EBA

06
2017

CONFRONTING THE CONTRADICTION - AN EXPLORATION INTO THE DUAL PURPOSE OF ACCOUNTABILITY AND LEARNING IN AID EVALUATION

Hilde Reinertsen, Kristian Bjørkdahl, Desmond McNeill

Confronting the Contradiction – An exploration into the dual purpose of accountability and learning in aid evaluation

Hilde Reinertsen
University of Oslo

Kristian Bjørkdahl
University of Bergen

Desmond McNeill
University of Oslo

Rapport 2017:06
till
Expertgruppen för biståndsanalys (EBA)

This report can be downloaded free of charge at www.eba.se

ISBN 978-91-88143-28-0

Printed by Elanders Sverige AB
Stockholm 2017

Cover design by Julia Demchenko

Acknowledgements: This study has benefitted greatly from the interest and time of a number of people and institutions.

The Expert Group on Aid Studies (EBA) took an early interest in our ideas and generously supported the project. EBA's secretariat, led by Sonja Daltung, has been a delight to work with due to their combination of professionalism and enthusiasm. Our project officers at EBA, Emma Östaker and Per Trulsson, assisted us throughout the process, from developing our project proposal; preparing the project's contract, schedule, and budget; accessing informants; organising Reference Group meetings; and organising the report launch.

Our designated Reference Group at EBA (led by Eva Lithman, former head of evaluation at Sida, with members Asbjørn Eidhammer, former head of evaluation at Norad; Penny Hawkins, former head of evaluation at DFID, UK; Karolina Hultergren, currently at Sida; and Lennart Wohlgemuth, also former head of evaluation at Sida) have been an invaluable source of critical insight, challenging questions, engaging discussions, and pleasant company. This final report is very much shaped by our interactions with and feed-back from our reference group.

Our informants, without whom this study would not have been possible, have generously shared their time, experience, and reflections, both during interviews and through comments on our final draft.

At the University of Oslo, the Centre for Development and Environment (SUM) hosted the project while the TIK Centre for Technology, Innovation and Culture provided additional office space and working community. Our work has depended on administrative assistance from the ever-efficient Gitte Egenberg and Charlotte Kildal (at SUM) and Frode Løvik (at TIK).

Numerous colleagues have contributed with comments on previous versions of our analysis, notably during SUM's staff seminar and during the conference panel "Analyzing Aid Evaluation" at the Annual Conference of the Norwegian Association for Development Research (NFU). Janet Vähämäki gave most valuable comments on our final draft.

Notwithstanding the many inputs we have received throughout the process from both academics and practitioners, all caveats of course still apply, and the responsibility for all remaining mistakes and limitations is our own.

Hilde Reinertsen is a postdoctoral fellow at the TIK Centre for Technology, Innovation and Culture at the University of Oslo. She is historian by training and holds a PhD in the interdisciplinary field of Science and Technology Studies. In her PhD dissertation, she analysed how aid evaluation was institutionalised in Norway during 1980-1992. She currently works on an EU project investigating the role of policy documents in Norway and the EU within the fields of aquaculture and the bioeconomy.

Kristian Bjørkdahl has a PhD in rhetorical studies from the University of Oslo, and is currently a senior researcher at the Rokkan Centre for Social Studies. Bjørkdahl typically combines rhetoric and reception studies, and his previous work includes studies of American neopragmatism; the animal movement; the discourse of development aid; science advice and public challenges to science; as well as the practice of nonreading.

Desmond McNeill, political economist, is Research Director at the Centre for Development and the Environment (SUM), University of Oslo, Norway. He has worked on development issues since 1969, in Africa, Asia and Latin America – as a researcher, consultant, and adviser (with Norad and DfID). He has undertaken evaluations and project reviews for Norad, SASDA and UNDP. He is the author of several books and numerous articles, most recently in the field of global governance.

Table of contents

Preface

Evaluation is a firmly rooted practice in international development cooperation. It is part and parcel of established routines in order to learn from experience and improve future undertakings. Evaluations also satisfy the need for accountability, i.e. ensuring that dedicated resources, whether they be human, financial or other, are well spent. At the same time, some argue that it is problematic for evaluation as currently practiced to contribute to both learning and accountability.

In this EBA report three researchers from the Centre for Development and Environment at the University of Oslo, Hilde Reinertsen, Kristian Bjørkdahl and Desmond McNeill have explored the dual nature of aid evaluation. They have looked at aid evaluations in Norway and Sweden over the last forty years. Their main conclusion is that "the dual purpose of accountability and learning in practice causes difficult trade-offs". And their blunt assessment is that "learning is crowded out by accountability". If we really want to promote learning, they suggest, maybe we should structure the evaluation process rather differently. In fact, maybe we should limit the engagement of external consultants and do away with formal evaluation reports! On the other hand, they also suggest, if the most important aspect of an evaluation is accountability, the evaluation process might be structured rather differently as well. However, in evaluation practice a clear distinction between learning and accountability is rarely made – and in the effort to kill two birds with one stone, those that commission evaluations end up only wounding the two birds with no clear view of the real benefit from this.

The authors also come to the conclusion that while aid evaluations clearly contribute to accountability, they to a much lesser extent contribute to learning. Unfortunately, it seems, the up-take of recommendations is rather disappointing. The question is, of course, whether there is enough capacity in the aid agencies and ministries concerned to capitalize on the lessons learned, or if the lessons learned through evaluations are simply not relevant enough.

The questions that Hilde Reinertsen and her colleagues raise pose an important challenge to those of us engaged in aid evaluation – not least the EBA itself. The expert group will in the process of launching this report hold a panel discussion around the findings and arguments in this report. A summary of the outcome from that discussion can be

found on our homepage (www.eba.se). We also intend to stimulate the discussion elsewhere in the hope that evaluation practice may evolve to better serve its dual purpose.

The report was produced in dialogue with a reference group under the leadership of Eva Lithman, member of the EBA. The analysis and conclusions expressed in this report are exclusively those of the authors.

Stockholm, May 2017

Gun-Britt Andersson

Sammanfattning

Ett av de viktigaste syftena med biståndsutvärderingar är att dra lärdomar. Men varför lär sig biståndsorganisationer inte mer av sina egna erfarenheter? Eller mer specifikt, varför lär de sig inte mer av sina egna utvärderingar? Det här är frågor som allmänheten, politiker och personer verksamma inom biståndsorganisationer och utvärderingar har ställt sig under mer än 30 år. Lärdomar är dock endast ett av de två allmänt vedertagna syftena med biståndsutvärderingar. Det andra viktiga syftet är ansvarsutkrävande. I den här studien undersöker vi varför det ofta är svårt att förena dessa två syften i praktiken.

Vår huvudsakliga slutsats är att *de dubbla syftena – ansvarsutkrävande och lärande – i praktiken medför svåra avvägningar*. Vår slutsats är baserad på en empirisk analys av nuvarande och tidigare metoder för biståndsutvärdering i Sverige och Norge. I analysen studerar vi utvärderingar på tre nivåer som vi menar är nära sammankopplade: utvärderingsrapporterna, de praktiska utvärderingarna och utvärderingssystemen mer allmänt, inklusive de bredare politiska och förvaltningsmässiga ramarna för utvecklingsbistånd. Vårt empiriska material består av djupintervjuer med seniora utvärderingschefer, en kartläggning av historiska dokument (utvärderingshandböcker, nyhetsbrev, rapporter osv.) och ett mindre urval utvärderingsrapporter. Dessutom gör vi en översyn av befintlig litteratur om bistånd (från akademisk forskning till publikationer baserade på praktiska erfarenheter) som särskilt tar upp ansvarsutkrävande och lärande som dubbla utvärderingssyften. Denna sammanfattning innehåller inte detaljerad information om källor och referenser, men sådan information finns tillgänglig i motsvarande avsnitt i den fullständiga rapporten.

Även om det här är en studie *om* biståndsutvärderingar så är det inte en metautvärdering av befintliga rapporter. Det är en mycket viktig distinktion. Vi tillämpar inte redan vedertagna kriterier för biståndsutvärdering, i stället är dessa kriterier en del av det vi undersöker. Vår metod är expansiv och vi utforskar den mängd texter, metoder, historia och sammanhang som finns inom biståndsutvärdering. Vår tvärvetenskapliga grupp består av författare med bakgrunder inom akademiska områden som retorik, historia, politisk ekonomi samt teknik- och vetenskapsstudier (STS) och vår utgångspunkt är att biståndsutvärdering är en egen och fascinerande form av kunskapsproduktion. Samtidigt är det en egen retorisk genre,

ett standardiserat praktiskt förfarande, ett väletablerat expertområde och ett myndighetsverktyg. Genom att kombinera dessa olika aspekter, och i vår frågeställning utgå från de två syftenas eventuella inneboende motsägelser, undersöker vi biståndsutvärdering utifrån alla dess praktiska dilemman, problem och oklarheter.

Det gör att det här inte är en vanlig biståndsutvärderingsrapport. Vårt mål är i stället att skapa något som kan fungera som en utgångspunkt för vidare diskussioner. Vi gör inte anspråk på att ha nått fram till den enda rätta slutsatsen, och vi förväntar oss inte medhåll från alla läsare. Erfarenhetsbanken inom bistånd är omfattande och vi har endast använt oss av en mindre del. Det finns med nödvändighet många andra exempel som både bekräftar och motsäger vår slutsats. Trots det vill vi att vår analys ska vara belysande och tankeväckande och att den kan fungera som en bra utgångspunkt för vidare diskussioner och undersökningar. Även om den samlade gruppen sakkunniga inom biståndsutvärdering huvudsakligen vidmakthåller (i publikationer och offentliga uttalanden) att det går att uppnå båda syftena pekar vår empiriska undersökning av texter och metoder för biståndsutvärdering, samt genomförda intervjuer med viktiga informanter, på att det tydligt finns problem, konflikter, avvägningar och motsägelser. Av litteraturöversynen framgår dessutom att det pågår en allt livligare debatt om de dubbla syftena. Vi hoppas att detta kan få personer verksamma inom utvärderingar och bistånd, samt beslutsfattare och allmänheten, att uppmärksamma och öppet debattera och diskutera de inneboende utmaningar som vi har identifierat.

Huvudresultat

Vår integrerade analys av utvärderingstexter, utvärderingsprocesser och utvärderingssystem visar att det kan vara svårt att förena ansvarsutkrävande och lärande, och att det ibland uppstår direkta motsägelser. Nedan presenterar vi våra viktigaste slutsatser på respektive nivå. Vår analys har vägletts av frågorna: Vem utarbetar och vem läser utvärderingsrapporterna? Hur tas de fram, och hur distribueras och används de? Vem lär sig av utvärderingarna, och på vilket sätt? Hur hanteras de olika aspekterna av ansvar och lärande i rapporter och system, samt av personal? Och hur har detta skiljt sig över tid och mellan Sverige och Norge?

Utvärderingstexten

En retorisk analys av ett urval utvärderingsrapporter visar att även om det tydligt framgår att de kan bidra till ansvarstagande så bidrar de i mycket mindre utsträckning till lärande. Den här genomgående slutsatsen kan dras över tid och i båda länderna. Även om rapporterna vid en första anblick kan *framstå* som annorlunda jämfört med 40 år sedan har ganska lite förändrats vad gäller struktur och innehåll. Det finns flera olika genrer av utvärderingsrapporter. Huvudrapporten är i allmänhet en väl etablerad genre där tre klassiska retoriska element kombineras: först fastställs vad som hände, därefter vem som ska berömmas eller skuldbeläggas och sist följer förslag på åtgärder.

Vårt urval består av 20 utvärderingsrapporter, och det första och andra retoriska elementet (fastställa vad som hände och fördela beröm eller skuld) ingår till stor del i beskrivningen och analysen. Därigenom uppfylls utvärderingens ansvarssyfte. Det tredje elementet (förslag på åtgärder) täcks genom de obligatoriska delarna ”rekommendationer” och ”erfarenheter”. Ofta är dock dessa avsnitt endast löst baserade på den föregående analysen. I de flesta av de studerade rapporterna bortser rekommendationerna från viktiga faktorer i sammanhanget, trots att sammanhangets betydelse uttryckligen betonats i tidigare avsnitt i samma rapport. Detta vidgar avståndet mellan beskrivningen och rekommendationerna och försvårar avsevärt möjligheterna att dra lärdomar från utvärderingar. Även om det kan tyda på att rapporterna helt enkelt håller en låg kvalitet är vår slutsats ändå att förbättrad kvalitet inte är lösningen. Det är även nödvändigt att undersöka i vilken omfattning som kvaliteten är avhängig processer och strukturer som ligger utanför själva rapporten, särskilt hur uppdragsbeskrivningen formuleras av rapportbeställare och vilka resurser som är tillgängliga för biståndsutvärderingen.

Utvärderingsprocessen

Både Sida och Norad har väl etablerade formella rutiner för att planera en utvärdering, förbereda uppdragsbeskrivningen, inrätta en utvärderingsgrupp, genomföra utvärderingen och följa upp den publicerade rapporten. Redan i utvärderingens inledande skede fastställs om den främst ska vara inriktad på ansvarsutkrävande eller lärande. Respektive syfte medför olika frågeställningar och metoder. Formella rutiner kompletteras dessutom med informella metoder. För att säkerställa samarbete, intresse, förtroende och, slutligen, lärdomar och användning är det mycket viktigt att skapa och bevara ett internt

engagemang för utvärderingen. Detta måste dock hela tiden vägas mot ansvarsprinciperna om kritisk distans och oberoende i fråga om ansvarsutkrävande eftersom ett för stort internt engagemang kan inverka på det externa förtroendet för utvärderingen.

Flera svåra dilemman kan uppstå: Ska utvärderingsgruppen genomföra en revision, eller ska de underlätta processen? Ska de utarbeta rapporten i huvudsak för extern styrning eller intern förändring? Ska gruppen prioritera internt eller externt förtroende? Transparenta processer och metodologisk stringens kan till viss del underlätta balansen mellan dessa skiljda frågeställningar men det går inte helt att undvika kompromisser. Möjligheten för lärande påverkas direkt av den roll som externa konsulter ges och självmant tar. Om tonvikten läggs i deras roll som kritiker kan det skapa en opedagogisk arbetssituation med defensiva människor, vilket i sin tur kan leda till att möjligheterna att dra lärdomar förloras. Dessutom uppfattas deras rekommendationer ofta som olämpliga. De kan ses som alltför detaljerade eller generella, eller alltför ambitiösa. Det mest grundläggande problemet med externa konsulter är att de som lär sig mest av processen inte har något ansvar för att tillämpa lärdomarna. Det är i sin tur knutet till den enkla frågan om *vem som utarbetar utvärderingsrapporterna*. Genom att det praktiska analys- och skrivarbetet i huvudsak görs utanför biståndsorganen upprätthålls utvärderingarnas ansvarssyfte, men det innebär även att viktiga kunskaper går förlorade i biståndsorganisationerna.

