanhålla det förväntade värdet med motiveringen att de vill att leverantörerna ska göra ett så bra jobb som möjligt med alla deltagare. På motsvarande sätt uppger de leverantörer som har högt resultat att det viktiga för att lyckas är att utgå från att alla är anställningsbara och att förmedla till personalen att deras uppgift är att göra ett så bra jobb som möjligt med alla deltagare.

De beskriver också att de kan sätta upp egna mål och analysera om de åtgärder företaget vidtar leder rätt även om de inte kan bryta ned AF:s rating. Det leverantörerna egentligen beskriver är alltså att de har ett systematiskt kvalitetsarbete där de eftersträvar att göra rätt så ofta som möjligt.

Leverantörerna har emellertid också framfört att det finns en förhållandevis stor grupp deltagare som står så pass långt från arbetsmarknaden att *Stöd och matchning* i praktiken är otillräcklig. De beskriver också att det är svårt att på förhand bedöma vilken typ av stöd den gruppen behöver för att kunna komma i arbete. Det är därför logiskt att både leverantörerna och AF har diskuterat åtgärder för att motverka risk för så kallad "parkering".

För leverantörerna är detta förstås en förtroendefråga eftersom det är för den målgruppen som risken för fusk är som störst. AF kan dock redan nu använda profileringsinstrumentet för att analysera om det finns systematiska resultatvariationer mellan olika grupper av deltagare.

Därmed kan risken för så kallad parkering motverkas genom att AF analyserar variationer i systemet och hur nära de önskade resultaten enskilda verksamheter presterar. Samtidigt förbättras möjligheterna att åstadkomma ett lärande i systemet av det slag som Svenskt Näringsliv efterfrågar.

Till skillnad från Australien har dock AF valt att inte inkludera uppgifter om vilken insats som har beviljats och om den faktiskt har utförts. Eftersom AF så kallade KA-system innehåller uppgifter om när deltagarna går vidare till arbete bör de uppgifterna användas för att säkerställa att de arbetssökande faktiskt har deltagit i insatser.

Australien har istället använt den typen av analyser för att sprida information om vilka åtgärder som fungerar för vilken målgrupp till fler leverantörer. De har också använt analyserna för att vidareutveckla de redskap som leverantörerna kan använda för att säkerställa att de upphandlade tjänsterna ska ha förutsättningar att leda till arbete för samtliga deltagare.

c. Skilj på att mäta för att värdera och mäta för att utveckla

Mycket i intervjuerna har kretsat kring att leverantörerna vill att oseriösa konkurrenter ska få lämna marknaden. Det är dock sannolikt att det också finns leverantörer som skulle uteslutas pga. av låg rating utan att för den skull vara oseriösa. I praktiken brukar det till exempel förekomma att verksamheter värderar sin egen kvalitet genom att jämföra sig med det nationella resultatet trots att det kan dras ned av att verksamheter som underpresterar har höga volymer. De kan då dra slutsatsen att de är bra trots att de egentligen är ganska medelmåttiga.

Ett sätt att hantera detta är att inte använda de genomsnittliga nationella resultaten som utgångspunkt för att sätta upp mål för enskilda verksamheter. AF bör istället kommunicera det förväntade värdet inklusive information om hur det nationella värdet påverkas av hur volymerna av deltagare fördelar sig mellan olika verksamheter.

Målet för *Stöd och matchning* måste vara att tjänsten ska leda till att så många som möjligt kommer i arbete även om detta innebär att leverantörer får en låg rating och utesluts för att de helt enkelt är inkompetenta. Det är dock vanligt att betrakta skillnaden mellan vad som är ”best practice” och hur de genomsnittliga verksamheterna presterar som ett implementeringsproblem.

Den typ av resultatstyrning som Australien har använt ställer betydligt högre krav på leverantörernas förmåga att ta eget ansvar för sitt systematiska kvalitetsarbete än de styrsystem som använder skall-krav för att utesluta leverantörer.

Den kanske största utmaningen att införa den typ av resultatstyrning som Australien har utvecklat kan därför vara att avstå från att toppstyra men ändå hitta strategier som leder till ett lärande i systemet. Detta är särskilt viktigt med tanke på att leverantörerna beskriver att det viktigaste för att lyckas är att bygga relationer med arbetsgivarna och samhället i övrigt. Det som ofta förbises med ett traditionellt implementeringsperspektiv är nämligen att verksamheten inte bedrivs i ett vakuum utan påverkas lika mycket eller mer av samhällsutvecklingen i övrigt.

