

# Utvärdering av utvärderingar – en analys av kvaliteten hos ett urval rapporter från SADEV



Beställd av Utrikesdepartementet UD-USTYR

Genomförd av Andante - tools for thinking AB  
Ekon.Dr. Kim Forss

Slutapport den 12/4 2012

## Sammanfattning

### Bakgrund

Institutet för utvärdering av internationellt utvecklingssamarbete (SADEV) bildades i januari 2006. Myndigheten har till uppgift att utvärdera det svenska internationella utvecklingssamarbetet samt reformsamarbetet i Östeuropa.

### Syfte

Syftet med den här studien är att göra en bedömning av kvaliteten i ett urval av SADEVs utvärderingar. Underlaget ska nyttjas av utrikesdepartementets enhet för styrning av utvecklingssamarbetet inom ramen för den löpande uppföljningen av SADEVs verksamhet.

### Metod

Studien bygger på en granskning av fem utvärderingar som SADEV publicerat under de senaste två åren. Granskningen bygger på en modell för kvalitet inom utvärderingsområdet. Enligt modellen definieras en bra utvärdering av att vara (1) användbar, (2) tillförlitlig, (3) effektivt genomförd, och (4) i enlighet med etiska normer.

*De fem utvärderingarna känneteckna i allmänhet av:*

- Väl valda och relevanta ämnen.
- Har genomförts i tid och levererats enligt beställning.
- När potentiella användare så att de kan användas.
- Utvärderingsfrågorna är fokuserade, relevanta och möjliga att utvärdera.
- Det finns överenskommelser med organisationer som berörs och processer för samverkan.
- Interna system för kvalitetskontroll.
- Ordentliga resurser ifråga om tid och pengar.
- En stark oberoende ställning
- Inga intressekonflikter som kunnat påverka dess omdömen.

*Många av de fem utvärderingarna har ett antal brister:*

- Användbarheten av slutsatser och rekommendationer är låg, de som framförs är allmänna, saknar konkretion, och är inte tillräckligt

praktiska för att ge vägledning i beslutssituationer.

- Bidrar i liten utsträckning till organisatoriskt lärande.
- Metodologiska brister vad gäller upplägg samt vad gäller urval, utformning av instrument för datainsamling, utveckling av skalor för att mäta och bedöma det som utvärderas.
- Transparensen är låg och metodkapitel innehåller inte information så att läsaren kan bedöma trovärdighet.
- Utvärderingarna bedömer inte resultat utan stannar vid att behandla processer och prestationer. Endast en av fem rapporter behandlar den typ av resultat som beskrivs som 'outcome', ingen behandlat det som benämns 'impact'. Ingen av de fem behandlar sido-effekter.
- Resurserna i tid och pengar används inte effektivt. Utvärderingarna har lagt för lite tid på empiriskt datainsamling.
- Dialogen med mottagare/användare har inte genomförts till slut och de synpunkter och/eller invändningar som framkommit reflekteras inte i rapporterna.

### Rekommendationer

Den här studien har inte haft i uppgift att granska orsakerna till kvalitetsproblemen. De förslag som lämnas är därför allmänna och tentativa. Utvecklingen av kvaliteten bör bygga på åtgärder inom åtminstone tre områden:

Det första gäller utvärderingsteam, som i högre grad bör bestå forskare/konsulter från samarbetsländer inom utvecklingssamarbetet och med internationell expertis inom respektive område

Det andra gäller kompetens där ett långsiktigt program för kapacitetsutveckling bör utvecklas och genomföras.

Det tredje gäller kvalitetssäkring, där de processer SADEV har institutionaliserat inte visat sig fungera. Det bör utvecklas nya rutiner så att kvalitetssäkring påbörjas tidigt, att referensgrupper har tid och möjlighet att påverka arbetet, samt att interna seminarier är kritiska och ifrågasättande samt uppmuntrar kritik, diskussion och lärande.

## Innehållsförteckning

<b>SAMMANFATTNING</b> .....	<b>2</b>
<b>INNEHÅLLSFÖRTECKNING</b> .....	<b>3</b>
<b>INLEDNING</b> .....	<b>5</b>
BAKGRUND OCH SYFTE .....	5
VAD ÄR KVALITET HOS EN UTVÄRDERING? .....	5
TIDIGARE STUDIER AV UTVÄRDERINGARS KVALITET .....	6
METODER.....	7
LÄSANVISNING .....	8
<b>SADEV REPORT 2010:5. UTVÄRDERING AV DEN SVENSKA STRATEGIN FÖR MULTILATERALT UTVECKLINGSSAMARBETE</b> .....	<b>9</b>
ANVÄNDBARHET.....	9
TILLFÖRLITLIGHET .....	10
ETIK .....	11
EFFEKTIVITET .....	12
<b>SADEV REPORT 2010:6. EVALUATION OF SWEDEN'S IMPLEMENTATION OF THE PARIS DECLARATION</b> .....	<b>13</b>
ANVÄNDBARHET.....	13
TILLFÖRLITLIGHET .....	14
ETIK .....	16
EFFEKTIVITET .....	16
<b>SADEV REPORT 2010:9. GENDER EQUALITY IN AND THROUGH EDUCATION</b> .....	<b>17</b>
ANVÄNDBARHET.....	17
TILLFÖRLITLIGHET .....	18
ETIK .....	19
EFFEKTIVITET .....	20
<b>SADEV REPORT 2011:1. MULTI-BI-FINANSIERING – UTVÄRDERING AV BEREDNING INFÖR BESLUT</b> .....	<b>21</b>
ANVÄNDBARHET.....	21
TILLFÖRLITLIGHET .....	22
ETIK .....	23
EFFEKTIVITET .....	23
<b>SADEV REPORT 2012:1. DEMOKRATISK UTVECKLING OCH ÖKAD RESPEKT FÖR MÄNSKLIGA RÄTTIGHETER – RESULTAT AV SVENSKT BISTÅND</b> .....	<b>25</b>
EFFEKTIVITET .....	25
ANVÄNDBARHET.....	26
TILLFÖRLITLIGHET .....	28
ETIK .....	29

<b>SLUTSATSER OCH REKOMMENDATIONER</b> .....	<b>30</b>
ALLMÄNT OM KVALITET .....	30
ANVÄNDBARHET.....	31
TILLFÖRLITLIGHET.....	32
ETIK .....	32
EFFEKTIVITET .....	33
REKOMMENDATIONER.....	34

<b>REFERENSER</b> .....	<b>36</b>
-------------------------	-----------

<b>BILAGA 1. UPPDRAGSBESKRIVNING</b> .....	<b>37</b>
<b>BILAGA 2. INTERVJULISTA</b> .....	<b>38</b>
<b>BILAGA 3. SYNPKTER PÅ UTKAST TILL RAPPORT</b> .....	<b>39</b>

## Inledning

### Bakgrund och syfte

Institutet för utvärdering av internationellt utvecklingssamarbete (SADEV) är en myndighet under regeringen<sup>1</sup>. Den bildades i januari 2006 och har nu varit verksam under sex år. Myndigheten har till uppgift att genomföra utvärderingar av det svenska internationella utvecklingssamarbetet samt reformsamarbetet i Östeuropa. Utvärderingarna ska avse såväl resultat och effekter av utvecklingssamarbetet som effektiviteten i användningen av biståndsmedel. Vidare ska utvärderingarna ha hög relevans som underlag för förbättringar av beslut om Sveriges internationella utvecklingssamarbete<sup>2</sup>.

Utvärdering är en kritiskt granskande aktivitet som lätt kan bli impopulär bland dem som granskas. Utvärdering av utvecklingssamarbete berör ett stort antal aktörer i Sverige och internationellt. Det är en starkt värdeladdad verksamhet. Ofta är det svårt att bedöma effekter, och det finns en livlig debatt om utvecklingssamarbetets nytta och värde.

I en sådan miljö är det av största vikt att utvärderingarna är väl genomförda. Utvärderingar ska bygga på vetenskaplig metod och ska vara tillförlitliga som beslutsunderlag och källor till lärande. Det är dock ganska sällan utvärderare granskas med samma metoder som utvärderare själva granskar andra. Det innebär dock inte att utvärderingar inte blir bedömda – de flesta som läser en utvärdering gör också en recension om den var bra eller dålig. Det är knappast någon hemlighet att det finns olika åsikter om SADEVs utvärderingar, konstigt vore det annars. Det finns även skilda åsikter om kvaliteten på Sidas utvärderingar och samma gäller de flesta institutionaliserade former för utvärdering.

Syftet med den här studien är att göra en bedömning av ett urval av SADEVs utvärderingar. Underlaget ska nyttjas av utrikesdepartementets enhet för styrning av utvecklingssamarbetet inom ramen för den löpande uppföljningen av SADEVs verksamhet. Underlaget ska också kunna nyttjas av Statskontoret inom ramen för dess översyn av utvärderingsverksamheten inom internationellt utvecklingssamarbete. Utvärderingens uppdragsbeskrivning framgår av bilaga 1.

### Vad är kvalitet hos en utvärdering?

Inom utvärderingssamfundet har man ägnat mycket tid och uppmärksamhet åt frågan om vad som utgör kvalitet hos en utvärdering. Redan 1974 bildades en kommitté (Joint Committee of Standards) med representanter från förvaltning, konsultvärlden och forskare för att definiera kvalitet. 1994 publicerades kommitténs förslag på kvalitetsstandard<sup>3</sup> och blev antagna av den amerikanska standardiseringsbyrån<sup>4</sup>. En utvärdering är med andra ord standardiserad på i princip samma sätt som ett eluttag eller en diskmaskin.

Kvalitet kommer från latinets '*qualitas*' som betyder egenskap. Ordet säger i sig inte vad de önskade egenskaperna ska vara. Det behöver preciseras vidare och det är vad the Program

---

<sup>1</sup> SADEV är en förkortning av det engelska namnet Swedish Agency for Development Evaluation

<sup>2</sup> Information från SADEVs hemsida; [www.sadev.se](http://www.sadev.se)

<sup>3</sup> The Program Evaluation Standards. The Joint Committee on Standards for Educational Evaluation. James R. Sanders, Chair. Sage Publication, Thousand Oaks.

<sup>4</sup> Approved by the American National Standards Institute as an American National Standard. Approval Date 15/3 1994. JCSEE PR-1994.

Evaluation Standards gör. De beskriver trettio olika egenskaper som en bra utvärdering bör ha. Dessa trettio egenskaper sammanfattas under fyra rubriker. En bra utvärdering ska vara (de ursprungliga 'kvaliteterna' inom parentes):

1. användbar (UTILITY)
2. tillförlitlig (ACCURACY)
3. etisk (PROPRIETY)
4. effektiv (FEASIBILITY)

Det första kriteriet, användbarhet, innebär att en bra utvärdering ska ha klara och tydliga rekommendationer, en analys som har ett nyhetsvärde för läsarna och som tillför kunskap, antingen för lärande eller för en beslutssituation (vilket även kan innefatta att med tillförlitliga metoder bekräfta ett resultat eller en uppfattning om hur t.ex. en organisation fungerar)<sup>5</sup>, eller för den kontroll som också är utvärderingens funktion. Det innebär att utvärderarna ska ha legitimitet och att de som berörs av utvärderingens resultat bereds utrymme att ge synpunkter.

Det andra kriteriet, tillförlitlighet, innebär att en bra utvärdering ska bygga på säkra och beprövade metoder, dvs. den ska ha intern och extern validitet och metoderna för datainsamling ska ge tillförlitlig information. Vidare ska utvärderingen ha ett upplägg som är relevant för problemformuleringen, och ett urval av t.ex. fallstudier, intervjuer, enkäter, som ger möjlighet till slutsatser om utvärderingsfrågorna. Detta ska också kunna utläsas i ett metodavsnitt så att läsaren har en chans att själv bedöma tillförlitlighet.

Det tredje kriteriet, etik, innebär att en bra utvärdering genomförs i en process där berörda parter legitima intressen respekteras och där uppgiftslämnare skyddas och där berörda parter integritet respekteras. Det innebär även att en utvärdering inte ska belasta den organisation som utvärderas mer än 'rimligt' och att det finns formella överenskommelser kring utvärderingen och dess processer. Utvärderingen ska leda till en rättvis bedömning och utvärderarna får inte ha några lojalitetskonflikter.

Det fjärde kriteriet, effektivitet, innebär bl.a. att utvärderingen ska arbeta effektivt, dvs. man ska inte lägga ned mer resurser än nödvändigt på att uppnå syftet. I begreppet 'feasibility' ligger även att utvärderingen ska vara möjlig att göra. Ibland kan det vara omöjligt att genomföra en utvärdering på grund av starka och upprörda känslor, organisatoriska konflikter mm. I så fall ska man inte heller ge sin in på det. Utvärderingar ska vara klara i tid och de ska vara relevanta.

### Tidigare studier av utvärderingars kvalitet

Det finns inte alltför många studier av kvalitet i utvärderingar av internationellt utvecklingssamarbete. Ett tidigt och intressant exempel utgörs av USAID (1988) som bedömde kvalitet samtidigt som ovan nämnda standards utvecklades. Det var den första praktiska tillämpningen av dessa, men följdes snart av en kvalitetsgenomgång av kanadensiska CIDA (Watson, 1990). Ett något modifierat format som dock följde de fyra huvudområdena ovan, om än inte de 30 underrubrikerna, gjordes i Norge (Samset, Forss och Hauglin, 1991). I Sverige gjordes två granskningar av utvärderingskvalitet, en avseende samarbete med Östeuropa (Forss, 1994) och en annan avseende det bilaterala utvecklingssamarbetet (Forss, 1994). Båda pekade på problem, särskilt när det gällde vetenskaplighet och frågan om att bedöma effekter. Detta ledde vidare till analyser av hur utvärderingsinstrumentet används (Carlsson, Forss, et al, 1997) och till en metodutveckling vad gäller meta-utvärdering, dvs. utvärdering av utvärderingar

---

<sup>5</sup> När sådana bedömningar görs nedan bygger de på såväl de formella aspekterna av användbarhet som en kontextuell prövning av om rekommendationer och den skrivna texten kan uppfattas som användbara av aktörer i systemet, men det är dock min bedömning – utifrån mina kunskaper om utvecklingssamarbete.

(Forss, Carlsson och Rebien, 1997). Ett stort urval av Sida utvärderingar granskades för några år sedan (Forss, Kruse, Mwaiselage, Nilsson och Vedung, 2008).