På motsvarande sätt är det viktigt att ställa sig frågan om *vem som läser utvärderingsrapporterna*. För de som genomför utvärderingen är det en stor utmaning att föra tillbaka lärdomar genom rapporten och relaterade samordnings- och kommunikationsåtgärder till organisationen. De upplever att få har tid att läsa utvärderingsrapporter och ta till sig innehållet. Vår analys ställer grundläggande frågor: Varför är det så viktigt att ta fram rapporter som så få kommer att läsa? Varför är det viktigare att anlita externa konsulter än att underlätta interna lärandeprocesser? Svaren på dessa frågor hänför sig till det bredare sammanhang där biståndsutvärderingarna genomförs.

Utvärderingssystemet

Biståndsutvärderingar ingår alltid i ett större sammanhang. En av de intervjuade beskrev träffande sammanhanget som fyllt av ”krafter med olika önskemål och intressen”. Under de senaste 40 åren har Sverige

och Norge vid upprepade tillfällen omorganiserat verksamheten för biståndsutvärderingar och man har valt olika sätt att balansera frågorna om integrering/distans, medverkan/styrning och ansvarsutkrävande/lärande. Precis som i fallet med utvärderingsrapporter och utvärderingsförfaranden finns ingen perfekt lösning. I stället krävs pragmatiska val mellan viktiga frågor som rent konkret innebär svåra avvägningar. Eftersom utvärderingsrapporter ger utomstående en inblick i vad som pågår i biståndsvärlden finns ett uppenbart demokrativärde och rapporterna är nödvändiga för att upprätthålla allmänhetens förtroende för biståndet. Men när ansvarsutkrävande definieras alltför snävt och enbart innebär rapportering av dokumenterade resultat kan det ske på bekostnad av lärandet.

Vid en jämförelse av Sveriges och Norges utvärderingssystem framträder två huvudsakliga särdrag: Det första är att det svenska utvärderingssystemet till stor del är decentraliserat, vilket medför att programbaserade utvärderingar även ses som en central del i utvärderingssystemet. Den centrala enheten har genomfört strategiska utvärderingar och tillhandahållit stöd vid decentraliserade utvärderingar. I Norge finns däremot en tydlig uppdelning mellan centralt framtagna utvärderingar och decentraliserade utvärderingar, som fram till helt nyligen kallades för programöversyner (heter fortfarande så på norska). För det andra finns det märkbara skillnader i hur de två länderna har valt att organisera arbetet med frågor som rör integrering och oberoende. Historiskt sett har den norska utvärderingsenheten flyttats från Norad till utrikesministeriet och åter till Norad. Under processens gång har enheten gått från en halvautonom till en integrerad modell och sedan tillbaka igen. Sidas centrala utvärderingsenhet har också genomgått tydliga förändringar – från att först ha varit en egen enhet inom organisationen till att utvecklas till en stark halvautonom enhet för att därefter åter integreras i organisationen. En annat centralt inslag i den svenska modellen, som fungerar som ett komplement till Sidas egen utvärderingsverksamhet, är att riksdagen och UD vid upprepade tillfällen inrättat externa organ (SASDA, EGDI, SADEV och EBA) som också haft i uppgift att utvärdera biståndet.

Valet av utvärderingssystem inverkar tydligt på hur och i vilka sammanhang utvärderingar kan användas som ett bidrag till antingen lärande eller ansvar, eller till båda delar. De visar således på olika sätt att hantera de centrala frågorna om ”vilkas ansvar” och ”vilkas lärande” det handlar om: Bör ansvarskedjan ”i hemgående riktning” väga tyngre

än ansvaret gentemot biståndsmottagarna, biståndsförmedlarna och de slutliga förmånstagarna? Kan lärande erkännas omfatta projektrelaterat lärande baserat på inkluderande utvärderingsprocesser, eller räcker inte det ur ett givarperspektiv? Det sätt på vilket givarländer väljer att hantera dessa viktiga frågor påverkar i sin tur utvärderingarnas roll.

Vem drar lärdom av utvärderingarna?

Eftersom det främst är externa aktörer som utarbetar utvärderingsrapporter och det är få som läser dessa är frågan *vem det är som drar lärdom av utvärderingsrapporterna*, eller generellt sett, av utvärderingarna? Vår analys visar att lärandet mycket väl kan ske på programnivå, i synnerhet bland externa konsulter, utvärderingsledare och handläggare som deltar i specifika utvärderingar. Detta betonades av alla våra informanter, oavsett om de arbetade med decentraliserade eller centraliserade utvärderingar. Det man kan kalla ”sidoinlärning”, bland aktörer som är involverade i specifika utvärderingar, rapporteras således vara vanligt förekommande. Men även inom dessa praktiska utvärderingar stötte vi flera gånger på exempel på hur lärandet kan begränsas av det laddade förhållandet mellan ansvar och lärande.

Begreppet ”sidoinlärning” berör främst dem som arbetar för givarorganen, antingen som anställda inom utvärderings-, program- eller styrningsverksamheten eller som externa konsulter. Partnerorgan, biståndsförmedlare och slutliga stödmottagare utgör ytterligare en rad relevanta grupper som befinner sig nära givarna. Frågan är om utvärderingen ska handla *om* dem, göras *tillsammans med* dem eller till och med *av* dem? Enligt vissa informanter involveras förmedlare och mottagare i själva verket i bästa fall i egenskap av berörda aktörer, men sällan som aktiva partner i utvärderingsprocessen. Andra informanter delade inte denna uppfattning och pekade på att man, utöver Sidas och Norads utvärderingssystem, även måste se till samarbetsorganisationernas egna utvärderingssystem, som inte utformats för att möjliggöra lärande bland givarna utan bland mottagarna. I sammanhanget avses därmed samarbetspartner och utförare.

De slutliga förmånstagarna spelar således endast en begränsad roll i givarnas egna utvärderingssystem. Eller som en informant uttryckte det: ”Vi utvärderar för vår egen räkning”. Ändå finns det betydande skillnader mellan lärande som sker på plats och hemmavid. Det är svårt att dra allmänna slutsatser av och sammanställa resultaten från

utvärderingarna och skapa ett brett lärande på organisatorisk nivå, vilket en av våra informanter kallade ”big learning”. Även om de centrala utvärderingsenheterna har försökt att möjliggöra detta på olika sätt under flera årtionden – med hjälp av årliga rapporter, nyhetsbrev, sammanfattande rapporter, offentliga databaser och uppföljningsplaner – så är det fortfarande svårt att åstadkomma s.k. ”big learning”.

Ett grundläggande problem är slutligen att det på samtliga nivåer i biståndssystemet finns en överdriven förväntan på vad som kan uppnås med hjälp av biståndsutvärderingar. Många förväntar sig att den expansiva ökningen av utvärderingsrapporter och annan tillgänglig dokumentation och information automatiskt ska leda till ökad kunskap och lärande. Denna linjära modell för lärande överensstämmer dock inte med de praktiska erfarenheterna på området. Den nuvarande situationen med ”big aid data” på biståndsområdet utgör inget botemedel för den utbredda upplevelsen att vi vet och lär oss för lite. Problemet förvärras av förstärkta krav på öppenhet och insyn, ansvar, revision och tillsyn, som samtidigt som de tjänar viktiga demokratiska syften idag används på sätt som inte nödvändigtvis fungerar väl i förhållande till lärandemålet. Man lär sig inte minst av att begå misstag och man måste räkna med att biståndsverksamheten inbegriper ett *flertal* misstag. En realistisk ansats skulle således vara att ha en hög tolerans för fel. I verkligheten är förväntningarna på biståndet mycket tuffare än så. Om man inte når målen inom ramen för en biståndsinsats, om finansieringen blir föremål för korruption och om effekterna inte blir de avsedda är medierna – och i vissa fall politikerna – snabba att skapa en skandal medan biståndsorganet tvingas försvara sig offentligt. Detta kan skapa fördjupad misstro både externt (mot biståndsorganet) och internt (mot det biståndsutvärderande organet).

Avslutningsvis visar det vi har beskrivit ovan på en ständig avvägning på olika nivåer mellan ansvarsutkrävande och lärande vid biståndsutvärderingar. I praktiken resulterar detta huvudsakligen i att det förstnämnda prioriteras på bekostnad av det sistnämnda. Enkelt uttryckt så konkurrerar ansvarsutkrävande ut lärandet.

Huvudrekommendationer

1. Vi måste tala öppet om avvägningen mellan ansvarsutkrävande och lärande.

2. Vi måste anpassa våra förväntningar både när det gäller biståndsinsatser och biståndsutvärderingar.

Med ”vi” avses här alla som medverkar i eller diskuterar biståndsutvärderingar, från utvärderingschefer, personer som arbetar med bistånd och beslutsfattare till forskare och den breda allmänheten. För att följa dessa rekommendationer menar vi att både de som arbetar med bistånd och andra som diskuterar bistånd behöver inse att ett antal ställningstaganden måste göras, oavsett deras ståndpunkt i den fråga som diskuteras i den här rapporten. Följande förteckning är inte uttömmande men återger de viktigaste ställnigstaganden som ofta görs idag, utan att direkt diskutera deras följder.

Ställningstagande 1: Behövs det en utvärderingsrapport i utvärderingen, och i sådana fall vilken slags rapport? Eftersom alltför många utvärderingsrapporter knappt läses bör man alltid besvara frågan om huruvida en rapport behövs, samt om så är fallet, vilket syfte den bör fylla och hur den bör utarbetas. Detta inbegriper att fastställa avsedda läsare och användare, vilket i sin tur bör ligga till grund för valet av rapportskrivare. Om syftet rör extern ansvarsutkrävande kan det räcka med en kortfattad rapport som beskriver existerande verksamhet och resultat. Om syftet rör internt lärande kan en offentliggjord och allmänt tillgänglig rapport ha kontraproduktiv effekt.

Ställningstagande 2: Skulle utvärderingen gagnas av att genomföras av en grupp externa utvärderare? Användningen av externa konsulter bör vägas mot deras kostnad och mervärdet uttryckligen motiveras. Det finns en direkt koppling mellan konsultens roll och syftet med utvärderingen. Om syftet är ansvarsrelaterat kan en begränsad revision vara bäst. Om syftet är lärande kan utvärderarna snarare underlätta processen genom att bl.a. erbjuda ett neutralt och externt perspektiv. Interna deltagare och externa aktörer måste aktivt inkluderas under hela processens gång, åtminstone genom en självutvärdering som ges lika stor tyngd som den externa utvärderingen.

Ställningstagande 3: Bör utvärderingsrapporten innehålla rekommendationer? Rekommendationer tas vanligtvis fram gemensamt av utvärderarna som en del av utvärderingsuppdraget. Utarbetandet av rekommendationer utgör oftast den svagaste länken i utvärderingsprocessen, trots att den också är den viktigaste. Det är i det skedet som den kartläggning och de analyser som tagits fram under utvärderingsprocessens gång kan omsättas i eventuella åtgärder.

Det är inte säkert att det är utvärderarna som är bäst lämpade att utarbeta rekommendationerna. Andra modeller kan vara mer användbara: Utvärderarna kan istället föreslå en rad scenarier som berörd programpersonal och beslutsfattare kan välja bland efter att ha satt sig in i de potentiella avvägningarna. Rekommendationer kan utarbetas av dessa parter, eventuellt med stöd av utvärderingsgruppen. Eller så kan utvärderingens avsedda användare då de erhållit rapporten få i uppgift att utarbeta rekommendationer som de sedan ansvarar för.

De tre ovannämnda ställningstagandena är praktiska uttryck för våra övergripande rekommendationer och gäller främst utvärderingsrapporter och frågor i utvärderingssammanhang. Samtidigt har våra rekommendationer även koppling till mer grundläggande frågor om utvecklingsbiståndets berättigande överlag och förväntningarna bland externa aktörer – beslutsfattare, kommentatorer och allmänheten – kring vad biståndsutvärderingar bör innehålla och användas till. Denna mer grundläggande fråga behandlas i den sista urvalsfrågan.

Ställningstagande 4: Bör system för ansvarsutkrävande prioriteras så högt som de gör idag bland givare, även om det sker på bekostnad av det interna lärandet? Det finns ett uppenbart demokratiskt behov av system för övervakning och utvärdering av bistånd eftersom de främjar ansvar samt öppenhet och insyn för skattebetalarna. Dock finns det i teorin inte någon gräns för hur omfattande sådana ansvarssystem kan bli, och de har blivit allt mer krävande över tid. Både inom och utanför biståndsvärlden behövs därför en debatt om valet mellan att utöka användningen av ansvarsfokuserade utvärderingssystem och att göra det möjligt att lägga större tonvikt på lärandet. De som efterlyser mer omfattande tillsynssystem samt starkare framgångsbevis bör således vara medvetna om den faktiska kostnaden för det de begär i form av ökade budgetutgifter, administrativt arbete, organisatorisk stress och minskad potential för lärande.

Summary

Learning is a key purpose of aid evaluation. So why do aid organisations not learn more from their own experiences? More specifically, why do they not learn more from their own evaluations? For more than 30 years these questions have been asked by the public, by politicians, by aid staff, and by evaluation professionals. Yet learning is but one part of the well-established "dual purpose" of aid evaluation: The other key purpose is accountability. In this study, we investigate how these two purposes are often difficult to reconcile in practice.

Our main conclusion is that *the dual purpose of accountability and learning in practice causes difficult trade-offs*. We base this conclusion on an empirical analysis of the current and historical practices of aid evaluation in Sweden and Norway. In our analysis, we study evaluation on three levels that, we emphasise, are tightly interconnected: the evaluation reports themselves, the practical evaluation processes, and the wider evaluation systems, including the political and administrative context of development aid at large. Our empirical material consists of in-depth interviews with senior evaluation managers, a mapping of historical documents (evaluation manuals, newsletters, reports etc), and a small sample of evaluation reports. In addition, we review the existing literature (from academic research to practice-based publications) specifically that which addresses the dual purpose of accountability and learning in aid evaluation. The numerous sources and references are not detailed in this executive summary, but readers will find them in corresponding sections of the full report.