Detta innebär också att ett lärande i systemet sker oberoende av vilka individer som idag jobbar i verksamheterna. Om ribban höjs kan verksamheten bli intressant för andra aktörer som har högre eller annan kompetens. Flera av de intervjuade leverantörerna har också efterfrågat att få gå vidare till nästa steg i dialogen med Arbetsförmedlingen om vad som krävs för att lösa matchningsproblemen. De menar att kvaliteten dras ned av att det finns för många oerfarna aktörer och att nivån i dialogerna måste anpassas efter dem.

Sammanfattningsvis har statens uppgift, i de avslutande rekommendationerna, förenklat beskrivits som att bygga en incitamentsstruktur som underlättar lärande i systemet men som också tillvaratar forskningsbaserad kunskap om vad som faktiskt är möjligt att åstadkomma.

Ratingen är tillförlitlig men har ändå inte bedömts vara en tillräcklig åtgärd för att ensam kunna leda till en förbättrad kvalitetskonkurrens och att fler deltagare går vidare till arbete eller studier. AF rekommenderas att utveckla sin resultatstyrning med Australien som förebild. Ett centralt utvecklingsområde från och med nu är en standardiserad kartläggningsprocess, likt det *profiling-system*, som använts i länder som kommit längre än Sverige vad gäller resultatstyrning.

Slutsatser

Svenskt Näringsliv poängterade i remissförfarandet inför sjösättande av *Stöd och Matchning* att ett tydligt resultatsystem likt det australiensiska är avgörande att få på plats för att förbättra resultaten i AF:s upphandlingar. Vi välkomnar att AF har arbetat kontinuerligt att få på plats ett *star rating* system och att leverantörernas prestationer redovisas.

Vi bedömer att systemet är tillförlitligt. Den metodik som används är tydligt inspirerat från Australien-exemplet. Det är positivt att även intervjuade leverantörer uppger att resultaten reflekterar de egna verksamheternas prestationer.

Icke desto mindre finns tydliga utvecklingsbehov:

För det första, ratingen måste digitaliseras, redovisas och spridas tydligare än vad som är fallet idag. I dagsläget återfinns information tio klipp bort på AF:s hemsida, i form av en pdf-fil. Sökningen på hemsidan digitalt är därtill för komplicerat.

För det andra, för att systemet ska vara träffsäkert och mer rättssäkert måste bedömningsverktyget som avgör individens behov av stöd göras mer standardiserat och objektivt. Detta är viktigt, eftersom det avgör vilket spår som den arbetslöse får ta del av och därmed mot vilken grund leverantörens prestation baseras på. Det blir dessutom allt viktigare i takt med att deltagare från Fas 3 och etableringsuppdraget kanaliseras till insatsen.

För det tredje, resultaten bör dessutom redovisas närmare det faktiska utförandet. I dagsläget är det uppemot ett års eftersläpning. En *just in time* princip bör vara ledstjärna. Ambitionen bör vara ett kvartals fördröjning.

För det fjärde, givet att en kartläggning genom ett bättre standardiserat *profiling* instrument kan säkerställas bör resultaten ha en tydligare inverkan på kontraktsförfarandet i kombination med större effekter av kundval.

Slutligen, allt detta sammantaget borgar för en uppskalning av systemet till fler programinsatser och till AF:s egna interna verksamheter. I förlängningen bör resultaten för hela AF-systemet kunna förbättras och kostnadseffektiviteten öka. Det vore önskvärt om AF-utredningen tar dessa synpunkter i skarpt beaktande inför delrapporten hösten 2017.

Fullt utfört gör vi ett grovt antagande att en effektivitetshöjning i nivå med Australien-reformerna i slutet av nittioalet skulle leda till kortare arbetslöshetstider och i förlängningen uppemot 15 000 färre arbetslösa.³⁰

³⁰ **Antaganden:** Nuvarande genomsnittsarbetslöshetstid: cirka 40 veckor. AF påverkan 16 procent, enligt Lämnatundersökning = 6,4 veckor. Effektivitetsökning enligt Australienreform: cirka 25 procent. = 1,6 veckors kortare genomsnittsarbetslöshetsperiod. Enligt VÅP 14 = En vecka mindre genomsnittsarbetslöshetsperiod ger 0,2 procentenhets lägre jämviktsarbetslöshet. Sammantaget = 0,3 enheter lägre jämviktsarbetslöshet, eller 15 000 färre arbetslösa (räknat på 7 procent och 350 000 arbetslösa). För kännedom: Riksrevisionen har räknat på en inverkan om 17 000 personer i en återrapport om effektivitetshöjning för AF-kontor i nivå med "bästa kontor" (2012).