Är det då så att det är en större uppmärksamhet på kvaliteten i utvärderingar när det gäller internationellt utvecklingssamarbete än på andra områden? Den mest omfattande analysen av kvalitet har gjorts av Schwartz och Mayne (2008) och där behandlas skilda politikområden. I Sverige har det gjorts studier av kvalitet i utredningsväsendet (Forss och Uhrwing, 2002) samt kvalitet på utvärderingar inom jordbrukspolitiken och inom folkbildning. Frågan om kvalitet är med andra ord aktuell överallt. Det är viktigt att det som ska utgöra underlag för policybeslut går att lita på. Konferenser som ordnas har på senare tid ofta haft kvalitet som överordnat tema – intressant nog hade både the American Evaluation Association, den svenska utvärderingsföreningen och den norska utvärderingsföreningen det som tema under 2011. Men det verkar som om uppmärksamheten på kvalitet har varit störst inom internationellt utvecklingssamarbete. OECD/DAC har utvecklat kriterier efter the Program Evaluation Standards, och inom EU kommissionen har EuropeAid lett ett liknande arbete. DFID har nyligen gjort en omfattande studie i syfte att stärka metoderna för effektutvärdering (Stern, Befani, Davies, Forss, Mayne och Stame, 2012).

## Metoder

Den här studien bygger på de metoder som utvecklats i de ovan refererade meta-utvärderingarna. Grunden är lagd genom de fyra kvalitetskriterierna från the Program Evaluation Standards och de trettio underkriterierna.

Jag har gjort ett urval av fem nyligen genomförda SADEV utvärderingar. Under åren 2010 – 2012 genomförde och publicerade SADEV 15 utvärderingar<sup>6</sup>, och det här urvalet representerar knappt en tredjedel av det antalet. Urvalet gjordes utifrån kriterierna att se på stora utvärderingar av breda teman i utvecklingssamarbetet. Jag ville också ta med den senast genomförda utvärderingen. Följande utvärderingar har ingått i studien:

- SADEV Report 2010:5. Utvärdering av den svenska strategin för multilateralt utvecklingssamarbete
- SADEV Report 2010:6. Evaluation of Sweden's Implementation of the Paris Declaration
- SADEV Report 2010:9. Gender equality in and through education
- SADEV Report 2011:1. Multi-bi-finansiering – utvärdering av beredning inför beslut
- SADEV Report 2012:1. Demokratisk utveckling och ökad respekt för mänskliga rättigheter

Den senaste rapporten är en syntesstudie som bygger på fem delrapporter som även de publicerats. Rapporten '*Gender equality in and through education*' är en av tre rapporter om jämställdhet, men de är sinsemellan oberoende och fristående som rapporter. De rapporter som inte valts handlar t.ex. om landstrategier/samarbetsstrategier, om biståndsstatistik, om svenska myndigheter som biståndskanaler, politiska organisationers biståndsarbete. De fem som valts utgör ett någorlunda representativt urval, även om varje utvärdering är unik och måste förstås i sin egen rätt. Kan man då säga att de kvalitetsproblem och de förtjänster som uppmärksammas här gäller andra studier med? Det finns mycket som talar för det, bl.a.:

---

<sup>6</sup> Uppgift från SADEVs hemsida, rubriken publikationer och underrubriken SADEV reports. De fem studier som ingår i utvärderingen av Demokrati och MR redovisas som separata rapporter men ingår i samma utvärderingsarbete under en budget och arbetsplan.

- De utvärderingar som bedömts är stora och viktiga, en del är regeringsuppdrag och andra är internationella åtagande. De borde utgöra bland det bästa SADEV åstadkommit och bör ha varit föremål för mer omsorg än andra utvärderingar.
- Kvalitetskontrollen bygger bl.a. på interna seminarier, så om problem slinker igenom är det mycket som talar för att det saknas en miljö som leder till att just sådana problem uppmärksammas. Då finns de sannolikt även bland andra utvärderingar.

Under arbetet har jag läst igenom varje rapport och därefter sammanfattat mina synpunkter på den, enligt de rubriker som följer ur Program Evaluation Standards. Det finns andra standards att välja, bl.a. de som utformats av OECD/DAC. Även andra biståndsorganisationer har standards för utvärderingar, t.ex. UNDP, DfID, USAID och EuropeAid. Det finns två huvudtyper av standards, sådana som i likhet med ISO huvudsakligen anger processkriterier och sådana som huvudsakligen bygger på en bedömning av substans. OECD/DACs kriterier är processinriktade, men erfarenheten av att tillämpa standards visar att det är viktigt att bedöma substans, dvs. att se på innehållet i analysen, slutsatserna och rekommendationerna – inte bara på de processer genom vilka de tillkommit. Min bedömning bygger på både processkriterier och på en analys av substans i utvärderingarna. Bedömningarna är gjorda på samma sätt som t.ex. bidrag till vetenskapliga artiklar eller konferensbidrag av utvärderingskaraktär granskas, dvs. genom en 'peer review'. Jag bör nämna att jag inte har några kopplingar till SADEV och inte varit involverad i några av dessa utvärderingar.

Efter analysen av rapporterna har jag kontaktat en representant för de utvärderare som genomfört studien, samt en eller två representanter för vad som kan sägas vara användarsidan, dvs. Utrikesdepartementet eller Sida. Syftet var att belysa hur utvärderarna resonerat kring metodval och genomförande, samt hur potentiella användare ser på användbarhet och hur man eventuellt använt rapporterna. Det är dock min bedömning av rapporterna som redovisas nedan. Jag har också utfört en intervju med SADEVs generaldirektör. Intervjuerna framgår av bilaga 2.

Ett problem i detta sammanhang var att i samband med tre av de fem utvärderingarna menade de jag kontaktade i deras egenskap av utvärderare att den rapport SADEV publicerat inte reflekterade deras arbete. Det fanns då utkast till rapport som inte godkänkts, och det som blev slutrapport har å andra sidan inte godkänkts av utvärderarna. Det har funnits någon konflikt inom SADEV och mellan/inom utvärderingsteam som jag inte haft tid, möjlighet eller intresse att fördjupa mig i. Min bedömning handlar om de rapporter som är godkända och publicerade i SADEVs namn.

Arbetet har genomförts under perioden 13/2 till 9/3 2012, med sammanlagt 18 arbetsdagar inklusive ett seminarium på utrikesdepartementet för presentation och diskussion av en första version av rapporten. Därefter sändes rapporten den 9 mars till SADEV för kommentarer och den 12 april överlämnade SADEV de kommentarer som återfinns i bilaga 3. Jag har valt att inte kommentera kommentarerna, och tycker mig inte ha anledning att rätta några faktafel i de sidor som följer.

## Läsanvisning

I texten som följer behandlas utvärderingarna en och en. Varje rapport diskuteras med avseende på användbarhet, tillförlitlighet, etik och effektivitet. Jag har valt att inte sammanfatta dessa egenskaper på någon enkel skala av typen "mycket bra – bra – mindre bra", även om det i princip låter sig göras. Jag tror dock att de viktigaste lärdomarna i den här analysen står att få i den substantiella bedömningen och diskussionen av starka och svaga sidor hos rapporterna. Den läsare som inte hinner gå in i varje rapport kan istället ta del av de allmänna slutsatserna och rekommendationerna i slutet av rapporten. Studien bör läsas parallellt med rapporterna. Jag utgår från att läsaren känner till rapporterna eller lätt kan skaffa dem. Jag har inga beskrivningar



av deras innehåll och antar också att läsarna är väl bekanta med såväl utvärderingsterminologi som begrepp i utvecklingssamarbetet.

## SADEV Report 2010:5. Utvärdering av den svenska strategin för multilateralt utvecklingssamarbete

Den här utvärderingen fick SADEV i uppdrag att genomföra genom regeringsbeslut i januari 2010, och den färdiga utvärderingen överlämnades till regeringen i augusti samma år. Utvärderingen är genomförd av 2 personer från SADEV.

### Användbarhet

De formella kraven på användbarhet är väl uppfyllda. Rapporten har ett rimligt omfång på 45 sidor huvudtext. Det finns en sammanfattning på knappt 4 sidor. Rekommendationer är samlade i det femte och sista kapitlet. Rapporten har en klar och tydlig struktur. Utvärderingen diskuterades vid ett möte till vilket UD, andra berörda departement och Sida hade inbjudits. UD har sedan skrivit en s.k. "management response" vilken i princip instämmer i stora delar av analysen, men påpekar att arbete i linje med rapportens rekommendationer antingen redan pågår, eller att det inte är lämpligt/relevant.

Det är därför nödvändigt att se närmare på rekommendationerna för att få en förståelse för deras användbarhet. Jag kommer inte att gå igenom alla de nio rekommendationerna utan lyfter fram några exempel. Den första rekommendationen är att *"fortsätta implementeringsarbetet så att Multistrategins element genomförs i sin helhet"*. Frågan är då om detta är en användbar rekommendation? I en situation där aktörerna lämnade den strategiska inriktningen skulle det utgöra en påminnelse. Om någon aktör hade uppfattningen att Multistrategin var behäftad med allvarliga fel, eller att de tre senaste åren visat på bristande relevans – ja, då hade en sådan rekommendation tillfört en viktig synpunkt till beslutsprocessen. I det här fallet är det egentligen ingen som inte delar uppfattningen att man ska fortsätta implementeringsarbetet så att Multistrategin genomförs i sin helhet. Det är ett exempel på en rekommendation som är självklar och därmed inte särskilt användbar. De två följande rekommendationerna är av samma karaktär; nämligen att (1) *"integrera arbetet med bedömningar och organisationsstrategier"*, och (2) *"överväga att låta bedömningen påverka finansieringen och prioriteringar"*. I båda fallen är det något som redan var praxis och som det fanns stor enighet om. Rekommendationerna i rapporten var inte fel, men de bidrog inte heller med nytt och användbart beslutsunderlag.

Andra rekommendationer är mer kontroversiella, som t.ex. att *"överväga att överföra huvudansvaret för fackorganens utvecklingssamarbete till biståndsförvaltningen"*. Nu säger rekommendationen visserligen inte att detta ska ske, utan att det ska övervägas. Frågan har dock övervägts tidigare och den rådande ordningen är ett resultat av detta. Det finns sedan tidigare en analys av fördelar och nackdelar med det huvudmannaskap som nu finns och för att rekommendationen ska ha kraft borde den ta sin utgångspunkt i vad som är fel i den nuvarande ordningen. Rekommendationen är heller inte specifik, den säger inte vem som ska överväga eller i vilka former – en intern utredning, uppdrag till managementkonsulter, Statskontoret, eller något annat. Till skillnad från ovanstående rekommendationer är detta ett förslag som tillför en synpunkt och förstärker vad som kanske, men inte nödvändigtvis, är en minoritetssynpunkt. Problemet vad gäller användbarhet är dock att den är generell och det är oklart vad som ska ligga i "övervägande" som går bortanför de övervägande som ligger till grund för den rådande

arbetsfördelningen. Detsamma kan sägas gälla rekommendationer och förslag som att *"utveckla samarbetet med utlandsmyndigheterna"* och att *"engagera det civila samhällets organisationer mer i arbete"*. För att vara användbar skulle sådana förslag behöva gå från den generella utsagan till mera konkreta anvisningar om hur det skulle ske.

Användbarhet skapas också genom att en rapport är tillgänglig och läsarvänlig. Den så kallade konceptuella användningen uppstår när läsare lär av analysen, till skillnad från den instrumentella användningen av rekommendationer. På nivå under kapitelrubrikerna är det svårt att följa logiken. I kapitel 3 analyseras genomförandet och författarna går igenom var och en av Multistrategins sju rekommendationer på cirka 2 sidor. Den texten delas in i tre huvudmoment; nulägesbeskrivning, analys och iakttagelser. Problemet är att nulägesbeskrivningen blir analys, och att beskrivningen leder till en värdering. Likaså i sektionen analys, där återkommer beskrivningen. Iakttagelser är snarare en form av preliminära rekommendationer. Detta leder till betydande överlappningar i texten och utrymmet räcker inte för det djup i både beskrivning och analys som en annan struktur kunde göra möjligt. En stor del av utvärderingens rekommendationer förhåller sig till dessa sju rekommendationer i Multistrategin, och på de cirka 2 sidor text som varje rekommendation har återkommer både beskrivning och analys från kapitel 3. Det innebär att de 45 sidorna inte utnyttjas på bästa sätt och det drar ned användbarheten genom att det blir svårt att följa en röd tråd i texten (och det räcker inte med en röd tråd i kapitelrubriker). Slutsatsen blir att användbarheten i rapporten är relativt låg, och även om det finns en tydlig struktur och omfattande rekommendationer är dessa inte i sak så konkreta eller nydanande att de kan spela någon väsentlig roll i beslut om Multistrategins fortsatta genomförande.

### Tillförlitlighet

Tillförlitlighet innefattar frågan om huruvida slutsatser och rekommendationer är väl underbyggda. Den övergripande slutsatsen är att genomförandet av Multistrategin i stort sett varit rätt i sak men att det går för långsamt. Det är egentligen en allvarlig kritik, men frågan är hur man underbygger ett sådant värdeomdöme. Utvärderingar kan i sådana fall bygga på två olika former av datainsamling och analys, och i bästa fall kombineras de två.

- Det ena är att fråga de som arbetar i förvaltningen och höra vad de anser. Utvärderingens intervjuprotokoll innehåller dock ingen sådan fråga, så det går inte att hänvisa till någon 'allmän mening' i saken. UDs management respons reflekterar en annan syn på saken.
- Den andra vägen innebär att utifrån kunskap om hur lång tid organisatoriska processer brukar ta, och bör ta, göra en bedömning. Detta är sannolikt vad SADEVs utvärderare gjort, men utan att klargöra vilka grunder man haft.

Eftersom det är ett omfattande reformarbete som pågår skulle man också kunna ha uppfattningen att det går ungefär som man kunde förvänta sig, eller att det faktiskt går ganska snabbt. Det är alltid svårt att göra sådana bedömningar och de måste motiveras avsevärt grundligare för att vara trovärdiga.

Det finns en rad omdömen som framstår som trovärdiga och som är viktiga. Utvärderingen redovisar arbetet med att rekrytera svenskar till internationella organisationer och sakförhållanden kring vad som skett – och inte skett, framstår som korrekta. Utvärderingen går igenom processerna med att göra organisationsbedömningar och organisationsstrategier, och redovisar även de samrådsförfarande som introducerats. Även detta är trovärdigt.

Utvärderingen riktar kritik mot organisationsbedömningarnas kvalitet, som det hävdas är låg. Frågan är hur processen att göra organisationsbedömningar ska förstås. Vad är det som händer? Det är givetvis helt otänkbart att en handläggare på utlandsmyndighet eller UD skulle kunna

göra en egen organisationsanalys av t.ex. Världsbanken av hög kvalitet. Det är ett arbete som tar forskare/utvärderare och managementkonsulter med många års erfarenhet lång tid att göra. En nyligen avslutad extern utvärdering av UNESCO sysselsatte 12 högt meriterade utvärderare i 10 månader till en kostnad av drygt 3 miljoner kronor. Vad handläggarna kan göra i sina organisationsbedömningar är att sammanfatta och använda den typen av studier. Ibland finns sådana, ibland inte, och det är det som avgör kvaliteten och tillförlitligheten i organisationsbedömningarna – inte de interna arbetsprocesserna.