While this is a study *of* aid evaluation, it is not a meta-evaluation of existing reports. This distinction is critical: We do not apply the already standard, accepted criteria of aid assessment; rather, these criteria are in themselves part of what we study. We take an expansive approach and explore the multitude of texts, methods, histories, and contexts of aid evaluation. Being an interdisciplinary team with backgrounds from the academic traditions of rhetoric, history, political economy, and science and technology studies (STS), we approach aid evaluation as a particular – and particularly fascinating – form of knowledge production: It is at the same time a specific rhetorical genre, a standardised practical procedure, a well-established field of expertise, and a tool of government. By seeing these aspects in

combination, while having the potential inherent contradictions of the dual purpose as our organising research problem, we explore the everyday life of aid evaluation with all its practical dilemmas, concerns, and uncertainties.

As such, this is not a typical aid evaluation report. Rather, we have intended it to be a conversation starter. We do not claim to have reached the one and only true conclusion, and we do not expect all readers to agree with us. The sheer multitude of aid experiences, of which we have captured just a few, necessarily entails that there are numerous other examples confirming and contradicting our conclusion. Yet we do hope our analysis is illuminating and thought-provoking and that it may serve as a useful starting point for further discussions and investigations. While the professional aid evaluation community largely maintains (in their publications and public statements) that the dual purpose is possible to achieve, our empirical investigations into the actual texts and practices of aid evaluation, including interviews with key informants, demonstrate that dilemmas, tensions, trade-offs, and contradictions clearly do arise. Furthermore, the literature review shows that the dual purpose is in itself a topic of increasingly lively debate. This should, or so we hope, prompt evaluation staff, aid staff, policy makers, and the wider public to acknowledge and openly take up the debate and discuss the in-built challenges that we have identified.

Main findings

Our integrated analysis of evaluation texts, evaluation processes, and evaluation systems shows how tensions, and sometimes direct contradictions, between accountability and learning arise. In the following, we present our main findings from each level. Key questions guiding our analysis have been: Who writes and reads evaluation reports? How are they produced, circulated, and used? Who learns from evaluations, and how? How do reports, staff, and systems negotiate between the diverging concerns of accountability and learning? And how has this varied over time and between Sweden and Norway?

The evaluation text

Our rhetorical analysis of a sample of evaluation reports shows that while they may clearly contribute to accountability, they to a much

lesser extent contribute to learning. This finding is consistent over time and between the two countries. Although the reports at first sight *look* different than they did 40 years ago, they have changed rather little in terms of structure and content. While there exist several sub-genres of evaluation reports, the main report genre is generally well-established and combines the three classic rhetorical elements: to establish what happened, to allocate praise or blame, and to propose what to do.

In our sample of 20 evaluation reports, the first and second rhetorical elements (establish what happened and allocate praise or blame) are largely covered through description and analysis. This contributes to fulfil the accountability purpose of evaluation. The third element (propose what to do) is covered through the mandatory sections of "recommendations" and "lessons learned". Yet these sections are most often only loosely based on the preceding analysis. In most of the reports we studied, the recommendations disregard critical contextual factors even when the importance of context is explicitly noted in earlier sections of the same report. This further deepens the disconnection between description and recommendations, which greatly impedes the potential learning from evaluations. While this could mean that the reports are simply of low quality, we conclude that improving the quality is an insufficient solution; it is also necessary to consider how the quality is contingent on processes and structures outside the report itself, notably by how the Terms of Reference (ToR) are formulated by those commissioning the evaluation report and by the resources made available for aid evaluation.

The evaluation process

Both Sida and Norad have well-established formalised routines for how to plan an evaluation, prepare the Terms of Reference (ToR), procure an evaluation team, lead the evaluation process, and follow up the published report. Already at the starting point in the evaluation process, key premises are established for whether an evaluation will contribute primarily to accountability or learning. The two purposes involve asking different sets of questions and applying diverging methods. Furthermore, the formal routines are complemented by informal practices. Building and sustaining internal engagement for the evaluation is critical to ensure cooperation, interest, trust, and, ultimately, learning and use. But this must constantly be balanced against the accountability principles of critical distance and

independence, as too much internal involvement may reduce the external trust in the evaluation process.

This situation poses important dilemmas: Should the evaluation team function as auditors or process facilitators? Should they write their report mainly for external control or internal change? Should they prioritise internal or external trust? Transparent processes and methodological rigour may enable some reconciliation between these diverging concerns, but they cannot completely avoid the trade-offs. The role assigned to and taken by the external consultants directly affects the learning potential. If they take on an exaggerated role as critics, they may end up conducting their work in an unpedagogical manner that makes people defensive, which in turn means that learning opportunities may be lost. Furthermore, their recommendations are often perceived to be inappropriate; they could be too specific, or too general, or too ambitious. Yet the most fundamental problem with using external consultants is that those who learn the most in the process have no responsibility for applying the lessons. This relates to the simple question of *who writes evaluation reports*. The fact that the practical work of analysis and writing is mainly done outside the aid agencies themselves clearly serves the accountability purpose of evaluation, yet it also means that important learning disappears from the aid agencies.

Correspondingly, asking *who reads evaluation reports* is illuminating. Feeding lessons learned back into the organisation by means of the evaluation reports and related efforts at synthesis and communication remains a considerable challenge for the evaluation staff. Their main experience is that few have the time to read evaluation reports and absorb their content. Our analysis prompts fundamental questions: Why is it so important to keep producing reports that few will read? Why is the procurement of external consultants more important than enabling internal learning processes? Answers to these questions relate to the wider context in which aid evaluations take place.

The evaluation system

Aid evaluation is always but one part of a larger context, what one of our informants aptly called "a power field of diverging concerns and interests". Sweden and Norway have repeatedly re-organised their aid evaluation activities during the past 40 years, choosing different ways of balancing the concerns for integration/distance,

involvement/control, and accountability/learning. Again, as in the case of evaluation reports and evaluation processes, there exists no perfect solution; rather, the balancing act involves making pragmatic choices between important concerns that in effect involve difficult trade-offs. Given that evaluation reports make visible to outsiders what happens inside the world of aid, they are of obvious democratic value and a necessary means for maintaining public trust in aid. But when accountability is too narrowly defined to mean merely the reporting of documented results, it may clearly come at the cost of learning.

Two main comparative features of the Swedish and Norwegian evaluation systems stand out: Firstly, in Sweden, the evaluation system is largely decentralised, which means that programme-based evaluations are also considered a key part of the evaluation system. The central unit has produced strategic evaluations and assisted in decentralised evaluations. In contrast, in Norway, there is a clear separation between the centrally produced evaluations and decentralised evaluations, which until recently was termed programme reviews (and still is in Norwegian). Second, there are notable in differences in how the two countries have chosen to institutionalise the two concerns of integration and autonomy. The Norwegian evaluation unit historically has moved from Norad into the Ministry of Foreign Affairs and back to Norad, and in the process shifted from semi-autonomy to an integrated model and back to semi-autonomy. Sida's central evaluation unit has also experienced clear shifts – from being first its own unit within the wider organisation, then expanded into a strong semi-autonomous unit before again being integrated into the wider organisation. Yet a key feature of the Swedish model that complements Sida's own evalaution work has been the repeated establishments by Parliament and the MFA of external agencies (SASDA, EGDI, SADEV, and EBA) that were also tasked with aid evaluation.

The choice of evaluation system clearly has implications for how and where evaluation may contribute to either learning, accountability, or both. As such, they are manifestations of different ways of answering the key questions of "accountability for whom" and "learning for whom": Should the accountability chain "homewards" be given more weight than the accountability towards aid recipients, aid intermediaries, and end beneficiaries? May learning be acknowledged to mean project-level learning based on inclusive evaluation processes, or is this insufficient from a donor perspective? How donor countries

choose to handle these important questions in turn directly affects what role evaluation may play.

Who learns from evaluations?

Given that mainly external actors write evaluation reports and few people read them, *who learns from evaluation reports*, or more broadly, from evaluation processes? Our analysis shows that learning may well happen at the programme level, notably for the external consultants, evaluation managers, and programme officers partaking in specific evaluation processes. All our informants emphasised this point, whether they were concerned with decentralised or centralised evaluations. What we may term "sideways learning", for actors involved in specific evaluation processes, is thus reported to be common. Yet also in these practical evaluation processes we repeatedly encountered examples of how learning might be limited by the tension between accountability and learning.

The notion of "sideways learning" mainly involves those working for the donor agencies, whether as evaluation staff, programme staff, policy staff, or external consultants. The role of partner organisations, aid mediaries, and end beneficiaries is yet another set of relevant groups one step removed from the donors. Is evaluation supposed to be *about* them, *with* them, or even *by* them? In effect, according to some informants, recipients and beneficiaries were at best included as stakeholders, but rarerly made active partners in the evaluation process itself. Other informants disagreed with this understanding and pointed beyond the evaluation systems of Sida and Norad, noting that one also needed to take into account the partner organisations' own evaluation systems, which were designed to enable learning not for the donors, but for the recipients, here meaning the partners and implementers themselves.

The end beneficiaries of aid thus hold only a limited role in the donors' own evaluation systems. As one of our informants stated: "We evaluate for ourselves." Yet the difference between learning on-site and learning at home is considerable: It is most challenging to generalise and synthesise findings from evaluations and achieve learning on a larger organisational scale, what one of our informants called "big learning". While the central evaluation units have sought to enable this in multiple ways during several decades – through the means of annual reports, newsletters, synthesis reports, public databases, and follow-up plans – "big learning" remains elusive.

Finally, a fundamental problem is that the aid system on all levels displays exaggerated expectations of what aid evaluation may accomplish. The expansive growth of evaluation reports and other available documentation and information makes many assume that increased knowledge and learning will automatically follow. Yet this linear learning model does not match the practical experiences in the field: The current situation of "big aid data" does not remedy the widespread experience that we know too little and learn too little. This problem is only deepened by intensified calls for transparency, accountability, audit, and control, which, while serving critical democratic functions, are currently operationalised in ways that do not necessarily harmonize well with the ambition to learn. One learns not least by making mistakes, and one must expect aid work to involve making *many* mistakes. A realistic approach would thus entail high tolerance for error. In reality, the expectations to aid are much stricter than this. If an aid effort fails to achieve its goals, if funds fall to corruption, or if the impacts are not what one had planned, the media, and – in some cases – politicians are quick to make a scandal of it, while the aid administration is forced to defend itself in public. This may deepen distrust both externally (to the institution of aid) and internally (to the institution of aid evaluation).

To conclude, what we have described above are all expressions, on different levels, of a persistent trade-off between accountability and learning in aid evaluation. In practice, the main result of this is a prioritisation of the former at the expense of the latter. To put it simply: Learning is crowded out by accountability.

Key recommendations

1. We must talk openly about the trade-offs between accountability and learning.
2. We must adjust our expectations to both aid interventions and aid evaluations.

The term "we" here points to everyone involved in doing and discussing aid evalution: from evaluation managers, aid practioners and policy-makers to researchers and the wider public. Following these recommendations would, we suggest, require that both those involved in aid and those discussing it on the outside must

acknowledge that regardless of their own position on the topic discussed in this report, a set of choices will have to be made. The following list is not exhaustive, but it captures the most important choices that are now often made without explicit discussion of their implications.

Choice 1: Does the evaluation process need an evaluation report, and if so, what kind? Too many evaluation reports are hardly read. One should therefore always answer the question of whether a report is needed, and if it is, what purpose it should fulfil and how it thus should be produced. This includes determining its intended readers and users, which in turn should inform who the writers should be. If the purpose is external accountability, then a consise report mapping existing activities and outcomes may be sufficient. If the purpose is internal learning, then a published, publicly available report may be counter-productive.

Choice 2: Does the evaluation process benefit from an external evaluation team? The use of external consultants should be weighed against their cost, and their added value should be explicitly justified. The role of consultants is directly related to the purpose of the evaluation. If the purpose is accountability, then a limited audit mission might be most beneficial. If the purpose is learning, then the team may rather function as facilitators of the evaluation process, providing a neutral outsider perspective. Internal participants and external stakeholders must be actively included throughout the process, at the minimum through a self-evaluation that is granted equal weight as the external evaluation.

Choice 3: Should the evaluation report include recommendations? Recommendations are commonly produced by the evaluation team as part of the evaluation assignment. The articulation of recommendations is often the weakest point of the evaluation process, yet it is also the most important one. This is where the mapping and analyses produced through the evaluation process may be translated into potential action. It is not a given that the evaluation team are best equipped to articulate recommendations. Other models may be more useful: The team could instead suggest a set of scenarios from which the involved programme staff and policy makers may choose, after being well-informed of the potential trade-offs thus involved. Recommendations may be articulated by them, possibly in a process facilitated by the evaluation team. Or the intended users of an evaluation may have the responsibility, upon receiving the report, to

articulate recommendations to which they in turn will be held accountable.

The three choices above are practical manifestations of our overall recommendations, and they pertain mainly to evaluation processes and concerns within the evaluation community. At the same time, our recommendations also connect to more fundamental questions about the legitimacy of development aid at large, and the expectations of external actors – policy-makers, commentators, the public – of what aid evaluation should be and what it should achieve. Our final choice addresses this more fundamental issue.

Choice 4. Should accountability systems be given the current high priority by donors, even when they come at the expense of internal learning? There is an obvious, democratic need for systems of monitoring and evaluation of aid, because they promote accountability and transparency to taxpayers. There is, however, in theory no limit to how comprehensive such accountability systems can be; and they have become steadily more demanding over time. There should thus be a debate, both within and outside the aid community, about the choice between enhancing the accountability-focused evaluation systems and allowing a greater emphasis on learning. Those calling for more comprehensive systems of control and stronger evidence of success should thus acknowledge the actual cost of their demands in terms of increased budgetary expenses, administrative work, and organisational stress, and reduced learning potential.

Introduction

"When will we ever learn?" This was the title of a report prepared one decade ago, in 2006, by the so-called Evaluation Gap Working Group which had been convened by the Centre for Global Development in Washington, DC. The report claimed that there was a lack of evidence about the effects of aid programmes, and that "[t]his absence of evidence is an urgent problem: it not only wastes money but denies poor people crucial support to improve their lives."[1] Across the international field of development aid, such questions continue to be raised, both by the public, policy makers, and researchers. But they are also raised internally, by aid staff and evaluation experts within the institutions of development aid. Indeed, the concern for results, effects, evidence, and learning in aid is constantly been discussed internally, and has been so for more than 40 years.