Källor

Arbetsförmedlingen, Rapport om fördjupad granskning av kompletterande aktörer, 2013

Arbetsförmedlingen, Gerdes, Christer, Hur påverkas kompletterande aktörers marknadsandelar av rating? Working paper 2014:2, 2014

Arbetsförmedlingen, Stöd och Matchning – så hittar du företagen som erbjuder tjänsten, 2016

Australian Government, Department of Employment, , Performance Framework Guideline <https://employment.gov.au/jobactive>, 2016

British Parliament, DWP's Commissioning Strategy and the Flexible New Deal – Work and Pensions Committee, 2007

Cumming, Lauren M., Payment-by-outcome in Welfare to Work, 2020 Public Services Trust at the RSA, 2011

Department for work and pensions, Job Services Australia: design and implementation lessons for the British context, 2011

Department of work and Pensions, Star Rating system, 2017, <http://webarchive.nationalarchives.gov.uk/20100515192520/dwp.gov.uk/supplying-dwp/what-we-buy/welfare-to-work-services/star-rating-system/>

Finn, D., The British 'welfare market': lessons from contracting out welfare to work programmes in Australia and the Netherlands, The Joseph Rowntree Foundation, 2007

Finn, D., Subcontracting employment services – The design and delivery of "outcome based" and "black box" contracts, The European Commission Mutual Learning Programme for Public Employment Services, 2011

Gerdes, C, Does Performance Information Affect Job Seekers in Selecting Private Providers in Voucher-Based ALMP Programs? Institute for the Study of Labor, Bonn, 2015

Grubb, Australia's quasi-market delivery of case management (2006) (Grubb är knuten till OECD)

Hamilton, Storbritanniens jobbmirakel 2010–2015 (2015)

IFAU, Arbetsförmedlingens service och arbetsmetoder om processtyrning i en målstyrd myndighet 2011

IFAU, Bruno Crépon Gerard J. van den Berg, Active labor market policies, Working paper 2016: 17, 2016

Jansson & Kilsved "Nya aktörer inom arbetsförmedling – Spotmarknad eller dynamisk förnyelse" (2009)

OECD, Activating Jobseekers: How Australia Does It, OECD Publishing 2012

OECD, Connecting People with Jobs – Activation Policies in the United Kingdom, OECD Publishing 2014

OECD, “Activating jobseekers: Lessons from seven OECD countries”, in OECD Employment Outlook 2013, OECD Publishing 2013

OECD, “Working Together: Skills and Labour Market Integration of Immigrants and their Children in Sweden” OECD Publishing, 2016

Svenskt Näringsliv, Så vann Tyskland EM-guldet i jobbskapande – Lärdomar för Sverige?, 2016

Svenskt Näringsliv (2014) RFI (remissvar) Stöd och Matchning

https://www.arbetsformedlingen.se/download/18.4da0cf661464157a70da72/1401955708509/Svenskt_n%C3%A4ringsliv_svar_p%C3%A5_RFI_2_Kompletterande_arbetsf%C3%B6rmedlingstj%C3%A4nst_20141.pdf

The National Audit Office, The work programme, 2016

Intervjuer och samtal

Leverantörer av Stöd och matchning

Academedi koncernen, Charlotte Vällfors

Adecco, Björn Hellgren

AlphaCE, Maria Matsson Mähl

Arbetslivsresurs, Anne Buckard och Anders Wallentin

Astar, Marie Söderberg

Clustera, Mats Emriksson

Jobbklar, Mateos Falk

Manpower Group, Lena Sunnefjord Borgström

Radix Kompetens, Martin Sjöström

Arbetsförmedlingen

Christer Gerdes

Elisabeth Häggström

Mikael Edin

www.svensktnaringsliv.se

Storgatan 19, 114 82 Stockholm

Telefon 08-553 430 00