Ett annat exempel på en tveksam rekommendation är den att överföra huvudmannskapet för utvecklingssamarbete genom FN-systemets fackorgan till biståndsförvaltningen. Det enda motiv jag funnit för den rekommendationen utgörs av två paragrafer text på sidan 18. Dessa paragrafer innehåller dock ingen analys av frågan och inga övervägande av fördelar eller nackdelar med den föreslagna lösningen. Texten innehåller även felaktigheter om fackorganens roll i utvecklingssamarbete och har en beskrivning av de svenska departementens roll som verkar yttlig (näst sista meningen, näst sista paragrafen, sid 18.) Författarna hänvisar till att arbetsfördelningen uppfattas som komplicerad av informanterna, men intervjuistan innehåller bara en intervju vid ett fackdepartement. Det är med andra ord bara en person av ett ganska stort antal potentiella respondenter på övriga departement som uttalat sig i frågan. Det är en rekommendation som inte bygger på en trovärdig analys.

Rapportens metodkapitel är av avgörande betydelse för att bedöma trovärdigheten. Metodkapitlet är lagom långt med tanke på rapportens omfång och innehåller väsentlig information om datainsamling. Det är bra att intervjuistor och intervjuguider återfinns i bilagor. Metodavsnittet inleder med att beskriva det övergripande angreppssättet som är programteoriutvärdering. Om utvärderingen vore en fullständig analys av Multistrategins effekter skulle det vara en utmärkt (kanske enda möjliga) design. Men så är det inte. Utvärderingen analyserar Multistrategins rekommendationer och om de är genomförda. Det är inte en programteoriutvärdering. Figur 1 på sidan 3 visar heller inte en programteori eftersom den inte utvecklar hur vissa orsaker (Multistrategin) förväntas få effekt. Programteori används för att göra explicit kausala samband, men några sådana framkommer inte i figur 1. Figurens element är heller inte konsistenta, distinktionen mellan processer (figuren visar här aktörer snarare än processer) och prestationer (som blandar processer och prestationer) är inte fullföljd. Utvärderarna påpekar att man inte kan vare sig bedöma eller redogöra för effekter, och då blir det ändå inte en programteoriutvärdering eftersom en sådan används för att tydliggöra kausala samband.

Ett metodkapitel bör också informera läsaren om eventuella problem som påverkar tillförlitligheten. Utvärderingen redovisar tydligt sitt urval av organisationer, intervjuer, liksom svarsfrekvensen på frågeformulär. Andra delar av rapporten behandlar avgränsningsfrågor, tidsaspekten, och omfattningen av frågeställningarna. Det är ju delvis metodfrågor, och har alltid metodkonsekvenser, och om de diskuteras i metodkapitlet eller på annat håll, spelar ingen större roll.

## Etik

En av frågorna när det gäller en utvärderings etik gäller uppgiftslämnarens integritet. Jag kan inte se att det har varit föremål för någon diskussion här. Metodkapitlet säger ingenting om de intervjuades anonymitet, men eftersom det är ganska få som är intervjuade kan den som känner miljöerna säkert knyta en del åsikter till personer. Å andra sidan är det knappast något problem. Som ett exempel är det bara två personer från ett fackdepartement som är intervjuade, och det hänvisas till synpunkter på fackdepartementets roll, så det kommer sannolikt från de personerna. Hade det varit ett mera känsligt område skulle det ha varit ett problem, men nu är det knappast det.

Ytterligare ett av de etiska kriterierna gäller hur de personer/organisationer som utvärderas kommer till tals med invändningar på rapporten. Utkast till rapporten sändes till UD-MU och Sida och det kom många synpunkter på felaktigheter, skilda bedömningar, saker att förtydliga, utveckla, mm. Även när det gäller den slutliga rapporten fanns sådana synpunkter. Det verkar inte som att utvärderarna har beaktat de synpunkter som lämnades i någon större utsträckning. Om utvärderarna inte håller med om kritiken kan man dock hantera det. I många organisationer är det vanligt att en slutversion av rapporten innehåller en bilaga med synpunkter från en remissomgång. Det ger möjlighet att korrigera faktafel och att tillföra ett perspektiv eller andra bedömningar. En del av en utvärdering är alltid en bedömning, som t.ex. frågan ovan om implementeringsgraden av Multistrategin. Det är viktigt att sådana åsiktsskillnader kommer fram i processen.

## Effektivitet

Det finns fungerande marknader för utvärdering och de flesta utvärderingar upphandlas genom öppna anbudsprocesser. Det är därför möjligt att jämföra tidsåtgång i denna utvärdering med andra strategiutvärderingar av liknande karaktär. Utvärderarna har läst bakgrundsinformation; referenslistan upptar ett 40 tal titlar och det är ett arbete som sammanlagt bör klaras av på cirka 5 arbetsdagar. Vidare har man genomfört 26 intervjuer, vilket med god marginal bör kunna genomföras på 13 arbetsdagar. En enkät till 18 respondenter bör kunna genomföras på cirka 3 – 4 arbetsdagar. Att analysera empiri och skriva en rapport på cirka 50 sidor bör kunna genomföras på cirka 10 arbetsdagar. Till allt detta kommer planering, utformning av intervjumallar, mm. Det kan uppskattas till cirka 10 dagar. Sammanlagt bör en utvärdering av den här karaktären kunna genomföras inom 40 – 50 arbetsdagar. Allt utöver det bör betraktas som mindre effektivt. Utvärderingen genomfördes under tiden mars 2010 till oktober 2010 och ianspråk tog 2 heltidstjänster. Budgeten var 920 000 kronor.

The Program Evaluation Standards talar om 'Feasibility' snarare än effektivitet. Det senare är en väsentlig del av feasibility, men begreppet omfattar också frågan om utvärderingen var möjlig. Var det ett rimligt uppdrag, fanns det tid och resurser att genomföra det, låg det rätt i tiden? Det finns en del som talar för att det inte var ett 'möjligt' uppdrag i den här bemärkelsen.

- För det första tidsaspekten: det är tveksamt om tre år är tillräckligt lång tid för att kunna bedöma effekter av Multistrategin, särskilt som det var tre mycket turbulenta år för biståndsförvaltningen. Utvärderingen hade inget mandat att förhålla sig till de organisatoriska processer som påverkat genomförandet, t.ex. att Sverige var ordförande i EU och att det skedde stora omorganisationer på Sida. Båda dessa faktorer påverkade arbetet med Multistrategin.
- För det andra avgränsningen: utvärderingen har inte i uppdrag att bedöma Multistrategin i sig. I princip utgår man från att den är perfekt. Den ifrågasätts inte och ställs inte i förhållande till andra aspekter av politiken. En strategiutvärdering bör även värdera själva strategin eftersom en del implementeringsproblem kan ha sin grund i element i strategin.
- För det tredje initieras ett antal begrepp, t.ex. kvalitetssäkring och effektivitetshöjning i det svenska multilaterala utvecklingssamarbete som är viktiga, men komplicerade. Det är lätt att det blir tom retorik kring dem, och frågan hur de kan operationaliseras i en utvärdering är svår. SADEVs utvärderare hade inga ingångsvärde att förhålla sig till och hade eller inga möjligheter att egentligen bedöma vare sig kvalitetssäkring eller effektivitet.

## SADEV Report 2010:6. Evaluation of Sweden's Implementation of the Paris Declaration

SADEV beslöt att delta i den internationellt planerade och genomförda utvärderingen av hur Parisdeklarationen genomförts. I arbetet deltog ett stort antal mottagarländer, givarländer och internationella organisationer. En syntesrapport sammanfattar arbetet. Den här behandlade studien har följt de gemensamma riktlinjerna med utvärderingsfrågor och metodanvisningar. Utvärderingen genomfördes av tre utvärderare från SADEV.

### Användbarhet

Det finns två huvudsakliga avnämare till utvärderingen. För det första är den gjord som en del i den internationella utvärderingen av Parisdeklarationen. SADEVs utvärdering borde komma till användning i den slutrapporten. Användning innebär då att den svenska studien bidrar till helheten, att den läses, citeras och ger ett väsentligt bidrag till den analys (och syntes) som det internationella arbetet utgör. De som skulle använda den svenska studien är då först och främst det internationella team som ledde den utvärderingen.

SADEV-utvärderingen har använts i syntesstudien, men den spelar inte en framträdande roll. Rekommendationerna i Parisdeklarationen är betydligt skarpare och kräver mer av såväl mottagarländer som givarländer. Den betonar kompetens- och effektivitet, målinriktning och samordning, betydligt mer än den svenska studien gör. Skillnaden tydliggörs genom citat av rekommendationer från vardera rapporten.

*10. Face up to and manage risks honestly, admit failures. Donor governments need to acknowledge frankly that development and development aid are inherently uncertain and risky and put in place measures to manage risks jointly with partners in the spirit of a mature partnership. In many donor countries, the period since the Paris Declaration has coincided with intense concerns about accountability for public spending that have at times translated into high levels of risk aversion. These tendencies have hampered good practice and many of the changes called for in the Declaration. But to try to avoid all risks in development cooperation is to risk irrelevance. There are demonstrated ways of promoting a mature public understanding of the uncertainties and risks of development and aid work and how to handle and learn from them, including both cutting edge initiatives and the effective use of tools like evaluation. Donors can learn from, and apply, some of these new techniques. There is also further evidence in this Evaluation to support the conclusions of other major assessments that the new approaches to development cooperation are in reality no more risky than traditional tightly donor-controlled projects, and that there are sound ways of managing the risks in the new models while also enhancing the development benefits.*

SADEV rapporten behandlar inte ovanstående fråga, men har en rekommendation (sid 50) som berör samma område: "The Government is recommended to advance the aid effectiveness agenda both internationally and at country level, and, recognising different contexts, operationalize the Paris Declaration and the Accra Agenda for Action."

Det är inte sannolikt att den internationella rapporten bygger på SADEVs rekommendation, snarare framstår den senare som en försiktig och begränsad tolkning av den förra. Den

internationella syntesrapporten är över lag mera specifik, mera direkt och även uttalad i sin analys och kritik av givarländernas genomförande av Parisdeklarationen. Den kritik som den internationella syntesstudien riktar mot givarländerna återspeglas inte i SADEV-rapportens analys och rekommendationer till den svenska biståndsförvaltningen. En parallell läsning av studierna har sina begränsningar, men det är inte möjligt att spåra några starka bidrag från den svenska studien till den internationella studien.

För det andra finns det potentiella användare i den svenska förvaltningen, och SADEV-rapporten riktar sina 19 rekommendationer antingen; (1) till regeringen, (2) till Regeringskansliet, UD och Sida, (3) till UD specifikt, eller (4) till Sida specifikt (sidan 50 och 51). De formella kraven på användbara rekommendationer är uppfyllda; de har en tydlig struktur, är samlade på ett och samma ställe, framgår av sammanfattningen, och är riktade till specifika aktörer. Innehållet i rekommendationerna är mer tveksamt. Några exempel:

- en rekommendation till regeringen är *"maintain reporting to Parliament on results of international development cooperation, with emphasis on challenges and opportunities to improve aid effectiveness"*. Jag är osäker på värdet av en sådan rekommendation och jag tror inte att regeringen övervägt att inte rapportera dessa frågor inom ramen för budgetprocessen, och gjorde den det skulle riksdagen sannolikt snart protestera.
- en rekommendation till regeringskansliet, UD och Sida är *"enhance and increase the use of skills of local staff"*, vilket är en självklarhet inom ramen för personalförsörjningen och organisationens arbete med kompetensutveckling och lärande. Det finns knappast någon som har en annan uppfattning.
- regeringen rekommenderas *"clarify the applicability of Paris/Accra in fragile and conflict/post conflict states .."* Det lär knappast vara regeringen som kan klargöra detta, utan snarare UD och Sida som besitter den kompetens som erfordras för ett sådant klargörande.
- Regeringskansliet, UD och Sida rekommenderas att *"maintain the dialogue between the Government Offices, MFA and Sida in line with the Joint Action Plan"*. Även där är det svårt att tänka sig att aktörerna skulle haft för avsikt att avbryta dialogen. Rekommendationen uttalar en självklarhet.
- En rekommendation till Regeringskansliet/UD lyder *"ensure that results frameworks are useful and easy to understand"*. Kanske borde rekommendationen snarare varit riktad till Sida, men den är under alla omständigheter en allvarlig förenkling och kan jämföras med den ovan citerade rekommendationen från den internationella rapporten som betonar att det inte finns några enkla lösningar, att problemen och biståndet är komplexa fenomen.

En närmare läsning av de 19 rekommendationerna visar ett de är överlappande och till stor del självklara och knappast tillför idéer och kunskap som inte redan finns i förvaltningen. Till del är de förenklingar och generaliseringar som skulle behöva mer konkretion för att vara användbara.

### Tillförlitlighet

Eftersom de 19 rekommendationerna samlas i ett kapitel på drygt en sidas text utan att de kopplas till analysen är det svårt att värdera hur de relaterar till slutsatser. Trovärdighet är ju en egenskap hos slutsatser snarare än hos rekommendationer. Slutsatserna ska bygga på utvärderingens datainsamling och analys, men rekommendationer bygger ofta på andra övervägande om t.ex. tillgång på resurser, alternativ användning av medel, mm. Det är dock en svaghet i den här studien att rekommendationerna är frikopplade från de 40 sidor text som föregår dem.

Hur tillförlitliga är då slutsatserna? Hela utvärderingen av Parisdeklaration, syntesrapporten liksom de olika landstudierna, organisationsstudierna och de tematiska studierna har genomgått en extern kvalitetskontroll. Den tar bl.a. upp *'Lack of standardized and defined criteria for rating judgements'* som en genomgående svaghet och det gäller även SADEV-rapporten.

SADEVs utvärdering bygger på intervjuer med ett femtiotal respondenter, huvudsakligen från Sida och UD. Den är starkt centrerad till den svenska delen av förvaltningen, enbart 10 intervjuer gjordes med respondenter vid ambassader. De var samtliga svenskar – inga lokalanställda. Det kastar tvivel på rekommendationen om kompetensutveckling av lokalanställd personal – var kommer den ifrån om inte en enda av de drygt 200 lokalanställda intervjuats? Allvarigare invändning är dock att ambassaderna spelar en nyckelroll i beslut om det bilaterala biståndet, och att bygga en så omfattande studie utan djupgående landbesök är tveksamt. Andra givare, t.ex. UNDP gjorde omfattande analyser av om fältkontoren arbetade i enlighet med Parisdeklarationen.