Our study is an exploration into these decades of hard work by evaluation managers aimed at answering the question of "does aid work?". Rather than trying to answer this question itself, we have sought to understand how aid organisations themselves have sought to answer it. More specifically, we have focused our study on *aid evaluation*. Indeed, numerous evaluations are always in progress across the field of development aid. The Swedish and Norwegian aid sectors – the two countries we focus on in this study – are constantly abuzz with planning processes, visiting consultants, circulation of drafts, and informal exchanges about the purpose, scope, and expectations of evaluation reports in the making. Yet if we take one step back from the vast and busy landscape of development aid, evaluation reports are curious objects indeed, and worthy of close attention in and of themselves. Who writes them, who reads them, who uses them – how, and for what purpose?

Clearly, evaluation reports are produced for a reason. We may think of them as *tools* that help us better see the effects of aid. Yet as tools, they are used by very different audiences and for very different purposes: External actors such as the public, the media, and NGOs use them to gain information about how aid funds are being used. Aid

[1] Center for Global Development 2006. *When Will We Ever Learn? Improving Lives Through Impact Evaluation.* Report of the Evaluation Gap Working Group. Quote from CGD's online presentation of the report: http://www.cgdev.org/publication/when-will-we-ever-learn-improving-lives-through-impact-evaluation (retreived January 2, 2017).

staff use them to gain insight into how things might be done differently. Policy-makers use them as input into policy decisions.

Within the field of aid evaluation, it is common to distinguish two basic purposes that evaluation is explicitly expected to fulfil: *accountability* and *learning*. These two are often referred to as "the dual purpose" or "the twin objective" of aid evaluation. Yet this conception of a dual purpose, we suggest in this report, conceals some inescapable dilemmas; even, perhaps, outright contradictions.

The problem

The dual purpose of accountability and learning is a well-established principle within aid evaluation, and has been so for several decades. Yet during the past 30 years, numerous reports and studies, from Sweden, Norway, and other key aid actors, have concluded that there is too little learning within development aid. Why, these studies ask, do aid organisations not learn more from their own experiences? And, more specifically, why do they not learn more from their own evaluations?[2]

While 'learning' is not always clearly conceptualised in the above-mentioned reports, it is often used, in practice, to mean *acquiring new knowledge that fosters change* – on programme, policy, or organisational level. Yet if a report finds that learning occurs at the individual staff level, this may be cast as problematic – i.e. that learning unfortunately "only" occurs at the individual level.[3] The main challenge is organisational learning: the literature on this topic is substantial, and has been a key part of the evaluation community's professional discussions during the past 30 years.[4] In this literature, most of the studies agree that evaluation is a key tool for enabling organisational

[2] Carlsson and Wohlgemuth 2001; ICAI 2014; Jones and Mendizaba 2010; Krohwinkel-Karlsson 2007, 2008; Norad 2016; Norwegian Ministry of Foreign Affairs 1993; Riksrevisionsverket 1988.

[3] Independent Commission for Aid Impact 2014: *How DFID Learns*. Quote from the policy brief: "DFID staff learns well as individuals. They are highly motivated and FID provides opportunities and resources for them to learn. DFID is not, yet, however, managing all the elements that contribute to how it learns as a single, integrated system." (http://icai.independent.gov.uk/report/dfid-learns/. Last retrieved 17.10.2016.)

[4] Berg 2000; Carlsson and Wohlgemuth 2001; Forss et.al 1994; Furubo 2003; Johnson 1991; Rist and Joyce 1995.

change and that learning has not been achieved unless it has caused some sort of change.

The question of how accountability relates to learning has been a topic of discussion for several decades, and with an increase in attention during the past years. Key actors within the OECD-DAC, the EU, and the World Bank have in recent publications explicitly discussed the relation between accountability and learning.[5] While this attests to the topic being on the table, it is at the same time remarkable how consistently evaluation practitioners and agencies conclude that the two purposes are indeed compatible. But we are not entirely convinced by this view and have designed this study to explore how the two purposes of aid evaluation relate in practice.

The question of why there is so little learning in development aid thus remains high on aid donors' agendas. But is the problem merely that they have not yet found the right approach? That the tool of aid evaluation is simply not used to its best potential? Or is the problem rather that the objective of learning is in itself compromised by its uneasy relationship to the other main purpose of aid evaluation, accountability? Might the problem lie not in how the tool of aid evaluation is being *used*, but in the tool *itself*? Perhaps the contradiction is located here; that we expect this tool to achieve too many things at the same time? Might there in fact be a direct trade-off between using evaluation for accountability and using it for learning? Put strongly: Does the concern for accountability in itself impede learning?

[5] For a short opinion piece, see blogpost by Caroline Heider (Director General of the World Bank's Independent Evaluation Group, IEG), March 22, 2016: "Facing Off: Accountability and Learning – the Next Big Dichotomy in Evaluation?." The blog comments on a Forum section in *Evaluation Connections* (the newsletter of the European Evaluation Society) of February 2016, titled "Forum: Is there a trade-off between accountability and learning in evaluation?" with four contributions including one by Heider. Accountability and learning is explicitly discussed in the reports "Evaluation for better results" (Asian Development Bank 2014, pp. 47-65), "Assessing the uptake of strategic evaluations in EU development cooperation (EuropeAid 2014, pp. 16-17, 39-40); and "Evaluation Systems in Development Cooperation", (OECD-DAC Evalnet 2016, pp. 23, 43-44). For a recent Norwegian contribution to the discussion of learning, see Norad 2016: *Kan lærdommer fornye utviklingspolitikken?* Evalueringsavdelingens årsrapport 2015/16.

Our hypothesis

In this study, we have reformulated the above question as a hypothesis: *The concern for accountability itself impedes learning; put strongly, the two are incompatible.* From this strong hypothesis we developed three separate "diagnoses" of where the problem might lie, organised around three levels of analysis.

First, the problem may lie in the *evaluation text:* Designed for multiple audiences, both internal and external, this document is expected to achieve the two largely contradictory goals of accountability and learning. By analysing the evaluation reports as pieces of text, we have asked: Might the problem of learning be solved by writing the reports differently? Second, the problem may be the *evaluation process:* The process of commissioning, conducting, and disseminating evaluation reports does not encourage the relevant audiences to use and learn from them. By analysing the evaluation process as an example of knowledge production, we have asked: Might the problem be solved by changing the way that evaluation processes are conducted? Third, the problem may be the broader *evaluation system*: Aid evaluation is but one element of the larger systems of so-called results-based management and performance reporting, which, while often recognising the importance of organisational learning, in practice appear to prioritise accountability as the primary concern. By analysing how evaluations are part of a political context, we have asked: Might the problem be solved by changing the expectations of what aid evaluation as such, and development aid more generally, may realistically achieve?

Our conclusion

Our analysis shows that at the project/programme level, the dual purpose may indeed be compatible, although practical challenges may clearly emerge also here. What is much harder to achieve, even possibly an unattainable goal, is what one of our informants dubbed "big learning": Learning on a more general level – in the headquarters, among policy makers, and in the wider public – about how and why development aid succeeds or fails, and what may done about it. Given that the relation between learning and accountability in practice differs at different levels and with different organisational arrangements, we have modified our initial hypothesis and articulated the following

conclusion to our study: *The dual purpose of accountability and learning in practice involve fundamental trade-offs*.

A key implication of our analysis is that aid evaluation as currently practised may not be the promising "all-purpose tool" that it is often expected to be. Increasingly sophisticated methods, regardless of their importance for the credibility and authority of aid evaluation, may in themselves contribute to deepen the increase the tensions between accountability and learning. Based on this conclusion, we argue that there is a strong need, both in the aid administrations and the wider public, to explicitly acknowledge that there are indeed important trade-offs between doing aid evaluation for external insight and control and for internal learning and change.

Our approach

This study is not a typical aid report. We have written the report as a conversation-starter, in an open style that we hope is engaging, thought-provoking, and constructive. We have limited the academic jargon and technical information, and made use of annexes for the benefit of those who want more details, including the theoretical framework and methodological design of the study. We hope our report may inspire our readers to further explore the rich literature on the topic of accountability and learning, and have used both footnotes and bibliographies to this end.

We have deliberately chosen *not* to operate with pre-defined theoretical definitions of the concepts of 'accountability' and 'learning' in this report. Both concepts are vague and broad with multiple meanings and a number of possible definitions. Our interest has been to explore how these concepts are used in practice by the actors themselves and in the documents we have analysed. Building on our findings in this report, we suggest a *definition of learning* that is practice-based: Learning entails actively aquiring new knowledge. It may thus happen when someone is actively partaking in an evaluation process. Presumably, the further removed one is from the practical, daily life of an evaluation process, the more difficult learning from evaluation becomes. This concept of learning enables a more open-ended interpretation which, rather than implying that learning is necessarily enabled by articulating 'lessons learned' and inducing concrete organisational changes, may be achieved through a

willingness to experiment, take risks, acknowledge failure, and adapt to one's circumstances. Similarly, based on our exploration of how the concept of accountability is used in practice, we suggest a *definition of accountability* that differentiates between the different sites of aid: Homewards to the donor countries and outwards to the beneficiaries. This ressonates with Ebrahim (2005), who suggests distinguishing between three forms of accountability: upwards (to donors), sideways (to peers), and downwards (to beneficiaries), and who argues that learning may be combined with the latter two forms of accountability, but not the first.[6]

The authors are all researchers working within Norwegian academic institutions. We bring an unusual mix of perspectives to this study: *Hilde Reinertsen* is a historian and researcher within the interdisciplinary field of Science and Technology Studies (STS). In her Ph.D. dissertation, she analysed how aid evaluation was established as a field of expertise within the Norwegian aid system during the 1980s, with a special interest in the role of documents and documentation practices in aid. *Kristian Bjørkdahl* is a rhetorical scholar who has published widely on science communication and rhetorical analyses, especially within the fields of environment and development. In his Ph.D. dissertation, he analysed the production and reception of a set of key historical texts within the field of ethics. *Desmond McNeill* is a political economist and senior professor with long-standing experience from the field of aid evaluation, both as a researcher and as a practitioner. Among his publications are the books *The Contradictions of Foreign Aid* and *Global Institutions and Development: Framing the World?*.[7]

In our analysis, we have combined our different analytical approaches of history, rhetoric, and political economy into what we hope is a refreshing analysis of aid evaluation.[8] We have made the study of evaluation texts our main priority: their historical

[6] Cf. Chapter 2 for further discussion of Ebrahim's definition of accountability.
[7] Reinertsen, H. 2016. *Optics of Evaluation. Making Norwegian Foreign Aid an Evaluable Object, 1980-1992.* Ph.D. dissertation, Faculty of Social Sciences, University of Oslo. Bjørkdahl, K. 2016. *Expanding the Ethnos. Rorty, Redescription, and the Rhetorical Labor of Moral Progress.* Ph.D. dissertation, Faculty of Humanities, University of Oslo. McNeill, D. 1981. *The Contradictions of Foreign Aid.* London: Croom Helm. Bøås, M. and D. McNeill (eds.) 2004. *Global Institutions and Development: Framing the World?* Routledge. All three authours have contributions in the forthcoming anthology (in Norwegian): Bjørkdahl, K. (ed.) 2017. *Rapporten. Sjanger og styringsverktøy.* Oslo: Pax.
[8] Cf. Appendix 1 for a more detailed description of our analytical approach.

development, their rhetorical properties, their processes of production and circulation, and the wider contexts of aid administration of which they are an integral part. Starting out with a strong hypothesis, we investigated three related sets of empirical sources: close reading of a selection of evaluation reports; interviews with past and present senior evaluation officials; and a mapping of the historical trajectory of aid evaluation in both countries. In addition, we conducted a literature review of international research literature and practice-based publications.

The comparison between Sweden and Norway enables an illuminating contrast, similar to the historical dimension, to the contemporary systems and practices of aid evaluation in the two countries. While being historically close collaborators within the field of development aid and also aid evaluation, the two countries have pursued often surprisingly different trajectories in practice. Given that local discussions of aid in the donor countries are often, perhaps paradoxically, domestically oriented, these differences are interesting to highlight in order to destabilise taken-for-granted ideas and practices of aid evaluation.

For our data collection, we started by going through the existing Swedish and Norwegian databases of aid publications to gain an overview of all existing evaluation reports in both countries.[9] The search also included other relevant documents, notably evaluation handbooks, manuals, annual reports and newsletters.[10] We then made a selection of 20 evaluation reports for close analysis based on the following criteria: historical breadth, thematic continuity, and diversity of form.[11] Thus, we wanted reports of different formats that covered similar topics across a wide time span in order to identify possible historical changes in the evaluation report genre. Based on this, we chose Swedish reports from the health sector and Norwegian reports on natural resources, notably energy and fisheries. The majority of reports were Swedish in order to reflect the analytical weight of the study. We also included two joint evaluation between

[9] For Swedish documents, we went through Sida's publication database and the online document archive at biståndsdebatten.se (http://www.sida.se/Svenska/Publikationer-och-bilder/publikationer/ and http://www.bistandsdebatten.se/dokumentarkivet/). For Norwegian documents, we went through Norad's evaluation database (https://www.norad.no/en/toolspublications/publications/evaluationreports/).
[10] Cf. appendix 2.1. for a list of evaluation manuals and guidelines.
[11] Cf. appendix 2.2 for a list of analysed evaluation reports.

Sweden and Norway in order to strengthen the diversity of reports and also accentuate publications from the most recent decade. While we do consider the selection to be sufficiently broad and systematic within the limited scope of this study, it is obviously not all-encompassing and a different sample might have given different results.

Our selection of interviewees was based on the same criteria of historical breadth and national comparison. We conducted in-depth interviews with seven highly experienced senior evaluation staff members (five in Sweden, two in Norway) who have held or currently hold key positions within Swedish and Norwegian aid evaluation. They have a combined experience ranging over 45 years, from 1971 to the present day. We used methods from oral history in order to explore the interviewees' individual professional trajectories, how these related to the wider changes in the evaluation field, and finally their reflections on the relationship between accountability and learning. We actively used our hypothesis to invite the interviewees to relate their own experience with the ongoing discussions in the literature. Here, we were interested in grasping evaluation *practice*, not general ideals and theories, and to identify potential practical dilemmas and contradictions.

While our sample enabled us to access the changing practices, tacit understandings, and informal processes of aid evaluation, it was unfortunately beyond the scope of this study to directly explore the perspectives of either the users of evaluation, partner organisations in developing countries, or the objects of evaluation (i.e. those being evaluated).