Intervjumallen är mycket omfattande. Bilaga 4 visar mallen som består av ett femtiotal frågor. Den är inte praktiskt användbar som intervjuguide. Enligt texten har SADEV valt att använda delar av den, men hur det gjorts och vad det innebär för behandlingen av svar framgår inte. Frågorna är generella och kan, var och en, leda till mindre avhandlingar. En fråga är t.ex. *'How has work to implement the Paris Declaration been implemented?'*. Det är vad hela utvärderingen handlar om, men man kan knappast begära att en intervju besvarar frågan. Intervjufrågor bör i regel vara specifika och konkreta. Några andra exempel på frågor som framstår som näst intill omöjliga att få tillförlitliga svar på är:

*'In what ways has the Paris declaration affected the structure of Sida's organisation?'*

*'What is the reporting on the Paris Declaration like, both externally and internally?'*

*'What do you think the other officers think of the Paris declaration?'*

Invändningen är inte att frågorna i sig är ointressanta, men för att belysa dem krävs ett betydligt mera fokuserat och konkret angreppssätt, och det är risk att sättet att fråga återspeglar att utvärderarna inte satt sig in tillräckligt i ämnet. Det är också tveksamt om det är bra att inleda en intervju med en fråga som *'What is the first thing you think about when you hear the Paris Declaration mentioned?'* Hur använder man svaret på en sådan fråga och hur påverkar den tillförlitligheten i vad som följer av intervjun?

Utvärderingen har även sänt ut en enkät till ett urval av 20 myndigheter som arbetar med bistånd. Det sägs ingenting om hur urvalet skett, men fem svarade att ämnet inte var relevant för dem. Man undrar då varför de ingick i urvalet. Tre andra svarade inte och utvärderingen bygger då i denna del på svar från 12 myndigheter. Svaren redovisas i bilaga 5, men framstår som otillförlitliga på grund av frågornas karaktär, sättet de ställs, och hur undersökningen är gjord. De refereras dock sällan i texten.

Utvärderingens slutsatser är att den svenska biståndsförvaltningen är motiverad att genomföra Parisdeklarationen (sid 21), inom förvaltningen har man kunskap om den, men utvärderingen konstaterar försiktigt att *'there is a certain competence gap at Sida between ideal and actual competence. The size of the gap varies considerably between individuals and teams'*. Tillförlitligt? Sannolikt ja, men det är en generell utsaga och närmast självklar i sin brist på konkretion. Slutligen konstateras att det inte finns några särskilda incitament att arbeta just med Paris deklarationen, däremot är personalen motiverad att göra ett gott jobb, och det är ju vad Paris deklarationen handlar om (!). Slutsatsen blir med andra ord *'The Swedish Government is both committed and ambitious regarding the Paris Declaration. Sida shares this commitment and ambition'*. På ett mycket generellt plan har detta uttalande trovärdighet, men det är också tydligt hur den internationella utvärderingen i sin analys är betydligt mer kritisk och pekar på utmaningar för såväl givare som mottagare av bistånd.

## Etik

Ett av kriterierna inom det här området gäller utvärderingens oberoende. Vad jag kan bedöma har utvärderarna varit oavhängiga från både UD och Sida och har kunnat genomföra sitt arbete utan påtryckningar och med god tillgång till information. De har haft möjlighet att resa, intervju, delta i möten och samla in data via enkät och dokumentanalys.

Utvärderingens intervjuer är genomförda under anonymitet. Intervjuprotokoll finns dokumenterade. Utvärderingen har intervjuat relativt många och det finns knappast någon möjlighet att spåra synpunkter i utvärderingen till enskilda uppgiftslämnare.

De etiska kriterierna innefattar även frågan om utvärderingen arbetat så att alla som borde komma till tals och få lämna synpunkter också har gjort det. Även om dryga 50 intervjuer är mycket, så är det en liten del av organisationen. Ambassader i flera viktiga biståndsländer är inte med, och det finns inga intervjuer med lokalanställd personal och heller avseende mottagarländernas synpunkter på hur Sverige lever upp till Parisdeklarationens avsikter. Det framstår som den allra viktigaste bristen. Utvärderingen borde ha byggt på fallstudier i landmiljö inkluderande omfattande intervjuer i samarbetslandets förvaltning.

En av de etiska kriterierna gäller hur de personer/organisationer som utvärderas kommer till tals med invändningar på rapporten. Utkast till rapporten sändes till UD och Sida och det kom en del synpunkter på saker att förtydliga, utveckla, mm. SADEV tog in del sådana synpunkter i slutrapporten, men inte allt. Det finns heller ingen bilaga med kommentarer från de utvärderade organisationerna (UD och Sida). När rapporten var klar gjordes en presentation på Sida och det är bra att utvärderarna är tillgängliga för diskussion efter genomfört uppdrag.

## Effektivitet

Det var ett frivilligt åtagande att delta i den internationella utvärderingen av Paris deklarationen. Genom att göra det demonstreras stöd för deklarationen och en vilja att arbeta i dess anda. SADEVs engagemang hade stöd från biståndsförvaltningen och var välkommet. Det fanns inga hinder av politisk eller administrativ art för att göra en gedigen utvärdering. Arbetet låg rätt i tiden och var synkroniserat med den internationella studien.

Utvärderingen av Parisdeklarationen pågick under tiden mars 2010 till januari 2011 och upptog 1,1 heltidstjänster – fördelat på tre utvärderare. Den totala kostnaden var 1 318 000 kronor. Utvärderingen hade med andra ord gott om resurser och med det i minnet framstår avgränsningarna vad gäller arbetsmetoder som onödiga. Det borde varit möjligt att genomföra ordentliga landbesök, att intervju samarbetspartner i mottagarländer, och även att inkludera de viktiga multi-bi projekten i utvärderingen.

Inledningsvis sägs att *'The evaluation is concerned with the evaluation criteria of effectiveness, i.e. the extent to which objectives were achieved.'* (sid 3). Därpå ges en analytisk modell som sägs vägleda utvärderingen där pilar leder från Parisdeklarationen till effektivt bistånd och *'reduced poverty'*. Rapporten säger klart och tydligt att det kommer att redovisas effekter. Därför blir läsaren förvånad när man på nästa sida läser att utvärderingen inte kommer att behandla effekter. Utvärderingen kan ju inte bedöma effektivitet om den inte behandlat effekter. Det är inte klart varifrån en sådan begränsning kommer. Givet utvärderingens resurser vore det möjligt att bedöma effekter och en sådan koppling skulle styrka slutsatserna om hur Parisdeklarationen genomförts.



## SADEV Report 2010:9. Gender equality in and through education

SADEV tog initiativet till att genomföra en serie utvärderingar kring jämställdhet eftersom det är en central komponent i biståndspolitiken. Detta är en av tre fristående utvärderingar på temat. Utvärderingen påbörjades hösten 2009 och blev klar i december 2010. Utvärderingen är genomförd av 3 personer från SADEV.

### Användbarhet

Användbarhet tas upp som det främsta syftet med utvärderingen; *"The evaluation aims at improving Swedish support to gender equality ..."* och *"The evaluation is primarily carried out for the use of the Ministry for Foreign Affairs, Sida, and partner countries"* (sid 2 och 3). Det är en hög ambitionsnivå och det är bra att användbarhet och användning uttrycks så klart och tydligt.

Utvärderingen behandlar tre frågor; (1) *"what factors are important in promoting gender equality in the education sector"*, (2) *"how does Sida use the dialogue with actors ..."*, samt (3) *"to what extent does Sida promote gender equality in the education sector through capacity building"*.

Jämställdhet inom utbildning och sektorns arbete för jämställdhet har behandlats i en lång rad forskningsrapporter, utvärderingar, strategi och policydokument. Det betyder i och för sig inte att ämnet är uttömt, men det ställer höga krav på innehållet i en ny publikation. Kapitel 3 i rapporten behandlar frågan mot bakgrund av litteratur på området. Kapitlet består av 10 sidor och är det längsta i rapporten. Det sammanfattas sedan i 11 punkter i slutsatskapitlet (sid. 29). Där sammanfattas t.ex. att *"poor school facilities are commonly reported and constitute an obstacle to learning achievements as well as to gender equality"* och *"Biases and stereotyping were identified as common problems"*. Det är knappast någon tvekan om att detta är rätt, men å andra sidan tillför det knappast något nytt för de handläggare med många års erfarenhet av fältarbete och biståndsadministration som är rapportens målgrupp – för att inte tala om företrädare i mottagarländer. För dem är detta och de nio andra slutsatserna förmodligen välkänt.

Appendix 2 i rapporten är mera konkret, där listas ett antal problemområden, och även förslag på lösningar (good practices). Å ena sidan är det en lovvärd ambition att ta de allmänna slutsatserna till en mera konkret nivå. Å andra sidan är det fortfarande långt kvar till en praktiskt användbar nivå. För att belysa med ett exempel sägs t.ex. att vad gäller problemet med den sociokulturella omgivningen är ett exempel på 'good practice' *"gender awareness training for parents and community members"*. Det finns säkert goda sådana exempel, men lika säkert många dåliga och bortkastade insatser med det syftet. Här hade referenser för vidare läsning kunnat vara användbara för att nå bortom det triviala. Den teoretiska delen av studien förs inte vidare i rapporten, man skulle annars kunna tänka att den skulle användas för att utveckla och testa kritiska hypoteser om både dialog och kapacitetsutveckling i handläggning av projekt och program, med jag har inte kunnat se någon sådan koppling.

Eftersom utvärderingen inte innehåller några rekommendationer beror användbarheten på vad den analytiska texten tillför och vad de ovan nämnda målgrupperna kan lära av den. Det finns två problem. För det första bygger utvärderingen på fallstudier i Tanzania och Kambodja. Utvärderingen erkänner att både dialog och kapacitetsutveckling beror mycket på omständigheterna i respektive land, men samtidigt görs inget försök att beskriva vad omständigheterna har för konsekvenser och hur det eventuellt påverkar generella iakttagelser. Det begränsar användbarheten. För det andra är fallstudierna formellt processinriktade och säger lite om dialogens innehåll. Utvärderarna refererar i några fall till dialogens innehåll och

problematiserar den (sid 25, sista paragrafen) men utan att gå på djupet. Det är dock genom en detaljerad analys av dialogens innehåll som utvärderingen skulle generera kunskap. Den beskrivning av dialog respektive kapacitetsutveckling som ges i kapitel 4 (4 sidor varje område) är alltför översiktlig och beskrivande för att vara praktiskt användbar.

Utvärderingen har valt att inte ge några rekommendationer och en utvärderings användbarhet bör inte bara ses i ljuset av dess rekommendationer. Utvärderingens första syftet går det knappast att utfärda rekommendationer för. Men när det gäller de två andra syftena, och med tanke på de omnämnda användarna, vore rekommendationer att föredra.

### Tillförlitlighet

Utvärderarna har identifierat väsentliga studier inom området jämställdhet och utbildning. Referenslistan tar upp cirka 25 titlar (exklusive material som land/samarbetsstrategier, metodreferenser, svensk policy, andra SADEV-studier). Bland de 25 titlarna återfinns några Sida-utvärderingar, en del utvärderingar från andra givare, rapporter från UNESCO och UNICEF, samt 12 titlar som kan betraktas som aktuell forskning på området. De är i regel publicerade mellan 2005 och 2010. Det reser tre frågor vad gäller tillförlitlighet: (1) Finns det viktiga referensarbeten som inte är omnämnda? (2) Skulle en inkludering av sådana ändra karaktären på problembeskrivningen och genomgången av 'good practice'? (3) Är studierna korrekt återgivna? Jag känner inte till några arbeten från senare år som borde varit med och kan därför inte se att slutsatserna kunde vara annorlunda. Vad jag kan bedöma ser det ut som om författarna har gjort en bra sammanfattning givet det mycket begränsade utrymmet. Det verkar vara en tillförlitlig sammanfattning av aktuell kunskap om problem och utmaningar vad gäller jämställdhet i utbildningssektorn.

När det gäller beskrivningen och slutsatserna om dialog respektive kapacitetsutveckling bör både reliabilitet och validitet diskuteras. Utvärderingens metodkapitel är översiktligt och säger ingenting om urval. Snarare är det så att de intervjuer som nämns (sid 4, näst sista paragrafen) täcker alla potentiella intressenter, men sedan är det bara en mindre del som återfinns i appendix 1 (intervjulistan).

Metodkapitlet anger att intervjuerna ska behandla "*the dialogue: what issues are discussed and why, forms for dialogue, and experienced challenges*" (sid 4) samt "*experiences from cooperation with Sida*" (sid 5). Detta är mycket svåra kvalitativa frågor och det är nödvändigt att se intervjumallar för att bedöma om intervjuerna kunde förväntas ge tillförlitlig information. Här bifogas dock inga intervjumallar. Det finns stor risk att sådana här intervjuer blir översiktliga och generella om utvärderarna inte är mycket väl insatta i problemområdet, förhandlingsprocessen i respektive land, och den politisk/administrative miljön. Det är risk att reliabiliteten är låg. Utvärderarna beskriver svårigheten med att t.ex. få ärliga synpunkter på Sida som dialogpartner (sid 5), men har ingen strategi för att lösa problemet.

Sidas arbete med att främja jämställdhet i utbildningssektorn beskrivs under rubrikerna "*Strategic documents*", "*The dialogue*" och "*Capacity building*". När det gäller strategisk dokumentation går utvärderingen igenom åtta samarbetsstrategier. Texten på en halv sida är övergripande och med summariska referenser. Så sägs t.ex. att "*Other country cooperation strategies mention girls' right to education, quality education with focus on girls etc. No strategy explicitly views the aim of support to education as a means to promote gender equality*". Det reser frågor kring hur bedömningen görs, vad exakt som ligger i "explicitly" och om det är giltigt även för andra än de bedömda strategierna. Det är en allvarlig kritik, men så som den framförs i utvärderingen har den både låg reliabilitet och validitet eftersom det inte framkommer hur bedömningen är gjord och inte heller hur representativa de valda fallen är. Vad gäller bedömningspromemorior konstaterar utvärderingen att tre sådana av åtta särskilt behandlar jämställdhet. Men det är också projekt som är särskilt inriktade på jämställdhet och i vilken

utsträckning andra behandlar frågan, eller borde behandla den, framgår inte. Det finns heller inga slutsatser av analysen.

Bedömningen av själva dialogen bygger på intervjuer med besök i två länder, samt telefonintervjuer med handläggare på de svenska ambassaderna i sex andra länder. De senare kan ge en del information om förekomsten av dialog, planer för dialogen, givarkoordination, mm. Det kan knappast ge insikter i det mera subtila förhandlingsspelet och heller inte andra aktörers synpunkt på dialogen. I förhållande till utvärderingens frågor ger ett sådant angreppssätt inte tillförlitlig och valid information – det hade behövts mera omfattande landbesök och kvalitativa intervjuer med flera aktörer.