Structure of this report

The report is structured according to the three levels of our hypothesis: The evaluation text, the evaluation process, and the evaluation system. *Chapter 1* is this introduction, which presents the starting point, main arguments, and analytical approach of our study. *Chapter 2* unpacks the potential contradiction between learning and accountability by mapping what existing literature has to say about this question. *Chapter 3* deals with the evaluation report as such: We here distinguish its specific genre and analyse a selection of evaluation reports to identify how they in practice handle the dual purposes of

accountability and learning. *Chapter 4* expands the concept of evaluation to investigate the whole evaluation process and identifies a set of key dilemmas that make the dual purpose of accountability and learning difficult to reconcile. *Chapter 5* contextualises the evaluation reports and processes both within the wider evaluation system and historically, by showing how evaluation has always, in different ways, been part of broader political systems of planning, accounting, and results assessment. In *chapter 6,* we conclude our study and reflect upon the practical implications of our findings.

Accountability vs. learning: a spectrum of positions

Our hypothesis going into this study was the following: *The demand for accountability itself impedes learning; put strongly, the two are incompatible.* In contrast, evaluation practitioners often state the opposite – that accountability and learning are "two sides of the same coin". Yet between these two extremes, there are a number of more nuanced positions available. In this chapter, we review the existing literature ranging from aid agencies' own evaluation manuals via practice-based publications to academic research. In the latter category, we have also included a few contributions from other fields than development aid that explicitly discuss the relation between accountability and learning.[12]

Based on our review, we have distinguished four main positions which are most commonly held in discussions on the relation between accountability and learning – ranging across a spectrum (see Box 1 below). As the table states, the different positions map broadly onto the spectrum between practice-based and academic literature, ranging from internal manuals to independent research journals.

Box 1: Spectrum of positions held on the relationship between accountability and learning

No.	Position	Typically found in
1	Complementary objectives	Evaluation manuals, practitioners' publications
2	A reconcilable dilemma	Practice-based research publications
3	A problematic trade-off	Practice-based research publications
4	An irreconcilable contradiction	Independent/critical academic research

[12] Cf. Appendix 2.3 for a list of publications on aid evaluation, accountability, and learning. A full list of cited academic literature may be found in appendix 4.

Position 1: Complementary objectives

It is commonly stated among evaluation practitioners and in evaluation publications that the purpose of evaluation is both accountability and learning; indeed, that these are two faces of the same coin.[13] In Norway, this is explicit in the Evaluation Department's mandate: "On the one hand, evaluation activities should promote the transfer of experience, and on the other, they should hold Norwegian development policy actors accountable for the management of funds. (…) A key objective is to identify lessons learned in a systematic way, so that they can be used in policy development and as the basis for operational activities."[14] No potential contradiction between accountability and learning is here acknowledged, although it is stated that they may be given different priority in individual evaluation processes.[15]

In Sida's evaluation manual, a sub-chapter is devoted specifically to the relation between accountability and learning.[16] This is described mainly as a matter of asking different kinds of questions and gaining different kinds of answers, which in turn relates to different levels of analysis and use: "An evaluation that is meant to satisfy the requirement for accountability may of course raise very different questions than an evaluation intended for learning."[17] Furthermore, learning is considered to entail more substantial analysis than does accountability. This approach echoes a distinction often made elsewhere between assessing whether aid staff are "doing things right" or "doing the right things": The Sida handbook conceptualises this

[13] ADB 2014; OECD 2016.

[14] "Instruction for evaluation activities in Norway's aid administration", p. 1. Revised version, approved 23 November, 2015. The Norwegian version uses the term "learn from experiences" ("lære av erfaringer"), rather than "transfer of experience". Among the stated objectives of evaluation are "systematise lessons learned" and "improving results through effective learning processes". The Instruction furthermore states that the purpose of evaluation is to "document the effectiveness and relevance of efforts to realise the Norwegian development policy", hence connecting the evaluation efforts directly to the implementation of Norwegian policy.

[15] «The emphasis given to each of these aims may vary from one evaluation to the other."

[16] Sida 2007. *Looking Back, Moving Forward. Sida Evaluation Manual.* 2. revised edition (1. edition 2004).

[17] "In general terms, what an evaluation for accountability seeks to find out is whether the organisations that are responsible for the evaluated intervention have done as good a job as possible under the circumstances. (…) When the purpose of evaluation is learning, on the other hand, the study is expected to produce substantive ideas on how to improve the reviewed activity or similar activities. Although learning, in itself, may be regarded as valuable, its real importance lies in the translation of new knowledge into better practice." Sida 2007, pp. 14-15.

distinction by means of the concepts "summative" and "formative" evaluations, to describe, respectively, evaluations for accountability, that mainly describe what has already happened, and evaluations for learning, that may be used to make changes of a more substantial nature. This distinction in turn resonates with another twin concept often employed in the literature on organisational learning, between so-called "single-loop" and "double-loop" learning: The former describes the feeding of information back into the specific project; the latter describes learning on a more substantial level.[18]

While Sida's evaluation manual points to differences between accountability and learning, it does not emphasise these differences as in any way at odds with one another. To the contrary, it emphasises that many evaluation questions may be relevant for both purposes and that different audiences may use the same evaluation for different ends: "It is not unusual that an evaluation, used by those who are responsible for the evaluated activity for improvement and learning, serves a purpose of accountability in relation to principals and the general public."[19] The manual points to what it terms "process accountability", which, it claims, blends the two purposes of evaluation: While the evaluation manual distinguishes between financial and performance accountability, and holds that evaluation is concerned with the latter, it here suggests that when "results are difficult to measure – a common situation in development cooperation" – a process-oriented accountability assessment may be useful.[20]

This concern for combining accountability and learning may be seen also among other donors. Already in 2001, the OECD-DAC Working Party on Aid Evaluation (today commonly referred to as EvalNet) asserted the importance of attending to both, while also emphasising that the two purposes clearly had diverging implications for the evaluation process.[21] In 2010 and 2016, EvalNet undertook reviews of the DAC members' evaluation systems.[22] The 2016 review identified "attention to both accountability and learning" as one newly emerging trend in aid evaluation.[23]

---

[18] Forss et.al 1994; Rist and Joyce 1995.
[19] Sida 2007, p. 15.
[20] Sida 2007, p. 15.
[21] OECD 2001.
[22] OECD 2010, 2016.
[23] OECD 2016, p. 23.

Some of those holding this position also considers accountability and learning to be mutually supportive.[24] In recent debates, the term "accountability *for* learning" is introduced to connect the two. Several agencies, including DFID and major NGOs (among them Oxfam), are replacing the acronym M&E (monitoring and evaluation) with MEL (monitoring, evaluation and learning) to describe the integration of learning into the regular performance measurement and evaluation efforts.[25]

Position 2: A reconcilable dilemma

Those holding this position do acknowledge that there may exist tensions between accountability and learning, but argue that it is possible to reconcile the two.[26] For example, a study of learning from evaluation of the EU's development cooperation agency (EuropeAid) highlighted the combining of accountability and learning as one of five "thorny dilemmas or challenges" that must be addressed in order to enhance uptake of evaluations: "In theory, there should be no contradiction between the two main objectives (…). They are, in principle, two faces of the same coin (…) The evidence collected shows, however, that this virtuous circle often does not occur".[27] This position is also held by Manning and White, who, in a discussion of so-called impact evaluations acknowledge that there may be tensions between accountability and learning that may give unwanted negative effects; yet they conclude that "performance measurement systems that use impact evaluation can make a serious contribution to both accountability and, in particular, decision-making."[28]

One notable contribution seeking to reconcile the two objectives in practice is made by Reeger et.al, who in a research article from 2016 state that: "Although evaluators are increasingly asked to facilitate and support learning, [the] call for accountability remains and, despite best efforts, often gains priority – hence the need to find ways to reconcile

[24] For example, in OECD 2016: "Accountability and learning are not mutually exclusive, rather they feed into each other, i.e. a learning culture improves the performance of development assistance, and ensures that organisations are held accountable" (p. 23).
[25] Cf. Grey et.al 2014 for a discussion of the concept "accountability for learning".
[26] Cracknell 1996; Lehtonen 2005; Manning and White 2014; Reeger et.al 2016.
[27] Bossuyt et.al 2014, p. 39.
[28] Manning and White 2014, p. 348.

the two."[29] Reeger et.al seek to enable this by making room for different forms of accountability, which in turn may accommodate the combination of accountability with learning. Building on the work of Alnoor Ebrahim, they distinguish between three forms of accountability: *Upwards* (towards donors), *downwards* (towards recipients), and *sideways* (towards other actors involved in the project). The potential for combination, they argue, lies in the downwards and sideways forms of accountability.[30] Furthermore, the authors separate between "goal-oriented evaluation, which is usually connected with accountability purposes" and "learning-oriented evaluation". In order to enable the latter, they argue that the evaluation methodologies must be adapted accordingly: "Thus, in order for evaluation methodologies to support learning, they should be participatory (...) and responsive (…) to the learning needs of evaluation stakeholders".[31]

Position 3: A problematic trade-off

This position holds that accountability and learning are not possible to combine without some negative effects. More specifically, the accountability concern comes at the expense of learning. The former UK Independent Advisory Committee on Development Impact (IACDI) has noted, "there is always a tension between the use of evaluation for accountability and its use for lesson-learning".[32] A recent evaluation by the evaluation department of the World Bank Group (the Independent Evaluation Group, IEG), that studied the Bank's systems for self-evaluation, takes the same perspective and concludes that there are indeed trade-offs between accountability and learning: "The systems' focus on accountability and corporate reporting – generating ratings that can be aggregated in scorecards and so on – drives the shape, scope, timing, and content of reporting, and limits the usefulness of the exercise for learning."[33]

---

[29] Reeger et.al 2016, p. 7. This article does not analyse evaluation of development aid as such, but it is still (and even: precisely for this reason) of very high interest to our study.
[30] Ebrahim 2005.
[31] Reeger et.al 2016, p. 10-11.
[32] IACDI 2010, p. 3. IACDI was disbanded in 2011 and replaced by a new agency, the Independent Commission on Aid Impact (ICAI), which reports directly to the International Development Committee of the UK Parliament.
[33] Independent Evalution Group 2016. *Behind the Mirror. A Report on Self-Evaluation Systems of the World Bank Group*. Washington DC: World Bank Group. Quoted from IEG's

Development scholar Des Gasper argues: "automatic choice of an audit form of accountability as the priority in evaluations can be at the expense of evaluation as learning".[34] Reeger et.al, while themselves concluding otherwise (see above), summarise this position nicely: "While today both accountability and learning are considered important motives for programme or project evaluation, the literature shows that it is not self-evident that evaluation focuses on both motives at the same time. Different scholars suggest tensions or trade-offs exist between accountability and learning as reasons for and results of evaluation. (…) These apparent tensions between accountability and learning pose challenges to evaluators."[35]

Basil Cracknell made this point already in 1996, arguing that the two purposes involve diverging methodologies, especially with regard to involving stakeholders, and that this divergence was not diminishing, rather widening.[36] Hence, in contrast to Sida's evaluation manual and Reeger et.al, who both note that the two purposes require different methodologies, both Cracknell and Gasper take the point one step further by arguing that the differences not only involving making different choices, but that these differences may also have problematic effects.

Position 4: An irreconcilable contradiction

According to authors holding this position, the trade-offs between accountability and learning are so substantial that the two objectives must be considered contradictory and not reconcilable in practice.

---

blog: "Learning from Evaluation: How can we Stay at the Top of the Game?" by Caroline Heider and Rasmus Heltberg, August 2, 2016. http://ieg.worldbankgroup.org/blog/learning-evaluation-how-can-we-stay-top-game (Last retrieved 18.10.16.)

[34] Gasper 2000, p. 17.

[35] Reeger et.al 2016, p. 7.

[36] "[A]id ministries are under greater pressure than other ministries to give priority in their evaluation work to the accountability objective rather than the lesson-learning objective. But this creates problems, because the evaluation approach needed for accountability (for example, random sampling; fair cross-sectional representation; the use of totally independent evaluators from outside the agency) is completely different from the approach needed for lesson-learning (for example, deliberate selection of projects with problems, or deemed of particular interest; use of own staff to ensure that the learning process stays in-house). (…) But as lesson-learning is the main objective, there is an increasing realization among donors that the only way to monitor and evaluate the impact of people-centred (that is, socially oriented as distinct from technological) projects is to involve stakeholders themselves in the process." Cracknell 1996, pp. 23-24.

Armitage writes, "there is an unresolved tension between accountability and learning dimensions of development evaluation which may be irreconcilable. This may explain the current trend towards performance-based models which markedly emphasize monitoring at the expense of learning".[37] Similarly, Serrat argues that "the two basic objectives of evaluations – accountability and learning – are generally incompatible."[38] He bases his conclusion on a study of the purposes of aid evaluation, and notes that the calls for accountability in the donor countries often make learning a secondary concern.

Eyben adopts the same perspective when she asks: "Why do donor governments have problems with learning and how can they help themselves do better?" Starting from the conclusions offered by Carlsson and Wohlgemuth on why learning is difficult,[39] she expands an argument by Curtis to criticise "a mind-set that seeks control through linear planning, supported by the instruments of performance management".[40] A key feature of this position is its critical stance towards the concept of accountability and to the wider systems of performance management and results-based management in general.[41] Indeed, in her piece, Eyben proceeds to argue that results-based management, which was being introduced in the UN and other international agencies during 2005, "may have paradoxical effects: First, it may distort or weaken recipients' accountability to their own citizens or intended end-users (…). Second, it may constrain transformative learning". Her conclusion is that this constitutes not merely a trade-off, but rather a contradiction: Analysing the UN Millennium Project Report from this position, she suggests "that donors have overemphasised target orientation to the detriment of relationships" and states, "*It* emphasises the need for more strategies and coherent planning; *I* respond that this is like recommending brandy as a cure for a hangover".[42]

---

[37] Armitage 2011. "Evaluating aid: An adolescent field of practice", *Evaluation* 17(3): 274.
[38] Serrat 2010. *Learning from evaluation*. Washington, DC: Asian Development Bank, p. 3. This quote is also used in a discussion of the relation between accountability and learning in a Norad evaluation report: *Use of Evaluations in the Norwegian Cooperation System.* Report 8/2012, p. 42-43.
[39] Carlsson and Wohlgemuth 2001.
[40] Eyben 2005: 98.
[41] For a more detailed review of the critical discussions of results-based management, see EBA report 2016:07: *Towards an Alternative Development Management Paradigm?*
[42] Eyben 2005, pp. 98 and 102-103. (Italics in the original)

Eyben's polemical point succinctly summarises her position: The wider management systems of which evaluation is a part in themselves make learning difficult. Hence, in order to remedy the lack of learning, which all authors agree is a problem, it is necessary to look beyond the issue of learning in itself and consider how learning fits into the broader systems of results-based management.