Landstudierna kunde å andra sidan ge mer trovärdig information och är i och för sig ett klokt metodval. Landstudier ger möjlighet att gå på djupet med kvalitativ och delvis motsägelsefull information, men utvärderingen utnyttjar tyvärr inte metodens möjligheter. I Tanzania finns en struktur för sektorförhandlingar och det sägs att Sida hade en dialogplan med punkter om jämställdhet (sid 22, nederst). På sidan 24 sägs å andra sidan under rubriken "*Gender equality in the dialogue content*" att "*Nevertheless, no evidence of Sida being a key actor was found*". Eftersom Sida var ordförande för givarsamordning vid den tidpunkt då tre av sex punkter i dialogplanen behandlade jämställdhet undrar läsaren vad som hände – och om slutsatserna är rätt.

Liknande motsägelser dyker upp på flera ställen utan att de motstridiga slutsatserna behandlas och leds i bevis. I slutsatskapitlet sammanfattar utvärderingen "*to engage in the dialogue is a demanding task for staff at the Swedish embassies*". Men eftersom just detta är handläggarnas huvuduppgift framstår det som en underlig slutsats om dialogen med mottagarländer. Vad är alternativet? Slutsatsen om kapacitetsutveckling är också märklig. Sidan 31 sammanfattar "*very little is seen of Swedish direct support to capacity development in education ...*" och det sägs bero på kanalerna som väljs, arbete genom UNICEF i Kambodja och budgetstöd i Tanzania. Men genomgången på sidan 26 – 28 ger många exempel på kapacitetsutveckling och det är fullt möjligt att beskriva Sidas arbete i Kambodja som inriktat på detta, även om det sker genom UNICEF. På sidan 28 står "*UNICEF is working with capacity development on all levels*". Hur detta då leder till slutsatsen på sidan 31 är svårt att förstå.

När det gäller reliabilitet och validitet framstår det som att en annan utvärderare mycket väl hade kunnat dra andra slutsatser från samma material (inte minst eftersom utvärderarna drar olika slutsatser i olika kapitel). Med den valda metoden samlar utvärderingen in spridda synpunkter på den komplicerade förhandlingsprocessen som dialog med andra givare och mottagarlandet är, men det finns en tyngdpunkt på intervjuer vid de svenska ambassaderna och på formella strukturer som gör att utvärderingen inte tränger på djupet med insikter om hur Sida använder dialog, vad som behandlas och varför.

## Etik

När det gäller de etiska kvalitetskriterierna så finns inga invändningar mot utvärderingen. Utvärderarna har haft kontakter med Sida och UD och diskuterat hur arbetet framskridit. Man har haft samtal om urval och kontakter med ambassader och haft hjälp med att ordna landbesök. Det finns ingenting i rapporten eller mina intervjuer som antyder någon intressekonflikt. Det finns inte heller något material från intervjuer som framstår som känsligt, och även om det ibland är uppenbart vem i intervjuistan som framfört en viss synpunkt, så är det knappast ett problem. Under utvärderingens gång uppstod en schism mellan utvärderarna och SADEVs ledning, vilket ledde till att några av dem avslutade sitt arbete och inte ville framträda med namn på rapporten. Detta är en aspekt av processen som jag inte fördjupar mig i, men som sannolikt haft betydelse för kvaliteten på den slutprodukt jag läst.

## Effektivitet

Inom begreppet "*Feasibility*" ryms inte bara effektivitet utan även fråga om ämnesval, de politiska förutsättningarna för undersökningen, relevans, mm. Ämnet för utvärderingen är väl valt och kombinationen av utbildningssektorn och jämställdhet bör ha särskilt hög relevans. Det fokus på användbarhet som utvärderingen väljer borde också göra den särskilt relevant.

Det finns ingenting som gör ämnet svårt eller olämpligt att genomföra just 2010. Tvärtom, med regeringens prioritering av jämställdhet är det väl valt. Det finns många insatser att bygga utvärderingen på. Däremot kan de begränsningar som utvärderingen lägger på sig självt ifrågasättas. Det sägs uttryckligt (sid 3) att man inte utvärderar resultat. Det vore mycket lämpligt och i högsta grad intressant att utvärdera resultat. En del av de insatser som diskuteras har pågått länge. På några ställen i rapporten antyds mycket viktiga resultat, men de varken ifrågasätts eller bekräftas. En resultatorienterad utvärdering på temat skulle säkert ha varit mera relevant – och det borde vara med möjligt med de resurser utvärderingen haft. Den här utvärderingen hade en budget på 640 000 kronor och genomfördes mellan november 2009 – September 2011 med 1,5 heltidstjänster. Även här borde det ha kunnat genomföras betydligt mycket mera utvärderande arbete under den tidsperioden.

## SADEV Report 2011:1. Multi-bi-finansiering – utvärdering av beredning inför beslut

Det multilaterala utvecklingssamarbetet är en stor del av det svenska biståndet. SADEV utvärderade regeringens strategi för multilateralt utvecklingssamarbete år 2010, och under 2011 följde en utvärdering av multi-bi bistånd. Utvärderingen genomfördes av två utvärderare från SADEV under 2011.

### Användbarhet

Utvärderingen har en klar och tydlig struktur och är förhållandevis lättläst. Två kapitel behandlar multi-bi projekt i de två länder utvärderingen använder som fallstudier; Mozambique och Kambodja. Det är långa kapitel på cirka 20 sidor vardera, men de innehåller central information och beskrivningar av beslutsprocesser. Det är genom en nära och uppmärksam läsning av dessa kapitel som läsaren kan få en viss förståelse för beslutsprocesserna.

Utvärderingen har en sammanfattning på drygt 2 sidor. I punktform summeras iakttagelserna från de två landstudierna och därefter de sex rekommendationerna. Det är en användarvänlig sammanfattning. I fyra bilagor återfinns metodredogörelser och information om de två fallstudieländerna. En lista på intervjuade personer återfinns under referenser i slutet av texten. Med 80 sidor text är det en ovanligt lång SADEV-rapport och på gränsen att vara alltför lång<sup>7</sup>.

De viktigaste vad gäller användning är dock rekommendationerna respektive de lärdomar och insikter som texten kan ge. Vad gäller rekommendationerna är det blandat beträffande hur användbara de är. Rent allmänt är det bra att varje rekommendation åtföljs av cirka en halv sida text som dels anger vem respektive rekommendation är riktad till och dels utvecklar motivet och ger mera konkreta anvisningar för att verkställa rekommendationen.

Den första rekommendationen är att program från multilaterala organisationer i högre grad bör vara finansierade av organisationerna själva. Utvärderingen anger att UD och Sida bör verka för detta (sid 72). Detta har varit en svensk position i närmare 30 år, så det är knappast någon nyhet. UD har haft detta som en stående synpunkt i förhållande till de multilaterala organisationernas biståndsverksamhet, och har verkat för en större andel basbudgetstöd. Gensvaret har dock varit minst sagt ljummet och i praktiken har de flesta länder i stället ökat multi-bi finansieringen. Om en organisation som UNICEF skulle ha krav på en större självfinansiering skulle det minska omfattningen av deras arbete drastiskt. Rekommendationen framstår som verklighetsfrämmande.

Den sjätte rekommendationen säger att utlandsmyndigheterna bör skapa rutiner för mottagning av externa program- och projektförslag. Sida bör instruera dessa att slå fast ett datum när förslag ska vara inkomna och vilka formella krav som ska ställas på dessa. Det är ett förslag som skulle förenkla arbetet för utlandsmyndigheterna, men försvåra det för samarbetspartner. I dagens läge när det internationella systemet går mot gemensamma rutiner och metoder och där givarsamordning står högt på dagordningen vore det konstigt om Sverige skulle kräva en ensidig anpassning från de internationella organisationerna till sina tider och sina beredningsprocesser. Det är sannolikt inte möjligt eller lämpligt att driva en sådan linje.

---

<sup>7</sup> Många organisationer har formella gränser för hur långa utvärderingsrapporter får vara och gränsen sätts ofta till mellan 20 – 40 sidor text. SADEV har inga sådana begränsningar och det framstår för mig som riktigt. Det är inte den formella längden som avgör hur läsvärd och lärbär en text är.

Rekommendation nummer 2 slår fast att kvalitet bör betonas som överordnat mål. Rapporten säger att UD bör klargöra för utlandsmyndigheterna och Sida att kravet på kvalitetssäkring är överordnat beslut om landramar och särskilda satsningar. Detta är också en debatt som förekommit länge. RRV (1990) behandlade konflikten mellan "utbetalningsmål" och andra mål (varav de flesta har med kvalitet att göra). Det är bra att SADEV fäster uppmärksamheten på att konflikten finns kvar, men det krävs betydligt mer för att lösa problemet än att UD 'klargör' enligt ovan. I princip är det redan klargjort, och det faktum att problemen kvarstår visar att de är djupgående, svåra att åtgärda, och att det finns en mängd incitament och styrsignaler bortanför skenbart enkla klargörande.

De återstående tre rekommendationerna är mera användbara, även om det återstår arbete innan det är praktiskt genomförbart.

När det gäller användbarheten i analysen skulle jag tro att den grundliga studien av beredningsarbetet vid de här två ambassaderna kan vara nyttigt och bidra till lärande. Beskrivningarna av de 12 projekten visar på svårigheterna med manualer och på varierande beslutssituationer och samarbetspartner, samt inte minst på hur man hanterar risk i osäkra beslutssituationer.

### Tillförlitlighet

Rekommendationerna baseras på slutsatser och som textavsnitten visar är de riktade mot UD, Sida och utlandsmyndigheterna. De är med andra ord avsedda att appliceras på beredningen av multi-bi insatser i allmänhet. Under 2010 finansierade svenskt utvecklingsamarbete knappt 600 multi-bi insatser. En stor andel var globala, regionala och sub-regionala program. Det var interventioner i en mängd olika länder, inte minst många inom humanitärt bistånd. Sett under den tid SADEV rapporten analyserar har det finansierats två till tre gånger så många insatser, dvs. mellan 1200 och 1800. Hur tillförlitligt är det då att generalisera från 12 projekt i två relativt stabila länder till denna massa? Det är inte särskilt tillförlitligt, men om man försöker ändå bör läsaren informeras och man bör diskutera problemet med generaliseringen.

Det sker också en rörelse mot större program, mot regionala och globala insatser, så den typ av projekt/program som utvärderingen behandlar minskar och urvalet av studieobjekt är därmed inte så relevant som det kunde varit. Rapporten har ingen diskussion om urvalet och inte heller om problematiken kring att generalisera från slutsatserna. Den externa validiteten är låg och borde därför behandlats/beskrivits i rapporten.

Utvärderingen bygger en modell över beslutsprocesserna. Det är en bra ansats och det gör tydligt vilken typ av information som samlas in och hur den används. Den gör det också möjligt att se svagheter i modellen och hur den tillämpas, som t.ex.:

- Det är en mekanistisk modell som inte fångar de komplexa och situationsberoende processer som utgör verkligheten. Det är därför inte konstigt att bedömningsmallen sedan visar en stor skillnad på idealen som den tregradiga skalan ger uttryck för, och verkligheten.
- Indikatorer i mallen ger ibland felaktig information. Ett av bedömningskriterierna är t.ex. om utlandsmyndigheten under beredningen anlitar experter. Om så skett är det bra, annars inte. Problemet är ju dock att en sakhandläggare på utlandsmyndigheten eller på Sida förväntas vara expert. Anställningsannonser och arbetsbeskrivningar visar tydligt att det förväntas expertkunskap. Det är sannolikt bara undantagsvis extern expertis ska behöva anlitas, och det är knappast något fel om man inte behöver anlita extern expertis (fast om man behöver det ska man givetvis göra det). I alla fall ger indikatorn och värdet på den ingen valid kunskap om beredningen.

- En annan indikator gäller om andra organisationer bedömts som genomförare. Det kan vara relevant att göra en sådan bedömning ibland, men inte alltid. En av fallstudierna avsåg t.ex. multi-bi stöd till FN's reformarbete på landnivå (DaO Mozambique). Det är förstås omöjligt att använda någon annan kanal än just dessa organisationer för deras reformarbete, men enligt mallen bedöms då beredningen som av låg kvalitet eftersom man inte gjort någon sådan analys – eller så får man 'lura' sin modell och ange hög kvalitet även om arbetet inte gjorts. Det är orimligt och leder till låg validitet i hela undersökningen när värdena på indikatorerna aggregeras.
- Det är oklart hur skalan för bedömningen ser ut. Metodkapitlet anger en tregradig skala 'ej tillgodosett', 'delvis tillgodosett' och 'tillgodosett'. I texten används genomgående en skala med begrepp som svag, låg, acceptabel eller god, där det inte är helt klart hur de relaterar till den tregradiga skalan. Bilaga 2 anger hur begreppet 'tillgodosett' förhåller sig till 'acceptabelt', men där finns en blandning av de två begreppen som fortfarande blandar bedömningskriterier.
- Metodkapitlet betonar att de olika kriterierna/indikatorerna inte är viktade. Det är fel, alla har samma vikt – vilket är en helt annan sak än att de inte är viktade eftersom det ger alla variabler samma betydelse. Samtidigt kan man ifrågasätta om alla de värden som ingår är precis lika väsentliga i bedömningen – det kan jag inte avgöra här, men ett metodavsnitt borde behandla frågan.

Till detta kommer problemen med reliabilitet. De 35 intervjufrågorna i bilaga 2 är allmänna och ger stort utrymme åt bedömningar. Det kan knappast vara på annat sätt heller med den valda metoden, men det innebär att värdena i tabellen på sidan 2 har låg reliabilitet.

## Etik

Det finns inte några särskilda invändningar kring etik i utvärderingsprocessen. De formella överenskommelser som bör finnas på plats har funnits. SADEV skapade en referensgrupp för utvärderingen. Utkast till rapport har lämnats för synpunkter. Det verkar dock inte som om dessa beaktats i någon större utsträckning när slutrapporten har publicerats. Det finns inte heller någon redovisning av eventuellt avvikande uppfattningar eller kritik från berörda organisationer. Utvärderingen reser inte några frågor kring intervjuades integritet, ämnets karaktär är inte av den arten. Rapporten har redovisats på ett seminarium på Sida. Det finns inga intressekonflikter.

## Effektivitet

Utvärderingen genomfördes under tiden november 2010 till december 2011 och arbetstiden uppgick till 1,5 heltidstjänster. Den sammanlagda kostnaden var 1 377 000 kronor. Varje landbesök upptar cirka 20 intervjuer, så det är rimligt att anta att varje sådant besök uppgått till 2 veckor. Utvärderarna har genomfört sex intervjuer på UD och åtta på Sida. Referenslistan tar upp dokument för vart och ett av de 12 projekten som analyserats och även annan bakgrundsinformation. Att skriva en rapport på närmare 100 sidor tar också en del tid. Det är dock svårt att se att så mycket mer än 3 – 4 månaders heltidsarbete ska behöva gå åt, och hade utvärderingen upphandlats hade det varit rimligt att specificera så lång tid för arbetet. Effektiviteten får bedömas som förhållandevis låg.