This fourth position builds on a growing academic literature that investigates the spread of accounting and auditing practices and rationales into new sectors of society. Here, Michael Power's concept of the "audit society" is a central reference.[43] Power argues that during the 1990s, Western European public administrations, with the UK as his case in point, changed their *modus operandi* from direct to indirect management of public services. This has had a number of practical and institutional effects that in turn have changed the relation between the state and its citizens: civil servants have shifted from providing social services to verifying that the services are delivered as expected, which in turn causes the imperative to "never trust, always check" and the subsequent build-up of agencies and systems for ensuring "control of control", i.e. internal units tasked with internal audits and reviews. This has caused a proliferation of documentation efforts, described by Power as an "audit explosion". [44] This in turn relates to a wider discussion about the context within which evaluation takes place, to be addressed in Chapter 5.[45]

Several scholars point to how ill-designed systems of accountability may have adverse effects on responsibility. Richard Rottenburg, in his analysis of the German Development Bank's project management routines in the 1990s, makes this precise point: With reference to Power, Rottenburg describes how aid agencies, by using the Logical Framework Approach, design aid projects in ways that make them responsible only for ensuring the execution of the project, but not for the project's actual results. He thus identifies an inherent tension between ensuring accountability to the donors and to the recipients.[46] Elinor Ostrom et.al, in a report written for Sida, makes a similar point,

[43] Power 1994, 1997.
[44] Power 1997, pp. 1-2.
[45] Cf. for example Carol Weiss, "evaluation is a rational practice that takes place in a political context" (1993, 94). Sida's Evaluation Manual makes a similar point: "it is important to keep in mind that there is a politics of evaluation, and that evaluation is often something else or something more than just a useful tool for social engineering." Sida 2007, p. 16.
[46] Rottenburg 2000. For a more elaborate discussion of Rottenburg's position, see Reinertsen 2016.

yet with reference to the political dimension of development aid, in noting that the current "tangle of relationships" within development aid in effect involves a fragmentation of responsibility.[47]

Chapter conclusions

We have identified a spectrum of different positions with regard to the relationship between accountability and learning. A key distinction between the positions lies in what is understood by accountability and the role of results-based management: Is results-based management in itself a problem or a necessary precondition for learning? The contrast between these two positions is most marked when comparing Eyben's stance with that of Norad's Evaluation Department, which considers one reason for the lack of learning to be a lack of sufficient programme documentation.[48] Sida's evaluation manual positions itself closer to the latter, although not quite so marked. This reflects Ray Rist's distinction from 1991 between two kinds of aid evaluation, placing Norway within an accountability tradition and Sweden within a tradition of organisational change.[49]

In short, our review shows that there is a recognition by many, both practitioners and academics, that the two objectives of accountability and learning in practice take one onto two diverging paths. As Basil Cracknell and several others have pointed out; the two objectives entail diferent methodologies, which in turn have important consequences for how the evaluation process is conducted, the content of the evaluation report, and how it may be used.

This last point will be of crucial importance as we now turn to our empirical analysis: What are the concrete, practical effects of the fact

[47] Ostrom et.al 2002: "The result of this tangle of relationships is that many individuals are responsible for ensuring the effectiveness and sustainability of aid, but no one is really responsible". *Aid, Incentives, and Sustainability: An Institutional Analysis of Development Cooperation*, Sida Studies in Evaluation 02/01.

[48] In its Annual Report for 2015 (in Norwegian), published in May 2016, the Evaluation Department states that: "The foundation of learning and improvement is laid in the specific program and initiative. It must be more clear what one wishes to achieve and what one actually does achieve must be better documented." (Our translation from Norwegian: "Grunnlaget for læring og forbedring legges i det enkelte program og initiativ. Det må bli tydeligere hva man ønsker å oppnå og hva man faktisk oppnår må dokumenteres bedre." (pp. 8-9). Available at https://www.norad.no/globalassets/publikasjoner-2016/evalueringsavdelingens-arsrapport-2015-16.pdf (last retrieved 04.01.2017).

[49] Johnson, Paul 1991. "Ray Rist Talks about the IIAS Working Group on Policy and Program Evaluation", in *Evaluation Practice* 12 (1): 45-53.

that aid evaluation is asked to satisfy these two different objectives? How may we see the tension or possibly contradiction play out in practice; and how is this tension sought reconciled by staff in their daily work? These questions will guide our analysis in the coming chapters, where we investigate Swedish and Norwegian aid evaluation in three ways: First evaluation reports, then evaluation processes, and finally the wider evaluation context. We begin by taking a fresh look at a set of evaluation reports to get an understanding of how they handle the dual purpose of accountability and learning.

Inside the evaluation reports

What is an evaluation report? This may seem a trivial question. Yet if we stop and think about it, the question is of fundamental importance. During the past 40 years, there has been a proliferation of such reports in development aid.[50] Why? What are their purpose and effects? In order to understand how evaluation may contribute to accountability and learning, it is necessary to take a close look at the reports themselves.

Evaluation reports have multiple purposes and there are high expectations to both their form and function. They are intended to shed light upon the life and effects of an aid programme. They should make it possible for actors far away to easily assess the programme, objectively and clearly. There is also a core concern that the evaluation reports should be *used*. Evaluation reports are expected not only to document and discuss the effects *of* aid, but also, in turn, to have their own effects *upon* aid. They are hence expected to feed back into the aid system. Indeed, no report is written in a vacuum; they are the products of practical evaluation processes and enter into the wider systems of aid evaluation and management. These key dimensions will be the topics of the next two chapters. First, in this chapter, we will critically examine reports as *texts*.[51] What do these texts tell us, how, and what are their potential implications?

In this chapter, we will present what we find to be the key characteristics of aid evaluation reports, based on our own previous research in combination with an in-depth analysis of 20 Swedish and Norwegian evaluation reports.[52] We will especially attend to the following features of the reports: their *genre*, including structure, format, and layout, and their *argument*, with special attention to the articulation of recommendations and lessons learned.

---

[50] Karlsson 2012, Reinertsen 2016.

[51] For studies analysing evaluation reports as texts, see Karlsson 2012, Stirrat 2000, Amba 1998, Moretti 2015, Gasper et.al. 2013, Reinertsen 2016, Winther 2016. See also the anthology: Bjørkdahl, K. (ed.), 2017a (forthcoming), in which reports are analysed both as a specific genre and as a management tool, notably the chapters Bjørdahl 2017b, 2017c; McNeill 2017, Reinertsen 2017.

[52] See Chapter 1 and Appendix 1 for discussions on methodology, including the selection of reports. For a detailed discussion of rhetorical methodology, see Bjørkdahl 2016.

40 years of evaluation reports

In both Sweden and Norway, the aid administration has produced evaluation reports for more than four decades. Sida established its first evaluation office in 1971, Norad in 1977.[53] Since then, both agencies have published evaluation reports every year in addition to several other kinds of reports and publications (cf. Box 2 below and Box 3 later).[54]

In Sweden, the majority of these reports are so-called "decentralized evaluations" – that is, evaluation reports commissioned by other actors than the central evaluation staff, such as programme officers in sectoral departments in the headquarters, in the embassies, and on-site in project management positions. In total, Sida published 50 evaluation reports in 2015. We have no similar number for Norway, since decentralised evaluations are not published in the same series as the centrally commissioned reports (of which there were nine in 2015). It is furthermore difficult to gain a symmetrical historical overview of the number of decentralised evaluations in the two countries since this term is fairly new, replacing the term "project reviews". In addition, we must expect that not all reports are included in the online databases. In both countries, the central evaluation units have been concerned with the quality of these decentralised evaluations: Norad's Evaluation Department published a report on the quality of decentralised evaluations in February 2017.[55] Sida's Evaluation Director is currently taking steps to improve the quality of decentralised evaluations within Sida, as his predecessors also did during previous decades.[56]

It is important to note the diversity of evaluation sub-genres with which evaluation practitioners operate. These range from evaluations

[53] Cf. appendix 3 for a short historical overview of how aid evaluation has been institutionalised in Sweden and Norway during these decades.
[54] In addition to the evaluation reports commissioned by Sida and Norad themselves, either by the central units or (in Sweden) by programme officers, come those commissioned by partner organisations, both in Norway/Sweden and in the recipient countries. We have delinated our study to reports published by Sida and Norad, thus leaving out those produced closer to the recipients.
[55] Norad 2017. For a presentation (in Norwegian), cf. https://www.norad.no/evaluering/planlagte-evalueringer/pagaende-evalueringer/evaluering-av-kvaliteten-pa-gjennomganger-og-evalueringer-i-norsk-bistandsforvaltning/ (retrieved 02.01.2017).
[56] See chapter 4 for a more detailed account of the current efforts and appendix 3 for a brief overview of past efforts. Cf. also the report by Forss et.al from 2008, titled *Are Sida evaluations good enough?*, that was commissioned by Sida's central evaluation unit (UTV).

limited to one single aid intervention to comprehensive studies of interventions within the same programme, sector, topic, or country, occasionally also involving several donors agencies (see Box 2 for a brief description of the most common different evaluation types).

Box 2: The diversity of evaluation reports

Type of evaluation	Description
Project evaluation	Covers one specific aid intervention. If commissioned by the project's own staff, it would previously be termed "project review" to distinguish it from evaluations commissioned by the evaluation unit in the headquarters. Project evaluations are increasingly termed "decentralized evaluations" to distinguish them from centrally commissioned reports.
Programme evaluation	Covers one distinct aid programme, commonly consisting of a set of related aid interventions. Distinctions were previously made between "programme review" and "programme evaluation" to distinguish between reports commissioned locally and centrally.
Thematic evaluation	Covers a larger number of aid interventions within one specific topic or sector, commonly across several countries.
Country evaluation	Covers the aid portfolio of one specific recipient country.
Real-time evaluation	Comprehensive process initiated by the central evaluation unit during the planning stage of a major new aid programme. Commonly includes comprehensive baseline studies, midterm reports, and final reports.
Strategic evaluation	Developed by the central evaluation unit, commonly more oriented towards key donor concerns. May be a thematic, country, or meta-evaluation.
Joint evaluation	Collaborative effort by two or more donors.

The report genre: Structure, style, and line of argument

The structure of evaluation reports has remained surprisingly consistent for the duration of the 40-year period, although they certainly *look* very different now. As one informant noted, the reports do indeed look more professional, yet the content may not necessarily be very different. The increasing methodological sophistication and development of sub-genres clearly demonstrates how evaluation has become an established field of expertise. Yet our analysis of the evaluation report as a genre – meaning the combination of structure, layout, format, style, and line of argument – also shows that the reports' core features remain virtually the same through the decades.[57]

In both Sweden and Norway, handbooks and guidelines for aid evaluation provide templates (of varying degree of detail) for how to structure an evaluation report.[58] These templates describe what a report should include – what elements it should contain, of what length, and in what order. In sum, the templates seek to create what we may call *the ideal evaluation report.* Ideally, an evaluation report will combine methodological strength and practical usability. Indeed, Norad's *Handbook of Evaluation Questions* asserted already in 1981 what has remained a distinguishing feature of this genre: "An evaluation report – no matter how good it might be – has little value unless it is being used";[59] the ideal report should therefore not only build a strong *analysis* and give clear *conclusions*, but also offer practical *recommendations* and more general *lessons learned*.

The dual concern for methodological rigour and practical use brings its own contradictions. As one of our informants pointed out: "The reports must be short in order to be read, but long in order to be trusted." The solution is often to include extensive annexes with more detailed descriptions and analysis of the data and methods, often published as separate documents. The possibility to publish and

[57] Miller 1984, Yates 1990, Devitt 2008. Cf. Reinertsen 2016, pp. 216-224 for a more detailed analysis of aid evaluation report as a textual genre.
[58] For Sweden: Sida manuals 1974, 1985, 2004, and 2007. For Norway: Norad manual 1981; MFA manual 1992; Norad guideline 2016. See Reinertsen 2016 for a detailed analysis of Norad's 1981 and 1992 handbooks. For both countries, the OECD-DAC Evaluation Norms and Standards have served as important frames of reference since 1991. Notably, the five DAC evaluation criteria (relevance, effectiveness, efficiency, impact and sustainability) have been widely influential.
[59] Norad 1981. *Håndbok for evalueringsspørsmål*, p. 30 (authors' translation from Norwegian).

circulate electronic versions of the reports makes this even more pertinent today. Furthermore, reports are expected to accommodate busy readers by including an executive summary, in effect enabling readers *not* to read the full report. Having a convincing conclusion, precise recommendations, and a clear executive summary remains an important characteristic of good evaluation reports. Ensuring the quality of these three elements remains a key concern of evaluation managers, according to our informants. Indeed, as one informant put it: "We know that reports are not read. So it's the summary of findings and recommendations that is important rather than the whole report."

In the following, we will discuss these issues in more detail. We begin by analysing how our selected reports build their analysis, craft their conclusions, and identify lessons learned. We then discuss how key features of the reports may be traced to the so-called Terms of Reference, i.e. the document in which the evaluation managers who commission the evaluation describe their expectations of the evaluation process and the subsequent report.

Crafting arguments and recommendations

In the classical rhetorical tradition, there are three basic rhetorical genres: the *forensic*, *epideictic*, and *deliberative*.[60] In short, what distinguishes each is their purpose, which in turn has implications for their composition, main concerns, and potential effects: The forensic genre was used for *judicial purposes* (in court). It deals with the past, tries to answer "what happened", establish causes and effects, and ultimately place blame. The epideictic genre was used for *ceremonial purposes* (e.g. funerals and festivities): It is oriented towards the present moment, gives elaborate descriptions of its object, and offers praise or blame, but with no immediate purpose beyond the ritual and its meaning for those present. The deliberative genre was used for *political purposes*, and directed towards the future. It makes the case for a future course of action, presents alternatives and argues for one of them, trying to answer the question "what ought we to do?"

While the evaluation report is obviously a modern genre, this classical three-part distinction serves to highlight how the different purposes have important practical implications for the composition of

[60] Cf. appendix 1 for a more detailed presentation of rhetorical analysis.

the text, as well as for its ostensible effects on (or use value for) readers. This in turn is relevant for our discussion of the relation between accountability and learning. For a report to attend to both in a satisfactory way is arguably possible in theory, but in practice this is a challenging task indeed.