Utvärderingens mål och ämne är i hög grad relevanta. Metodval och urval av objekt att studera begränsar dock relevansen väsentligt. Det finns betydligt mera angelägna beslutsprocesser gällande multi-bi projekt att granska än landbaserade projekt av den karaktär utvärderarna valt. Kvalitativa beslutsprocesser är svåra att bedöma, och om utvärderarnas kunskap inom området

är begränsad behövs mycket mera tid för fältstudierna. En ofta använd tumregel<sup>8</sup> är att 50 – 60% av en utvärderings tid och resurser bör ägnas empirisk datainsamling, så den här utvärderingens tidsallokering framstår inte som optimal.

Utvärderingen ligger rätt i tiden och det finns ingenting som tyder på att den kommer för sent eller för tidigt för att vara användbar. Det var också lämpligt att följa upp Multistrategi utvärderingen med en utvärdering inriktad på beslutsprocesser. Utvärderarna behandlade dock bara processerna och tog inte tillfället att även analysera resultat. Det vore annars ett bra komplement till processanalysen och kunde kopplat processen till dess effekter. Många av de studerade projekten var utvärderade tidigare och kunde enkelt kompletteras med verifierande slutomdöme. Bristen på resultatanalys framstår som ett förbiseende – 'an opportunity lost'.

---

<sup>8</sup> se t.ex. uppdragsbeskrivningar för utvärderingar inom utvecklingsarbete och för en generell diskussion även Forss (2009)



## SADEV Report 2012:1. Demokratisk utveckling och ökad respekt för mänskliga rättigheter – resultat av svenskt bistånd.

SADEV fick i december 2009 uppdrag av regeringen att utvärdera resultat av det bistånd vars huvudinriktning är demokratisk utveckling och ökad respekt för mänskliga rättigheter. Utvärderingen skulle vara klar den 15/2 2012. Flera personer har varit engagerade i studien, förordet nämner projektledare och tre andra utvärderare på SADEV. Avrapporteringen av utvärderingen består av en rapport (SADEV 2012:1) som sammanfattar analys och slutsatser av hela utvärderingsuppdraget, och fem delstudier (SADEV 2012: 2 – 6). Den här analysen fokuserar på syntesrapporten (2012:1) för bedömningen av användbarhet, trovärdighet, etik och effektivitet, men gör referenser till de andra för bedömningen av tillförlitlighet, etik och effektivitet.

### Effektivitet

Biståndet för demokrati och mänskliga rättigheter uppgår till ungefär en tredjedel av det bilaterala biståndet, och under den tid som SADEV utvärderar (1998 – 2010) finansierades 13.850 insatser till ett belopp av 36,2 miljarder kronor. Det är utan tvekan ett mycket svårt utvärderingsuppdrag. Det kan inte råda någon tvekan om att det är relevant och intressant i sig, men frågan är om det uppfyller andra krav på 'feasibility'. När SADEV fick uppdraget hade organisationen funnits i tre år och det var tre turbulenta år med byte av chefer och hög personalomsättning. En del av de kvalitetsproblem som påpekas i de utvärderingar jag studerat fanns förmodligen sedan tidigare. Det framstår som tveksamt att lägga ut ett så komplext utvärderingsuppdrag i det stadium av kapacitetsuppbyggnad organisationen då befann sig.

Utvärderingen inleder med en tolkning av uppdraget (sid 8) som leder till en precisering av frågor och som ger ett fokus i förhållande till uppdraget. Det anger att måluppfyllelse och hållbarhet, samt relevans på två nivåer, samt en bedömning av '*direkta och indirekta, avsedda och oavsedda, positiva och negativa*' effekter (sid 8). Det är en bra början, men i det här fallet hade en mera precis formulering av utvärderingsfrågor varit önskvärd.

Utvärderingen genomfördes under tiden januari 2010 till februari 2012 och ianspråk tog 4 heltidstjänster. Kostnaden för hela utvärderingen uppgick till 4 053 000 kronor. Det är en väl tilltagen budget och det går att åstadkomma mycket med de resurserna. Det går dock knappast att besvara en allmän fråga om resultat av svenskt bistånd inom DMR när detta består av 13 850 insatser, spridda på 100 länder, och av mycket olika karaktär. Frågan om hur man begränsar uppdraget och får tillförlitliga svar på frågor om resultat utifrån den budget man har borde ha utvecklats längre och kommit till uttryck i syntesrapporten. Kanske borde det ha förmedlats mera precisa resultatfrågor till SADEV i samband med att uppdraget utformades.

Utvärderingsteamet har hunnit med mycket under den tid arbetet pågått. De har genomfört fem landstudier som rapporterats var för sig, och varje rapport bygger på en analys av utvecklingen vad gäller DMR i landet, en analys av det svenska biståndet och hur DMR passar in i landstrategierna, samt utvärderingar av ett antal projekt/program i respektive land. Utvärderingen har gjort en översyn av forskning på området, gjort syntesstudier, och även policyanalyser samt ett antal intervjuer på UD och Sida. Rent oavsett de iakttagelser vad gäller kvalitet som framförs så har den s.k. inre effektiviteten i arbetet varit hög.

SADEV erhöll uppdraget i december 2009, men det arbete som redovisas i rapporten 2012:1 och de bakomliggande fem delstudierna har genomförts sedan våren 2011, dvs. under mindre än ett

år. Det ingår inte i den här studien att undersöka varför uppdraget inte kom igång under 2010, och heller inte vad som eventuellt skett under den tiden. Men det faktum att det utvärderingsarbete som redovisas i förhållande till uppdraget pågått under knappt ett år, snarare än under drygt två år, som regeringsuppdraget förutsedde, kan ha påverkat en del av de kvalitetsproblem som diskuteras nedan.

## Användbarhet

Rapporten 2012:1 som sammanfattar uppdraget består av drygt 60 sidor text. Det finns ingen sammanfattning och det finns heller ingen läsanvisning. Rapporten är skriven på svenska. När en rapport är mycket kort och sammanfattande kan det ibland vara onödigt med en sammanfattning, men i det här fallet skulle förståelsen av materien underlättats med en bra sammanfattning.

Kapitel 5 har rubriken 'Slutsatser och rekommendationer'. Det är bra att läsaren får dessa moment tydligt påpekade, men det är inte att förväxla med begreppet 'sammanfattning'. Kapitlet går först igenom de viktigaste slutsatserna och presenterar därefter 10 rekommendationer. Varje rekommendation är uttryckt som en uppmaning, följt av en eller två paragrafer motiverande text. Det är en bra struktur för presentationen. Det hade varit lämpligt att presentera en slutsats i direkt anknytning till rekommendationen, så hade kapitlet som helhet framstått som mindre splittrat och en del överlappande text kunde skurits bort. Som ett exempel citeras en del av texten om vad som bidragit till goda resultat och motsvarande rekommendation några sidor senare (sid 56 och 58):

### ***Slutsats: "Sverige har skicklig personal på utlandsmyndigheterna***

*Tjänstemän på utlandsmyndigheterna påverkar viktiga aktörer i stationeringslandet och skickliga ambassadörer har haft avgörande inflytande på DMR-området. I samarbetsländerna har Sverige ofta haft och har rollen som givarsamordnare".*

### ***Rekommendation: "Satsa på kunnig och välunderrättad personal på utlandsmyndigheterna***

*Tekniskt kompetent och politiskt välunderrättad personal på utlandsmyndigheterna är nödvändiga för att få resultat. De som arbetar med DMR-bistånd behöver vara insatta också i politiska frågor. Att ge personalen möjlighet att lära av varandra och dela med sig av erfarenheter är ett viktigt incitament i sammanhanget. Dialog är inte förbehållen ambassadören - även tjänstemännen måste vara förberedda för att i anslutning till DMR-stöd ta upp angränsande politiska frågor. I tillägg till DMR-kunskap bör god kommunikationsförmåga prioriteras när personal utbildas inför uppdrag utanför Sverige. Kommunikationsstrategier bör tas fram på utlandsmyndigheter där dialogen är central."*

Är det en användbar slutsats och rekommendation? Det är alltid svårt för en utomstående att veta, men för mig framstår det som att det är en nyttig påminnelse att det personliga engagemanget och kunskaperna vad gäller ambassadörer och tjänstemän, och samverkan dem emellan, kunde vara en underskattad faktor. Det är bra att bli påmind om den och det kan vara användbart att betona detta. Jag kan inte påminna mig att det lyfts fram lika tydligt i andra utvärderingar på området. Detsamma gäller slutsatsen innan, som handlar om vikten av närvaro och dialog, och som utmynnar i en likande rekommendation. Dessa var rekommendationer 1 och 2, och rekommendation nummer tre (sid 58) är: 'Öka kunskapen om samsambetsinstitutionerna'. Det kan vara värt notera att Regeringens policy för DMR säger att: 'Vidare anges att Sverige ska ha kunskap om varje samsambetslands politiska situation, specifika historia, traditioner, maktstrukturer, värderingar och kultur för att åstadkomma bästa möjliga resultat i utvecklingssamarbetet.' så det utvärderingen säger bekräftar en redan beslutad inriktning.

Rekommendation 4, 5, 6 och 7 handlar om att bättre planera insatserna genom målformulering, riskanalys, partnerskapsavtal och genom att utveckla det lokala ansvaret. Rekommendation 8 fäster uppmärksamheten på jämställdhetsarbete, och rekommendation 9 och 10 handlar om utvärdering – systemen behöver utvecklas och det behövs fler utvärderingar av långsiktiga effekter.

I några fall kan användbarheten av rekommendationen ifrågasättas. Det är mycket vanligt att utvärderingar, särskilt tematiska utvärderingar och syntesstudier, slutar i att det behövs mer utvärdering på området. Frågan vem som ska göra dem, hur de ska bli bättre, mm. besvaras inte. Flera av de ovan nämnda rekommendationerna framstår som väldigt generella och ganska vanligt förekommande i utvärderingar av utvecklingssamarbete. Det hade behövts mer detaljerade anvisningar för att de ska vara användbara.

På en mera abstrakt nivå indikerar rekommendationerna ett utvecklingssamarbete av traditionell bilateral karaktär, projekt med en mottagande organisation som använder medel för ett syfte, som kommer till ett slut och som efter projektet har uppnått 'hållbarhet'. Det innebär att det moderna biståndet där FN organisationer är centrala aktörer och ofta samordnar internationella insatser, i Parisdeklarationen anda av givarsamordning, ömsesidigt resultatansvar, mm. inte är reflekterat i rekommendationerna (se t.ex. sid 50). Det gör sannolikt rekommendationerna mindre användbara.

Användbarheten av en så omfattande studie som denna ligger inte bara i rekommendationerna. Det skulle vara ett mycket högt pris att betala 4 miljoner kronor för de 10 rekommendationer som finns på sidorna 58 – 60. Analysen i sig bör vara användbar, och de landstudier som presenteras av DMR bistånd i Guatemala, Kenya, Serbien, Vietnam och Kambodja borde bidra med lärande. I rapporten presenteras de med 1– 5 sidor text vardera. I varje land gjordes ett urval av några projekt, men läsaren får inte klart för sig hur stor populationen var och hur den relaterade till annat svenskt bistånd och till bistånd från andra länder och organisationer. Syntesrapporten för inte någon diskussionen om hur situationen vad gäller demokrati och mänskliga rättigheter utvecklats i landet och kan följaktligen inte heller analysera de av Sverige finansierade insatserna och därmed blir det ingen analys av resultatkedjan från insats till effekt. De landstudier som publiceras som separata rapporter har mera fullständiga analyser, men så som de sammanfattas ger de inte läsaren rätt bild av vad som skildrats i dem och väcker därmed inte nyfikenhet att läsa dem. Det är synd, för några av dem innehåller mycket bra studier av DMR biståndet, med en analys av det svenska biståndets effekter och hur det, tillsammans med andra aktörer, bidragit till situationen vad gäller DMR i landet, även när denna försämrats. Särskilt landstudien av Guatemala bidrar med förståelse för DMR biståndets möjligheter och begränsningar.

Kapitel 3 är den mest omfattande delen av syntesstudien med 27 sidor text. Det är en viktig redogörelse för strukturen på det svenska DMR biståndet och där redovisas erfarenheter från andra utvärderingar. Strukturen på analysen utgör dock ett problem. Det finns åtminstone tre principiellt olika sätt att dela upp och analysera portföljen av insatser. Utvärderarna kan; (1) använda kategorier som återkommer i policy dokument, (2) använda kategorier i enlighet med Sidas databas och OECD/DAC kriterier, eller (3) konstruera egna kategorier. SADEV valde det senare och det innebär att slutsatser som dras måste omtolkas till det ena eller andra kategoriseringssystemet och även bedömas hur relevanta de är i förhållande till det. Det framgår inte att de nykonstruerade kategorierna är begreppsmässigt klarare eller leder till färre begränsningar än alternativen, och det är ett val som försvårar användbarheten. Om man väljer att utveckla egna kategorier bör de vara väsentligt bättre än de som används redan, annars försvårar de användningen av analys och slutsatser.

## Tillförlitlighet

Metodavsnittet i rapport 2012:1 är på en och en halv sida. Det innehåller relevant information, särskilt om urvalskriterier för landstudierna, men halva avsnittet handlar om strukturen på rapporten och inte om metoder. En del av avgränsningsproblematiken har dock behandlats i tolkning av uppdraget, men det är en fråga som har metodkonsekvenser och även bör diskuteras i ett metodologiskt perspektiv. Det är många frågor avseende metod och metodologi som inte diskuteras, vare sig i rapporten eller i bilageform, t.ex.;

- val av ansats, dvs. hur utvärderingen avser belysa kausala samband mellan insatserna och resultat;
- hur de olika momenten av meta-utvärdering, statistisk undersökning och landstudier är avsedda att komplettera varandra och varför just denna kombination av datainsamling är användbar;
- frågan om olika typer av resultat och hur utvärderingen hanterar dem;
- inledningen säger att *'Resultaten av biståndet redovisas i termer av prestationer och effekter (direkta och indirekta, avsedda och oavsedda, positiva och negativa)'*, men metodavsnittet säger ingen ting om hur dessa aspekter undersöks och det syns heller ingenting sådant i rapporten;

Med tanke på att uppdraget är så komplext hade ett fylligt metodkapitel varit påkallat för att etablera rapportens trovärdighet.

Kapitel 2 presenterar det svenska biståndet för demokrati och mänskliga rättigheter. Där konstateras att *'under perioden 1998 – 2010 finansierade Sida nästan 13 850 DMR-insatser till ett belopp om 36,2 miljarder SEK'*. Det är en viktig uppgift, men då inräknas enbart insatser som har DMR som huvudinriktning. En insats klassificeras beroende på om ett målområde är 'principal objective', 'significant objective' eller 'not relevant'. En insats skulle t.ex. kunna ha jämställdhet som 'principal objective' och demokrati och mänskliga rättigheter som 'significant objective', men skulle i så fall inte finnas med i SADEVs redovisning. Det är ett allvarligt problem och det innebär att bilden av hur mycket bistånd som ges till området inte är tillförlitlig. Det är en begränsning som gjordes i uppdraget och det är i och för sig inte fel, däremot måste man diskutera vad det har för konsekvenser och vad det blir för bild av helheten man presenterar.

Det leder också till andra problem vad gäller slutsatserna i studierna. En av slutsatserna är att det är ett svagt genomslag för jämställdhet (sid 57). Men en insats för t.ex. nätverksbyggande mellan kvinnliga parlamentariker kan i första hand klassificeras som jämställdhet (principal objective) och i andra hand som DMR (significant objective). Men då kommer den inte med i utvärdering eftersom de bara behandlar insatser som har DMR som 'principal objective'. Kanske är jämställdhetsområde särskilt svårdefinierat eftersom insatserna ofta berör både jämställdhet och ett annat målområde. Det innebär att slutsatserna om jämställdhet på sid 31 – 34 samt 41 och 42 inte är tillförlitliga.

På sidan 13 presenteras också en övergripande slutsats att *'Största mottagare av DMR-stöd är multilaterala organisationer, i huvudsak inom FN-familjen.'* Jag antar att detta avser multi-bi projekten, dvs. insatser som finansieras av det bilaterala utvecklingssamarbetet med där någon organisation som Världsbanken, UNDP, UNICEF eller annan varit ansvarig för program/projekt. Men de är inte mottagare, det är samarbetsländer eller regioner som är mottagare. De multilaterala organisationerna är kanaler för utvecklingssamarbete – inte mottagare.

Slutligen skulle jag vilja problematisera slutsatsen om kunnig och kompetent personal och ambassadörernas roll som nämndes ovan. Landstudierna, särskilt den om Guatemala, ger goda belägg för slutsatsen, men det är inte lika klart vad ambassadörernas roll varit i de andra länderna. Det framgår inte av kapitel 4 i syntesrapporten. Steg 1 i analysen är då att i de länder

där processanalys och resultat har påvisat ett positivt samband, där är slutsatsen trovärdig. Steg 2 är att det är en rimlig slutsats och den stämmer med sunt förnuft. Steg 3 i SADEV rapporten är att framföra rekommendationen att satsa på kunnig personlig och skickliga ambassadörer. Som nämnt har det ett värde att uppmärksamma detta. Frågan är dock hur giltigt det är för de 100 länderna där DMR bistånd genomförts. Är det uteslutet att resultaten är goda ändå och att de kan bero på en rad andra faktorer; t.ex. en hängiven Resident Coordinator, eller karismatiska ledare i jämställdhetsrörelsen eller andra organisationer i det civila samhället? Innebär det för framtiden att DMR bistånd inte bör genomföras i länder där det inte finns en ambassad, eller vilka blir begränsningarna? Det är en analys och en diskussion som måste föras betydligt längre än vad som görs i rapport 2012:1. Även om slutsatsen verkar stämma med sunt förnuft så finns det en komplex verklighet där det behövs fördjupade resonemang för att nå trovärdiga – och användbara – slutsatser.

## Etik

Det finns inte några särskilda invändningar kring etik i datainsamling och analys när det gäller rapporten 2012:1. De problem av etisk karaktär som framstår som särskilt viktiga skulle kunna gälla uppgiftslämnarens integritet och behov av skydd om de uttalar sig kritiskt om MR frågor i miljöer där de kunde kännas igen och råka ut för repressalier. Det förekommer dock inga intervjuer som underlag i syntesrapporten och i landrapporterna framgår att de intervjuade finns i ett organisatoriskt sammanhang som t.ex. företrädare för ombudsmannamyndigheter, rättsväsendet och organisationer i det civila samhället. Intervjulistorna i fallstudierapporterna indikerar att centrala aktörer kommit till tals; både sådana som varit mottagare av bistånd och andra.

De formella överenskommelser som bör finnas på plats har funnits. SADEV skapade en referensgrupp. Utkast till rapport har lämnats för synpunkter. Det finns visserligen ingen redovisning av eventuellt avvikande uppfattningar eller kritik från berörda organisationer, men om det hanterats via remissprocessen är det i så fall bra.

## Slutsatser och rekommendationer

### Allmänt om kvalitet

Begreppet kvalitet är mångfacetterat. Olika aspekter av kvalitet kan peka åt skilda håll. Vissa kvalitetsegenskaper är bra, andra mindre bra. I texten ovan presenteras omdöme om var och en av de fem utvärderingar som granskats, men som läsaren lagt märke till är dessa omdöme nyanserade snarare än förenklade. Vissa egenskaper är ganska bra, andra tvärtom. I princip kan egenskaperna adderas ihop till ett enda slutomdöme, och hade det varit fråga om t.ex. att godkänna en avhandling, ge ut en publikation på ett förlag, så hade beslutssituationen tvingat fram ett omdöme av karaktären 'ja' eller 'nej'. Den här texten har inget behov av en sådan förenkling och jag tror att både de svaga och de starka sidorna hos SADEVs utvärderingar blir tydligare på det sättet.

Sammanfattningsvis finns det *ett antal goda egenskaper* hos de fem SADEV-studierna som har granskats:

- Ämnena är väl valda och i högsta grad relevanta i biståndspolitiken.
- Utvärderingarna har genomförts i tid och levererats enligt beställning och nått potentiella användare så att de kunnat användas i beslutsprocesser.
- Utvärderingsfrågorna är i regel preciserade och väl formulerade för den fortsatta utvärderingsprocessen; dvs. de är fokuserade, relevanta och möjliga att utvärdera, samtidigt som de är så pass öppna att de ger möjlighet att utveckla processen efter resultat som uppnås under väg.
- Det finns överenskommelser med de organisationer som berörs och processer för samverkan, t.ex. för inledande diskussioner, datainsamling, preliminära rapporter, och presentation av slutrapporter.
- Det finns interna system för kvalitetskontroll, med seminarier och referensgrupper.
- Utvärderingarna har haft ordentliga resurser både ifråga om tid och pengar för att besvara utvärderingsfrågorna. Alla har disponerat minst en årsarbetskraft, någon betydligt mer.
- Utvärderingarna har en stark oberoende ställning och det har inte framkommit några intressekonflikter som kunnat påverka dess omdömen.

Det finns alltså goda förutsättningar och det är viktigt att bygga på de här egenskaperna. Men genomgången av de fem utvärderingarna har även pekat på *ett antal brister* som förekommer i större eller mindre grad hos dem alla:

- Användbarheten av slutsatser och rekommendationer är låg, till stor del beroende på att de rekommendationer som framförs är allmänna, saknar konkretion, och inte är tillräckligt praktiska för att ge vägledning i beslutssituationer.
- Bidragen till organisatoriskt lärande är låga eftersom de kunskaper som förmedlas är väl kända för de flesta och sällan innebär några nyheter.
- Rapporterna har omfattande metodologiska brister vad gäller upplägg, dvs. hur man avser att utvärdera kausala samband mellan insatser och effekter.
- De tekniska metodfrågorna vad gäller urval, utformning av instrument för datainsamling, utveckling av skalor för att mäta och bedöma det som utvärderas har varit bristfälliga.
- Transparensen är låg och metodkapitel saknar ofta grundläggande information för att läsaren ska kunna bedöma trovärdighet.

- Det är bara en av de fem utvärderingarna som kommer i närheten av att bedöma resultat, men det borde varit en självklar del i åtminstone tre av de andra utvärderingarna.
- Kopplingen till forskning på områdena som utvärderas respektive till andra utvärderingar är i en del fall svag och i den mån den finns så används den inte tillräckligt mycket för att generera och pröva hypoteser i den pågående utvärderingen.
- Resurserna i tid och pengar används inte effektivt. Utvärderingarna har lagt alldeles för lite tid på empiriskt datainsamling i allmänhet och för lite tid i fält i synnerhet.
- Dialogen med mottagare/användare har inte genomförts till slut och de synpunkter och/eller invändningar som framkommit reflekteras inte i rapporterna.

Nedan följer en genomgång i förhållande till var och en av de fyra kvalitetskriterierna.

### Användbarhet

De fem SADEV studier som analyserats här har alla haft en hög ambitionsnivå när det gäller användning. Det har varit ett uttalat syfte att bidra till biståndsförvaltningens lärande, att lyfta fram exempel och ge rekommendationer. De motiveras av en önskan att vara användbara. Alla utom en av rapporterna har sammanfattningar som är läsarvänliga. Alla utom en har klara och tydliga rekommendationer, ofta i särskilda kapitel. Tre av fem anger också vilka aktörer rekommendationerna riktar sig till.

SADEV har försökt arbeta nära UD och Sida. Kontakter har inletts tidigt i utvärderingsprocessen, men har dock inte lett till att potentiella användares intresse styr utformningen av arbetet – inte ens där så vore befogat utifrån ett användarperspektiv.

Rapporterna har en struktur som ger översiktlighet och som bidrar till att läsare får en överblick och uppfattar en röd tråd i framställningen. Rapporterna är som regel korta. I några fall är språket otympligt. Två rapporter är skrivna på engelska och tre på svenska. Genom att skriva på svenska når man färre potentiella användare – en betydande andel av de anställda på ambassaderna talar inte svenska och det gör inte partner i samarbetsländerna heller. Valet av svenska som huvudspråk drar ned användbarheten väsentligt.

Varje rapport – med ett undantag – har rekommendationer. De är av olika karaktär, men en stor del av dem framstår som;

- svagt underbyggda i den föregående analysen;
- självklara genom att det som föreslås redan äger rum och är beslut/processer som föreligger oavsett vad SADEV-rapporterna säger;
- saknar konkretion vilket gör att det är långt från den allmänna rekommendationen till praktisk användning;
- direkt felaktiga, eller bygger på felaktiga föreställningar om utvecklingssamarbetet och dess aktörer.

Därmed uppnår rapporterna inte syftet att vara användbara, och det beror på det substantiella innehållet i rekommendationerna. Strukturer och processer är på plats och har bra utformning, men har i de här fem fallen inte fyllts med ett konkret, kritiskt provande, nyttigt och användbart innehåll.

## Tillförlitlighet

Utvärdering bygger på vetenskaplig metod och det är genom rätt metodval och en korrekt tillämpning av metoder som en utvärdering kan ge tillförlitliga svar på utvärderingsfrågor. De fem SADEV-rapporter som analyserats har mer eller mindre allvarliga brister vad gäller metod.

De saknar en övergripande struktur som väver samman utvärderingens syfte och frågor med val av upplägg, grunder för värdeomdöme och slutsatser, och de attribut som kännetecknar det utvärderade fenomenet. En av rapporterna säger sig redovisa ett sådant upplägg (programteori), men problemet är att det faktiskt inte är en programteori som redovisas och används. De fyra andra har ingen diskussion alls av upplägg och det går inte heller att utläsa en sådan implicit. De saknar därmed övergripande vetenskaplig styrning.

De områden som utvärderas är oftast omfattande och den empiriska datainsamlingen bygger på urval – av länder att besöka, projekt/program att analysera, människor att intervjua, dokument att analysera, respondenter för enkäter, mm. De urval som gjorts i de här utvärderingarna har i varierande grad varit;

- alltför små;
- inte beskrivna i metodkapitel eller text för övrigt;
- inte gjorda enligt explicita kriterier, och
- missvisande i förhållande till det utvärderade fenomenet.

I tre fall utvecklas modeller som används för att strukturera komplexa situationer, t.ex. beredningsprocesser för multi-bi insatser och för genomförande av Parisdeklarationen. Det är bra att bygga modeller och det borde vara mer vanligt förekommande även i de andra rapporterna. Ett problem är att variablerna i modellerna inte är valida. I ett fall var t.ex. den interna validiteten låg eftersom utvärderingen byggde på felaktiga föreställningar om valmöjligheter vad gäller kanaler att arbeta genom, kompetens och kunnande i beredningsprocesser, mm.

Utvärdering handlar om att bedöma och det är viktigt att grunderna för bedömning är tydliga. I en del fall används mått och skalor. Ett problem som påvisades här var att skalorna inte var riktigt konstruerade, dvs. de saknade en enhetlig begreppsapparat. Då blir tillförlitligheten lidande. Ett liknande problem var att intervjumallar var alltför omfattande och frågorna alltför svävande, vilket i sin tur leder till låg validitet och låg reliabilitet.

Tillförlitligheten i de fem studerade rapporterna är låg på grund av låg reliabilitet, låg intern och extern validitet, och genom att det saknas genomtänkta upplägg för utvärderingarna. Det i sin tur bidrar till att användbarheten är låg. Det är särskilt viktigt att läsare kan bedöma tillförlitligheten själv genom metodkapitel. Dessa bör redovisa svårigheter och problem som kan påverka bedömningen. Rapporterna har metodkapitel och de ger insikter i arbetsprocesserna, men de saknar genomgående analyser av urvalsförfarande, kriterier för olika metodval, och diskussioner av konsekvenser vad gäller tillförlitlighet. Det får läsaren göra själv och får ingen hjälp på vägen.

## Etik

De fem utvärderingar som granskats har haft en stark oberoende ställning och det finns ingenting i rapporterna som tyder på att deras oberoende komprometterats eller att det för övrigt funnits några intressekonflikter som påverkat analys, omdöme och rekommendationer. Utvärderarna har inlett kontakter med de utvärderade organisationerna tidigt och presenterat uppdrag och syfte. De har presenterat preliminära rapporter och gett möjligheter till



återkoppling. Däremot har kritiska synpunkter inte beaktats och i ringa grad lett till förändringar i rapporterna.

En av de standards som utvecklats för området etik gäller att utvärderarna ska ha erforderliga insikter och kunskaper på det utvärderade området och att utvärderingen inte ska leda till ökad arbetsbelastning för den utvärderade organisationen. De intervjuer som genomförts i den här studien pekar på att några av utvärderingarna inte tillförst den kunskap om utvecklingssamarbete som är en nödvändig förutsättning för att genomföra analys, dra slutsatser och utveckla användbara rekommendationer.

Det har inte framkommit några problem vad gäller uppgiftslämnarens integritet; utvärderingarna har inte samlat in känslig information, och i den mån sådan eventuellt skulle finnas kommer den inte till uttryck i rapporterna.

I några fall har utvärderingarna dragit slutsatser och lämnat rekommendationer utan att berörda personer intervjuats eller haft möjlighet att lämna synpunkter. Det gäller t.ex. lokalanställd personal vid utlandsmyndigheterna, deras kompetens, roller, och kapacitetsutveckling. Det är ett etiskt problem av underlåtenhetskaraktär. En utvärdering ska interagera med berörda organisationer och personer. På samma sätt kan det framstå som ett etiskt problem att så lite av utvärderingarna bygger på förhållanden i de länder där utvecklingssamarbete äger rum och att så få i förvaltningar, frivilligorganisationer, forskarsamfund, m.fl. kommer till tals i utvärderingarna.