In general, the evaluation reports attend well to the rhetorical purpose of establishing "what happened", discerning causes and effects, giving praise, and placing blame. This is typically achieved by telling a formalised story of a practical aid effort and its consequences. But the purpose of accountability is, we would argue, poorly balanced with and weakly adjoined to the deliberative function, i.e. the question "what ought we to do". We have observed varying degrees of this mismatch in the great majority in our sample of reports.

In order to illustrate this point, it is necessary to look more closely at the details of specific reports. A Sida report from 2007, titled "Healthy Support?", may serve as a key example.[61] In this report, the authors spend the main part of the text documenting how Sida's support to the health sector in Angola came about, how the motivations and the means of the effort changed over time, what the outcomes were and why, and how these outcomes in turn affected the effort. The authors note that the long, drawn-out war in Angola hampered most of the Swedish effort, and in the section on "lessons learned" refer repeatedly to the war as a main causal factor of what must in most ways be deemed a failure. The wider implications of this finding, however, are hardly dealt with. Of a total of 140 pages, the authors spend only a half page responding to the question, "Is improved health possible in war and absolute poverty?". Their main point is that "Health depends on a broad spectrum of social, economic and cultural factors, which have made Angola's health indicators some of the worst in the world" and that "Angola's catastrophic health situation cannot be seen as an isolated problem".[62] While obviously true, this points to the difficulty of isolating the aid effort itself, which is arguably a prerequisite for arriving at recommendations that might lead to change and improvement. If one cannot *isolate* the effects of one aid intervention, one cannot really *evaluate* the effects of that intervention.[63] Using the Angola effort again as an example, the only

[61] Sida 2007. *Healthy Support?* Sida Studies in Evaluation 07/50.
[62] *Healthy Support*, p.105.
[63] For a detailed analysis and discussion of this point, cf. Reinertsen 2016.

apparent lesson appears to be that the entire "broad spectrum of social, economic and cultural factors", in short the whole of Angolan society, must change before one can hope to succeed with aid efforts for health, and consequently that the answer to the authors' own question, "is improved health possible in war and absolute poverty?", is "no". The authors do not come to this conclusion, however, but state simply that the war has been the main cause of failure.

If the Angola report is an example of a report that does not (fully) develop the implications of its own description and diagnosis, the by now old, yet still well-known Norwegian report on the Lake Turkana Fisheries project is an example of a report with a clear mismatch between description and recommendations.[64] This report too concludes that the project has been in large part a failure, but its (many) recommendations follow only very loosely (if at all) from the descriptions it makes. While the bulk of the recommendations have to do with concrete circumstances of the operation – notably how the project is organised, and how the relations between the aid officers and local staff are set up – it becomes apparent in the report's descriptive part that the main reasons for the effort's failure are most likely to be found elsewhere: on the one hand, in the natural phenomenon of drought (Lake Turkana had been drying up), on the other hand, in endemic, counterproductive practices of the recipient culture (corruption, systemic gender imbalances, etc.).[65]

We have found examples of the same problem also in recent reports. This historical continuity suggests that the report genre and its writing process has changed less than many would perhaps acknowledge. Among more recent examples is the comprehensive 2007 report that evaluates Norway's sector portfolio within so-called "power-related assistance".[66] Using Nepal and Mozambique as case countries, the report assesses both cases to be relative successes though it also acknowledges that certain objectives have not been met. Summing up the reasons for these failures, the evaluation identifies three main causes: First, civil wars that have plagued both countries. Second, institutional changes within the power sectors stood in the way of effective resource use, including capacity development. Third, there was a lack of capacity and political will to implement strategies

[64] DUH 1985. *Lake Turkana Fisheries Development Project.*
[65] DUH 1985, pp. 7-17.
[66] Norad 2007. *Evaluation of Norwegian Power-Related Assistance*.

and plans. Elsewhere, however, the report singles out corruption as an "overriding concern," noting that the partner countries are "considered to suffer from severe corruption problems" (p. 8). Although this is probably descriptively true, and also seemingly a relevant analytical cause for failure, we again find a rather poor link from description and analysis to recommendations. To counteract corruption, they advise "implementation of existing anti-corruption measures" which includes a strengthening of monitoring. However, it is not clear how this recommendation can avoid tripping on what the evaluation has already identified as a major cause for failure, namely the lack of capacity and political will. Put simply: Given that they have already highlighted a lack of will to implement existing measures as a central problem, how can "implement existing measures" be a realistic recommendation? The evaluation's elaboration of this recommendation ("better business ethics need to be fostered") is not very helpful. Indeed, how does one effectively "foster better ethics"? The passive form of the sentence only further obscures the basic question of "who is going to do what".

The mismatch we have found between description/analysis and recommendations is apparent also in other reports about rather successful projects, like the ones that evaluate the Vietnam-Sweden Health Cooperation. In a report from 2001, for instance, the authors note the high degree of aid dependency in Vietnam, and state that this "remains a concern".[67] At the same time, one of their recommendations is to "*Sustain* the process and basic structure" of the project, as "Sida has a role to play and is a wanted partner".[68] Here, the recommendation not only does not follow from the diagnosis given, but appears quite simply to contradict it. Similarly, in a new report five years after, the high degree of aid dependency in Vietnam is underlined once more, and again, the recommendation is that "Sweden ought to stay in the health sector in Vietnam". To justify this apparently contradictory recommendation, the authors refer to the fact that, "To leave the sector would be a lost opportunity to leverage Sweden's comparative advantage as a long-time trusted partner in health".[69] The problem here is not that the report proposes the wrong solution – there might well be good reasons to continue an intervention even when aid dependency is a problem. Our point is

[67] Sida Evaluation 01/03, p. 2.
[68] Sida Evaluation 01/03, p. 3.
[69] Sida Studies in Evaluation 06/02, p. 7.

simply that the recommendation does not follow logically from the preceding analysis, as the report neither recognises the potential contradiction nor justifies its own position.

These are only a few examples of how the different functions of the evaluation reports – mainly the forensic (description and analysis) and the deliberative (recommendations) – are disconnected. In the texts, the balance between description/analysis and recommendations is such that the first is heavily prioritised. Furthermore, the recommendations are often not well connected to the foregoing description and analysis. Hence, the recommendations often appear as little more than add-ons. As we have seen, the mismatch may range from a failure to develop obviously important points, to an inadequate basis for recommendations, to recommendations that seem to actively contradict the description.[70]

Many of these forms of mismatch appear to stem from an unwillingness to factor in circumstances that would seem to challenge either the institution of aid itself or concrete projects or efforts more specifically. For instance, in the case of the Vietnam reports, the recommendation to sustain the effort in the face of high aid dependency was, as we have seen, motivated not by a desire to learn – change and improve – but by a strategic consideration concerning Sida's or Sweden's position in the international aid landscape. These reports thus evidence a variety of evasion strategies that seek to avoid recommendations that the reports themselves make seem natural and sensible. The apparent consequence of this move is a certain conservatism; instead of making recommendations that incorporate the entire range of relevant factors, they make recommendations that fit roughly into the frame of development aid as it is currently practiced. That does not necessarily mean that development aid *in general* is conservative, i.e. opposed to learning, only that the aid evaluation reports are probably not what supplies the field with an impetus for change.

When the reports in this way become less interesting – or actually less useful – as instruments of learning, the *ceremonial* function emerges as more prevalent than we would first expect. In their ceremonial function, the reports respond primarily to the call for

[70] Stirrat (2000) makes a similar point, arguing that there is often a missing link between analysis and recommendations in development evaluations.

accountability, in that they establish legitimacy and transparency. Simply by describing what has happened, and by putting these descriptions into (increasingly formalised) formats, the evaluation system has made a contribution to justifying itself. However, their specific answers are not necessarily useful in decision making. This reflects the argument made by German evaluation researcher C. Schwartz, who describes "evaluation as modern ritual".[71]

In sum, our analysis has found that recommendations are often not well founded in rigorous analysis. This, we suggest, may seriously hamper learning. Furthermore, given that few reports factor in context, they make their recommendations less relevant.

Identifying "lessons learned"

In addition to offering specific recommendations for the evaluated aid programme, evaluation reports are also most often expected to identify so-called "lessons learned". As described in Sida's evaluation manual, the point of this feature is to highlight what may be learnt from the evaluation and be of use "elsewhere".[72] These sections are asked to elaborate how findings from a specific evaluation can be transferred to other contexts, and thus made into a more general insight. Identifying this may clearly be a difficult task: What qualifies as a lesson? Who may be said to have learnt "it"? When is a conclusion local, specific, or context-dependent, and when is it generalisable, that is, potentially relevant for actors beyond the specific programme? While these questions clearly are not easy to answer, they are important for understanding the function of the "lessons learned" section.

In our sample of reports, few include sections explicitly articulating "lessons learned". Our analysis here may therefore not be as conclusive as for the recommendations, which were provided by

[71] See Schwartz, C. 2006 for an elaboration of this point. Drawing parallels between evaluation and audit, Schwartz quotes M. Strathern: "Like a ritual, audit tries to persuade participants of the way the world is without acknowledging its own particular perspective." (Strathern 2000, p. 287, quoted in Schwartz 2006, p. 254.)

[72] Sida's evaluation manual, p. 29. Annex B "Format for Sida Evaluation Reports" states (p. 102): "Lessons learned are findings and conclusions that can be generalised beyond the evaluated intervention. In formulating lessons, the evaluators are expected to examine the intervention in a wider perspective and put it in relation to current ideas about good and bad practice."

nearly all the reports. One exception is the Sida report *Health cooperation at the crossroads* from 2006, which includes a short chapter titled "Lessons learned". The report explains the purpose of the chapter by stating that it "brings out conclusions with a broader application than the [programme] itself." The report states in the passive voice that "the following lessons have been learnt", of which three concern the process and two the policy context. Of these, one directly addresses learning itself: "By not utilising lessons learned from pilot models, opportunities for informing policy were lost". [73] Here, use of the passive voice and the linear relationship between lessons and decisions suggest a one-directional view of learning, as though the articulation of lessons learned would automatically also mean to learn it. But if we consider the actual content of this particular "lesson", we might see how the failure of the organisation to learn from reports appears quite endemic, and also, how the function of accountability seems to be crowding out the function of learning. Indeed, the lesson (supposedly) learned in this case is, in effect, that the lessons that someone were supposed to have learnt from the pilot had *not* in fact been learnt.

While our sample is too small to conclude more firmly about the function of 'lessons learned' sections, it is relevant to point out that the wider system of aid evaluation, as we will further explicate in chapter 5, depends on these for synthesising learning on a broader level. An attempt to synthesise, and actually use, "lessons learned" runs, however, into the same problem as the one we have identified for recommendations: The writing down of lessons learned may risk becoming a sort of ritual, in which one respects the demand to *document* "lessons learned" but not the ambition to actually *learn* them.

On the basis of our sample, we thus claim that the evaluation reports mainly concentrate on the question of "what happened", offer weak answers to the question of "what we ought to do", downplay the importance of context, and to only a little extent contribute to the purpose of learning. Why is this so? In order to answer this question, it is necessary to study the document that precedes every evaluation report: the so-called "Terms of Reference" (ToR).

[73] Sida Studies in Evaluation 06/02, p. 36.

The influential "Terms of Reference"

The document titled "Terms of Reference" (ToR) defines the evaluation assignment. It is written by those commissioning the evaluation and is normally included as an appendix to the evaluation report. The document details what is to be the object of the evaluation; the questions to be answered; the evaluation criteria to prioritise; the format of the evaluation report; the scope, timeline, and budget of the assignment; and the required professional profile of the evaluation team. In addition, the ToR sometimes also explicitly outlines how the report itself should be structured in order to attend to the evaluation questions.[74]

In our sample, the Terms of Reference are very specific, and tend to delineate the object of the evaluation very narrowly. The consequence of this, we will argue, is that the reports in effect contribute to producing accountability while oftentimes getting in the way of learning. There are three reasons for this: First, the ToRs specify in much greater detail how a report should document "what happened" than they specify how analysis and recommendations should be developed. This is directly reflected in the attention devoted to each in the corresponding reports. Second, the problem stems from what several reports indeed themselves assert, namely that it is "impossible to isolate the effects" of an aid programme.[75] Yet most of the ToRs do not acknowledge this problem and give little room for including contextual and other factors. Third, and related to this, the ToRs give little room for incorporating uncertainty, risk, and contingency. As a consequence, the evaluation reports may often identify weighty contextual and contingent factors, yet are not given the opportunity to factor them into the evaluation.[76]

There is one notable exception to this general tendency: The Terms of Reference document for the report *Supporting Child Rights* (a major joint evaluation commissioned by Sida and Norad in 2010) explicitly calls for a "thorough analysis of contextual factors" and for designing a learning-oriented process. While first affirming the dual purpose of evaluation (both "to summarise results in order to account for the efforts and resources invested" and "moreover to contribute to the

[74] According to Stirrat (2000), this is a common feature of Terms of Reference documents.
[75] *Healthy support*, p. 13.
[76] Cf. Stirrat 2000 and Reinertsen 2016 for more detailed discussions of this point.

continuous learning and development of policies, strategies and methods"), this ToR document assert that "organisational learning presupposing a participatory evaluation process" is one of two "corner stones for the conduct of this evaluation". The ToR therefore demand the bidding teams to demonstrate their capacity for designing such learning-oriented evaluation processes, from facilitating workshops with involved stakeholders (also including children themselves) at several stages of the evaluation process, including the articulation of recommendations.[77] In this way, the ToR envision the evaluation process itself to contribute to realising the strategy it is designed to assess, thus emphasising the internal learning process over the external accountability function. Yet as we will show in the next chapter, which explores the evaluation process, this explicit ambition proved difficult to realise in practice.

Chapter conclusions

Both development aid and aid evaluation create great expectations. These expectations rest heavily on the evaluation reports themselves. Indeed, reports are arguably the primary way of showing those who finance aid – in different capacities – what has been done, and what the consequences of the effort have been. Reports thus serve a very central function in ensuring the accountability of the aid system. But as we have argued, this rhetorical work does not necessarily contribute to learning, and in the way it is apparently practised, often contradicts this ambition. In the reports themselves, there is a systematic deprioritising of the learning function, while the attempts to contribute to learning – through sections with "recommendations" and "lessons learned" – often fail to factor in relevant context. In principle, there is no inherent contradiction here; a text may serve both forensic and deliberative functions. But for this combination to be a success, the transition from *is* to *ought* – from description to

[77] *Supporting Child Rights*, p. 236-237, 243-244 (Appendix 7: Terms of Reference). «The objective is to enhance the sharing of experiences (including preliminary evaluation findings) between consultants and staff as well as between staff within the organisations. Creating opportunities for reflection on the organisations' own practices as well as results of those practices found in the evaluation, is an important aspect of deepening the understanding of support for children's rights among stakeholders. The sharing of experience is furthermore key to formulating relevant recommendations. Recommendations are, after discussion, to be formulated jointly by consultants and stakeholders».

recommendation – needs to be far more firmly established than what is the case in our sample of reports.