## Effektivitet

Den grupp av standards som behandlas under den här beteckningen är mera omfattande än effektivitet, även om detta är en central del av bedömningen. Andra delar avser relevans och 'evalueringsbarhet'. Ämnesområde för de fem utvärderingarna är väl valda och återspeglar centrala diskussioner i politiken för utvecklingssamarbete. De har hög relevans både för den politiska ledningen, för biståndsförvaltningen, och för en bredare biståndsintresserad allmänhet.

De ämnen som valts har omvandlats till utvärderingsfrågor. De frågor som styrts respektive utvärdering har för det mesta varit relativt fokuserade och klara. Frågorna har även varit relativt öppna, vilket i princip gjort det möjligt att vara flexibel under utvärderingsprocessen, anpassa den efter resultat och utveckla analysen. I ett jämförande perspektiv är det inte ovanligt att uppdragsbeskrivningar listar en stor mängd detaljerade frågor, men i det avseendet har SADEV utvärderingarna varit exemplariska. Genom väl valda frågor har svåra områden gjorts utvärderingsbara – därmed inte sagt att det som återstår att göra är lätt, men jag bedömer inte att det i något avseende varit omöjligt.

I efterhand kan det konstateras att utvärderingarna haft väl tilltagna resurser. Däremot kan jag inte bedöma hur budgetprocessen hos SADEV gått till och samverkat med arbetsplanen, och inte heller hur och när respektive beslut tagits. Däremot borde den tid och de resurser som redovisats för respektive utvärdering ha varit tillfyllest för att lösa uppgifterna om det från början var klart att man skulle arbeta inom ramen för en sådan resurstilldelning.

Däremot visar genomgången av de fem studierna att man fått ut relativt lite utvärdering för pengarna. Studierna har samlat in förhållandevis lite datamaterial. Det har varit få intervjuer, begränsade enkätundersökningar, i princip ingen observation. Syntesstudier bygger på ett litet dataunderlag. Litteraturgenomgångar täcker en liten del av tillgänglig forskning på områdena. Utvärderingarna lägger förhållandevis lite tid på fältarbete i Sveriges samarbetsländer. Att – som i flera fall - inom ramen för ett till två årsarbetskrafter inte lägga mer än cirka åtta veckor på datainsamling i samarbetsländer framstår som direkt olämpligt. Rapporterna framstår i en del fall som ofullständigt genomarbetade och reflekterar inte att så mycket tid lagts ned.

## Rekommendationer

Uppdraget för den här meta-utvärderingen var att bedöma kvaliteten i fem SADEV-studier. Jag har inte haft möjlighet att söka förklaringar till de problem som har uppmärksammats och har heller inte gått in på hur vanligt förekommande de är. Det är därför risk att rekommendationerna slår in öppna dörrar och har låg relevans. Det är svårt att utfärda rekommendationer om man inte har någon insikt i vad som är orsak till problemen. Rekommendationerna som följer är dock exempel på tänkbara åtgärder som är relativt uppenbara med tanke på de omdömen om kvalitet som redovisats ovan.

Det första gäller utvärderingsteam. Det är bra att utvärderingarna bygger på team snarare än på ensamt arbetande utvärderare. Dock är dessa team uppbyggda kring fast anställda och svenska medarbetare på SADEV. Det vore lämpligt att:

- utse utvärderingsteam med forskare/konsulter från samarbetsländer
- utse team med internationell expertis inom respektive område

Det andra gäller kompetens. De svagheter avseende metod som identifierats ovan kan eventuellt ha sin grund i att den utbildning de anställda har – eller som de beretts inom ramen för sin anställning - inte täckt områden som:

- val av upplägg ('approach', 'design') för komplexa uppdrag, så att val av experimentell metod, programteori, mm. kunnat beakta de möjligheter som står till en utvärderares förfogande och vad de valen får för konsekvenser
- hur man använder teori i ett utvärderingsuppdrag för att bygga på ackumulerad kunskap och utveckla frågor med hjälp av hypoteser<sup>9</sup>
- strategier för urval inom ramen för omfattande och komplicerade utvärderingsområden
- utformning och användning av modeller för utvärdering.
- utformning av datainsamling, intervjumall, enkäter, strukturer för observation mm.

Ett långsiktigt program för kapacitetsutveckling bör utvecklas och genomföras.

Det tredje gäller kvalitetssäkring. Den kvalitetssäkring som SADEV har institutionaliserat har inte fungerat. Det bör utvecklas nya rutiner för kvalitetssäkring, som bl.a. innebär att

- kvalitetssäkring påbörjas tidigt i respektive utvärdering, redan i samband med att upplägg och urval bestäms
- att referensgrupper har tid och möjlighet att påverka det praktiska arbetet
- att interna seminarier är kritiska och ifrågasättande och att organisationskulturen uppmuntrar kritik, diskussion och lärande.

---

<sup>9</sup> med det menas hypotesprövning som det tillämpas i samhällsvetenskaplig forskning – inte hypotesprövning i statistisk bemärkelse.



## Referenser

Carlsson, J., Forss, K., Metell, K., Segnestam, L., and Strömberg, T. (1997) Using the Evaluation Tool - A survey of conventional wisdom and common practice at Sida. A Study commissioned by Sida.

Forss, K. (1994) Development Cooperation with Eastern and Central Europe - a review of evaluation results. A study commissioned by the Swedish Secretariat for Analysis of Development Assistance.

Forss, K. (1994) Några aspekter på kvaliteten inom utvärderingar av svenskt biståndssamarbete. Genomgång av databas upprättad vid Sekretariatet för Analys av Utvecklingssamarbete. En rapport beställd av Sekretariatet för Analys av Utvecklingssamarbete.

Forss, K., Carlsson, J. and Rebien, C. (1997) The Quest for Quality - or can evaluation results be trusted? Evaluation, Vol (4), 481 – 502. 1997

Forss, K., och Uhrwing, M. (2001) Att bedöma kvaliteten i svenskt utredningsväsende – modell och metoder. Ett metodutvecklingsarbete beställt av Regeringskansliet.

Forss, K. Kruse, S-E., Mwaiselage, A., Nilsson, A. and Vedung, E. (2008) Can Evaluations be Trusted? An Inquiry into the Quality of Sida's Evaluation Reports. Commissioned by Sida.

Forss, K. (2009) *Utvärdering som hantverk*. Lund: Studentlitteratur.

Forss, K., M. Marra and R. Schwartz, Eds. (2011). *Evaluating the complex: Attribution, contribution and beyond*. New Brunswick and London, Transaction Publishers

Joint Committee on Standards (1994) *The Program Evaluation Standards*. Sage: Thousand Oaks

Samset, K., Forss, K., and Hauglin, O. (1991) Learning from experience: A study of feedback from the evaluation system in the Norwegian aid administration. A report commissioned by the Ministry of Foreign Affairs, Oslo.

Schwartz, R. and Mayne, J. (2008) *Quality Matters*. New Brunswick and London, Transaction Publishers

Stern, E., Befani, B., Davies, R. Forss, K., Mayne, J. and Stame, N. (2012) Broadening the range of designs and methods for impact evaluations. Commissioned by DFID.

USAID (1988) Review and Quality of A.I.D. Evaluations FY 1987 and FY 1988. Internal Document, Washington D.C.

Watson, P. (1989) A Study of the Quality of CIDA's Bilateral Evaluations. Commissioned by Cida, Ottawa.

## Bilaga 1. Uppdragsbeskrivning

### Konsulttjänst, korttidsuppdrag angående:

#### Uppdraget

Konsulten ska göra en bedömning av ett urval av Institutet för utvärdering av internationellt utvecklingssamarbetets (SADEV) utvärderingar. Underlaget ska nyttjas av Regeringskansliets enhet för styrning av utvecklingssamarbetet UD-USTYR inom ramen för den löpande uppföljningen av SADEV:s verksamhet. Underlaget ska också kunna nyttjas av Statskontoret inom ramen för myndighetens uppdrag om en översyn av utvärderingsverksamheten inom internationellt utvecklingssamarbete (UF2011/xxxx/UD/USTYR), i den mån Statskontoret finner lämpligt.

#### Aktiviteter

- **Inläsning och bedömning:** Analys av utvärderingar utifrån uppdraget
- **Rapport:** Utarbetande av rapport till UD.
- **Möten och kontakter:** Presentation av rapport och frågestund på UD. Övriga behov av möten och kontakter för uppdragets genomförande.

## Bilaga 2. Intervjulist

### SADEV

Gunilla Törnqvist  
Bengt Grude  
Agne Sandberg

### Sida

Lina Lindblom  
Peter Linnér  
Helena Reuterswärd  
Torsten Wetterblad  
Joakim Molander (mailkorrespondens)

### Utrikesdepartementet

Anna Centerstig  
Harriet Pedersen  
Thomas Brundin

### Övriga

Lennart Widell (mailkorrespondens)  
Martin Sjöstedt

## Bilaga 3. Synpunkter på utkast till rapport

SADEV' generaldirektör översände med mail den 12 april 2012 följande kommentarer till rapporten.

"Möjligheten att kommentera rapporten uppskattas. Intrycket är gediget, rapporten är lättläst, och SADEV instämmer i delar av slutsatserna. Metodvalet är adekvat och SADEV finner rekommendationerna såväl bekräftande som användbara.

I utvärderingssamfundet har, så som påpekas i rapporten, mycket tid ägnats åt vad som utgör kvalitet hos en utvärdering. Det är vanligtvis svårt och inte alltid ens önskvärt att redovisa det fullständiga underlaget för utvärderarens bedömningar. I detta fall lyckas utvärderaren, trots avsaknad av redovisning av mer djupgående analyser, göra flera korrekta iakttagelser och lämna pertinenta rekommendationer.

Rapportens rekommendationer pekar på behov av utveckling inom tre områden som SADEV understryker vikten av: kvalitet, kompetens och kommunikation. De är inbördes beroende och nödvändiga för utvärderingarnas användbarhet.

I ex post utvärderingar är det inte ovanligt att slutsatser och rekommendationer bekräftar redan gjorda iakttagelser, så även i detta fall. Det förtar inte rekommendationernas värde men gör att mer konkreta rekommendationer hade välkomnats. Det gäller inte minst kvalitetssäkringen av SADEVs utvärderingar vilken kantats av problem. Ett antal åtgärder är motiverade. Till exempel har myndigheten dragit slutsatsen att mer genomarbetade projektplaner behövs där bland annat mål, roller och ansvarsförhållanden uttrycks tydligare. SADEV har också inför nya projekt engagerat välrenommerade utvärderare som särskilt ska genomlysa förslagen till projektplaner, fungera som rådgivare, samt kritiskt granska utkast till rapport. Möjligheter till kvalitetskontroll genom anlitan av anonyma sakkunniga undersöks också. Dessutom ska i både kvalitets- och kompetenshöjande syfte utvärderingsteamerna i ökad utsträckning bestå av utvärderingskunniga personer från utvecklingsländer och till teamen tillföras vetenskaplig expertis. Under 2011 inleddes en kompetensutveckling med fokus på utvärderingsmetodik av SADEVs personal. Ekonomistyrningsverket får också i uppdrag att lämna förslag till förbättrad intern styrning och kontroll. Med ökad kompetens inklusive nyrekryteringar och fungerande system för kvalitetssäkring bör en mer "akademisk", öppen intern granskning av utvärderingarna uppnås.

Kommunikationens betydelse för kvaliteten i utvärderingarna hade förtjänat mer analys. Till exempel synliggörs inte uttalade önskemål och resultat av dialog mellan beställare och utvärderare i regeringsuppdragen. Avseende uppdraget om den svenska strategin för multilateralt utvecklingssamarbete redovisade SADEV bland annat att det inte kunde bli annat än en uppföljning. I arbetet med Demokratisk utveckling och ökad respekt för mänskliga rättigheter stod det klart att rapporteringen skulle begränsas till ett antal exempel som kunde illustrera resultat. I tolkningen av SADEV Report 2012:1 föreligger också ett missförstånd. Rapporten är inte en syntesstudie utan en rapportering till regeringen ("kappa") i vilken SADEV beskriver hur myndigheten gått

till väga och vad som gjorts inom ramen för uppdraget. Vidare finns i Språklag (2009:600) regler avseende användningen av svenska språket i offentlig verksamhet som gör att SADEV i utvärderingsarbetet behöver arbeta med flera språk mer eller mindre parallellt; regeringsuppdrag rapporteras alltid på svenska.

I SADEV:s uppföljning av rapporter hösten 2011, *Sammanställning av uppföljning av SADEV:s rapporter*, konstateras bl. a :

*"Respondenter har uppskattat när kontakten med SADEV under rapportens produktion och vid publicering har varit tät, medan detta har efterfrågats när kontakten inte varit tät. För flera av rapporterna har mottagande organisation gjort management response på rekommendationerna.*

*Flera respondenter anger att rapporterna har varit användbara för sina organisationer i det dagliga arbetet. För flera av rapporterna hade mottagare på känn det som presenterades i rapporten och respondenter påpekar att det har varit positivt att rapporterna har kunnat påvisa detta med fakta. Organisationerna har därigenom kunnat använda rapporterna som underlag för diskussioner och för att kommunicera det som rapporterna har berört. UD har ofta angett att rapporten har varit användbar och informativ, medan Sida har angett att de redan hade kännedom om ämnet och/eller att rapporten har bidragit med faktamässigt underlag för deras aningar." Det kan illustrera hur olika intressenter kan uppfatta användbarhet.*

Urvalet av rapporter är begränsat och har inneburit att utvärderarna i tre fall inte längre finns på SADEV, medan en kvarvarande medarbetare på SADEV finns för de övriga två utvärderingarna. Det ledde till svårigheter för utvärderaren som konstaterar att det "funnits någon konflikt inom SADEV och mellan/inom utvärderingsteam" (sid 8) som han inte haft tid, möjlighet eller intresse att fördjupa sig i. Mot bakgrund av SADEVs roll som första fristående myndighet att utvärdera utvecklingssamarbete och bistånd och då utvärdering vanligtvis har ambitionen att klargöra orsaksförhållanden är det olyckligt att djupare analys saknas. Befintligt material som utvärderaren inte studerat skulle till exempel visa tecken på intressekonflikter i utvärderingarna, att för projekten specifikt anlidade visstidsanställda personers arbete i tid överskridit det ursprungligen överenskomna, samt bristande efterlevnad av projektplan och bristande hänsyn till referenspersoners kommentarer. Sådana erfarenheter har väglett och fortsätter att vägleda SADEV i kompetensförsörjning och utvecklingen av system för kvalitetssäkring och därmed effektiviteten i arbetet. Mer djupgående analys avseende såväl *propriety* och *feasibility* hade välkomnats, även i ljuset av SADEVs uppdrag och roll som myndighet och därmed sammanhängande implikationer på tidsanvändning.