In practical terms, we believe this leads to a choice between two alternatives: Either, aid evaluation reports should admit that what they do best is, as it were, to *document*, to describe what has happened. In this case, ToRs should not ask for recommendations at all, because one would have to come to the realisation that recommendations are best developed elsewhere, in other (perhaps even non-textual) aid and aid evaluation practices. In other words, the first option would be to take seriously the idea that we have been asking too much of the evaluation reports and decide to cultivate the function they already serve well, namely description and accountability. The other option is completely opposite, i.e. to consolidate the ambition that evaluation reports should be tools to serve the function of accountability *as well as* that of learning. This would have major implications for the ToR and the report itself. One way to solve the problem of weakly founded recommendations might be for the ToR not to ask for recommendations. In other words, evaluation should be seen to a greater extent as an *explorative* undertaking, where one might make findings that lead to controversial recommendations, recommendations that create discussion, dialogue, and an impetus for change.

Inside the evaluation processes

The writing of an evaluation report involves a comprehensive process and numerous different actors. As we showed in the previous chapter, the so-called "Terms of Reference" document frames the subsequent report in important ways by describing the evaluation assignment in detail. This means that the initial work done by evaluation staff or programme officers to define the evaluation assignment is critical for what the evaluation report ultimately says – and how it may be used. This is just one of several such aspects of the evaluation process to which we will attend in this chapter.

At multiple points during the evaluation process, aid staff make practical decisions regarding the balance between accountability and learning – whether explicitly or not. While seeking to ensure a proper distance between the evaluation team and the evaluated programme, they at the same time seek to ensure that the evaluation report will be as useful as possible. As we will show, this concern for neutrality (in the service of accountability) may often be at odds with the concern for utility and organisational learning.

A well-established field of expertise

Our informants assert that there have been substantial changes to the evaluation process, and especially to the report-writing process. One suggested calling this "professionalisation"; another hesitated using this term, saying that the process has always been professional, but the content and what is expected have changed. All agreed that there has been a *formalisation* of the process, notably because of the tender-process by which consultants are commissioned for evaluation assignments.[78] Furthermore, there has been a *standardisation* which means that the evaluation methods and processes are increasingly similar across countries, notably through the use of internationally agreed-upon principles and guidelines developed by the OECD-DAC.[79] Finally, this joint cooperation through the DAC and other

[78] As public agencies, Sida and Norad are obliged to adhere to EU regulations of public procurement. Both countries in addition have national directives for disbursement of public funds.

[79] OECD 2010: *DAC Quality Standards for Development Evaluation*. OECD 1991: *DAC Principles for evaluation of development assistance.* This document is still widely cited. Less

international expert arenas has encouraged increasing *coordination* between evaluation units across the aid agencies. Aid evaluation is today a well-established international field, constituting what we may consider a distinct "epistemic community",[80] that includes a large number of consultancy firms taking assignments in multiple countries, several long-standing research journals, strong international networks (including OECD-DAC's EvalNet and the Nordic evaluation network), international training programmes, and global, regional and national associations of evaluation.

Another historical transformation is related to the major and rapid developments in computer and internet technologies, which involves the ability to write, edit, and circulate large documents and to build databases for storing and retrieving these documents online. Indeed, new evaluation reports are now commonly published only in digital form. In sum, this has dramatically increased the velocity of (parts of) the evaluation process. Given that the circulation of texts may now be instantaneous, it may have wider reach and grant far easier access. At the same time, the sheer increase in the volume of evaluation documents available also makes it harder to gain an overview of the field and creates challenges for evaluation units to synthesise the many different findings. This, in turn, has potentially negative implications for learning from evaluation.

The critical early stage

Both Sida and Norad have formalised systems for how to plan an evaluation, write the Terms of Reference document, and procure an evaluation team.[81] In practice, already at this point key premises are established for whether an evaluation will contribute most to accountability or to learning. Chosing either accountability or learning as the main concern for an evaluation has a number of practical implications: the two prompt different questions and different methods. As several of our informants explained (and which is also

prominent, but of great historical importance, is the *Methods and Procedures in Aid Evaluation* (OECD 1986).

[80] Haas, Peter M. 1992. "Introduction: Epistemic Communities and International Policy Coordination", *International Organization* 46 (1) 1-35.

[81] This is described in Norad's guideline for evaluations (2016) and Sida's evaluation manual (2007).

confirmed by our analysis of evaluation reports in chapter 3): Accountability is concerned with finding out *what* has happened and whether this is consistent with the initial programme plans and budgets; while learning has more to do with *why* things happened as they did. Accountability would therefore require a more formalised and standardised process, more yes/no-questions and the use of checklists. Learning, on the other hand, would entail asking more open questions, notably "how" and "why", and lead to a more inclusive process.[82]

The specific way in which these initial questions are articulated is important, given that they will be a key part of the Terms of Reference document, which determines the further process and the final report in direct ways. These factors in turn have important repercussions throughout the evaluation process and for the makeup of the ultimate report. The Terms of Reference document is, as one informant put it, "completely decisive": It requires that staff think through and articulate what needs to be known. The very wording of the ToR frames what the consultants are expected to deliver, given that the relation between the agency commissioning the evaluation and the team taking on the assignment is formalised through a legally binding contract.

In practice, the work that goes into writing the Terms of Reference documents differs widely, depending on one's position in the aid system. If the evaluation is prepared in the central evaluation units, then much care and experience goes into this process. When the evaluation is prepared by a programme officer in a field office, a sector department, or a partner organisation (what is commonly called a "decentralised evaluation"), one may not expect staff to have the same methodological expertise. Oftentimes, according to one informant, programme officers ask colleagues who have previously managed evaluation processes to share a past Terms of Reference document, ideally one that was particularly good, upon which they may model the one they are going to write.[83] This may have a large impact upon what and how one evaluates, and therefore also what one may learn: If the Terms of Reference document was developed for a programme

[82] These differences are also discussed in the evaluation literature (cf. Chapter 2), especially by Cracknell (1996), who describes an expanding divergence between the methods used for accountability purposes and for learning purposes.
[83] Cf. Reinertsen 2016, chapter 4 for a detailed analysis of a similar process, in which a ToR document was modelled upon another ToR from a similar project but in a different country.

situated within a different context and with a different history, how may it be expected to handle the distinct properties and challenges of the programme currently under evaluation? Indeed, to train programme staff in writing good Terms of Reference documents, especially with regard to articulating evaluation questions, is currently (and has also previously been) a key priority of the Sida's central evaluation unit.

While the preparatory work of an evaluation process is based on the formalised procedure of public procurement, there are also important informal aspects to this process. These informal aspects are, according to many of our informants, critical for ensuring that the evaluation will be useful and contribute to learning. As one informant noted, you need distance (institutionalised through the tender process and the external consultants) to ensure credibility, yet you need local grounding to make it useful. Several informants highlighted this point, and emphasised the need to build support for an evaluation from the very start as a way to increase the prospects of it being used once it was finished. This meant to spend time involving the relevant actors and build engagement for the *evaluation idea* itself, even before the ToR was written.

In explaining the need for informal grounding of the evaluation idea and the further evaluation process, most of our informants referred to Michael Patton's approach of "utilization-focused evaluation", which precisely emphasises that in order for an evaluation to be used, there must be an explicit need for it; if there is not, then the evaluation should not be done in the first place.[84] Indeed, the one explicitly learning-oriented evaluation in our sample (the major joint evaluation *Supporting Child Rights* from 2011) builds directly on Patton's approach. As shown in chapter 3, the Terms of Reference document for this evaluation explicitly called the bidding teams to design a learning-oriented participatory process. In the final synthesis report, the team describes its methodology, contrasts the utilisation-focused approach to "conventional evaluation practice [that] takes an arm's length posture to the evaluation object and the stakeholders involved in order to buttress independence and impartiality. […] The former has tended to be divorced from the users to the extent that the findings are compiled in unread reports. The latter, on the other hand, is more likely to create ownership of the evaluation process and

[84] Patton 1984, 2008, 2015.

findings among the stakeholders because they have been actively involved." While this ToR explicitly stated an ambition to involve not only aid staff but also the end beneficiaries, including children, the team in practice experienced time constraints and unforeseen events that restricted stakeholder involvement mainly to the pre-defined primary users of aid staff in Norway, Sweden, the embassies, and relevant country authorities.[85]

The informal work to build and maintain internal engagement for specific evaluation processes may indeed run against the well-established principle of organisational distance, as formulated in the OECD-DAC's Quality Standards for Development Evaluation: "The evaluation process is transparent and independent from programme management and policy-making, to enhance credibility."[86] One way to formalise internal participation in the evaluation process, which is often used for more comprehensive evaluation assignments, is to establish a reference group with representatives both from internal units and external actors. This enables the evaluation managers and the evaluation team to involve stakeholders at multiple stages of the evaluation process, from the drafting of ToRs to commenting on draft inception reports and draft final reports. This, several of our informants maintained, would help secure the evaluation's relevance and utility and also contribute to learning. Yet the more informal efforts at embedding the evaluation in the organisation may clearly run counter to the principle of arms-length distance between evaluators and evaluated. Indeed, efforts at maintaining internal engagement for the evaluation, which potentially enhances learning and use, may ultimately reduce external trust in the evaluation process. This, we would argue, demonstrates a critical trade-off between accountability and learning: both purposes clearly have strong merits, yet choosing one over the other has practical implications that may make either internal or external actors distrustful.[87]

---

[85] *Supporting Child Rights,* p. 35-36.
[86] *DAC Quality Standards for Development Evaluation,* section 1.2, p. 6.
[87] Schwartz & Struhkamp (2007) argue that evaluation may both build and destroy trust. In their two case studies of evaluation processes within German universities, "ambiguity replaced transparency, confusion replaced systematicness and suspicion replaced legitimation." 2007, p. 336.

Balancing internal and external concerns

Working with external consultants remains a key challenge for evaluation managers, both in centralised and decentralised evaluation processes. It involves multiple practical dilemmas for how to handle the relation between internal and external concerns: Maintaining this boundary is necessary for the credibility and legitimacy of aid evaluation, yet if the distance is made too wide, it may be to the detriment of the evaluation's quality, utility, and potential for learning. The central evaluation units in both Sweden and Norway have developed different way of handling this challenge. In Sweden, the central unit holds framework agreements with a limited set of consultancy firms (currently three) who then compete for the individual evaluation assignments. In Norway, the central unit has increasingly taken on a larger part of the evaluation work after the unit's mandate was amended in 2015.

According to our informants, several problems commonly emerge from using external consultants. First, consultants may have a limited understanding of the context (on both the donor and the recipient sides), and must therefore either spend disproportionately much of their time getting to know the relevant specifics of what they are evaluating or risk making misleading analyses and offering potentially inappropriate recommendations. Second, constraints caused by the evaluation assignment's available budget and calendar time mean that the consultants normally have very little time available, which makes it all the more difficult to build trust among their informants and to do in-depth analyses. These are both good reasons for the active participation and assistance of evaluation managers and project officers in retrieving project documentation and granting access to informants. Indeed, there is a case to be made for going even further and involving those responsible for the development intervention directly, by requiring that they prepare a self-evaluation as a part of the process. Clearly this will not be neutral and objective, but it can be efficient in terms of obtaining factual information, as well as giving the evaluators very valuable insights. Self-evaluation is quite common among UN agencies, such as IFAD and ILO, and indeed the World Bank. It has been recommended also for DFID: "The need to give priority to enhanced self-evaluation is highlighted by the NAO

(National Audit Office) opinion survey".[88] What we are referring to here is more radical: the inclusion of a self-evaluation as an integrated part also of larger independent evaluations.

How much should evaluation managers be involved in the practical evaluation process? This dilemma was raised by all our informants, and was a concern at all stages of the evaluation process and both at the central and decentralised levels. All our informants had experienced one or more of the following problems: The evaluation team might be weak on evaluation methods; they might submit reports of low quality; or offer recommendations with little practical value. As one informant stated with reference to the report writing process: "It is hard to strike the right balance – you want to ensure they write a good report, but without encroaching upon their independence." While the evaluation team is responsible for the content (including the decision on what methodology to use), the evaluation manager is supposed to assure the report's quality at several stages in the processes: when assessing the team's bid for the tender; when negotiating the contract; and by critically reviewing (and approving) first the inception report, then drafts of the full report, and ultimately the final report ready for publishing.

The specifics of how evaluation reports should be written were of special concern to our informants. One key aspect that also has a strong bearing on the reports' potential contribution to learning was the articulation of recommendations. One of our informants noted that evaluators often exaggerated their role as external *critics*, posing their critique in an unpedagogical manner that made people defensive, hence losing opportunities for learning. Another stressed that the main challenge was the articulation of recommendations: too often, the recommendations were too specific, too general, or too ambitious, which all made them of little practical value. Hence, adjusting the consultants' expectations of their own role, and also the role of the evaluation report in the subsequent follow-up stage, might be a necessary part of evaluation managers' work.

[88] Picciotto, R. 2008. "Evaluation independence at DFID: An independent assessment prepared for IACDI", p. 10. Cf. also *Evaluation manual* (2015) from the Independent Office of Evaluation of IFAD (International Fund for Agricultural Development); "Evaluation policy" of the International Labour Organization (ILO); and a self-evaluation (2011) by the World Bank's Independent Evaluation Group (IEG).

A major question in the evaluation process regards the role of the recipients and beneficiaries of the evaluated aid programme. Is the evaluation supposed to be *about* them or *with* them? Or indeed, should it rather be *by* them? Most often, according to our informants, the end beneficiaries were not included in the evaluation process.[89] "We evaluate for ourselves", as one informant frankly stated. Other informants explained that in decentralised evaluations, aid recipients would more often be included as stakeholders. The relation to aid recipients, notably partner organisations, might also be complicated by the evaluation process, as it might be difficult for evaluation managers to explain that the evaluation process was not only about accountability, but also about learning. As one informant noted, partner organisations would often expect an evaluation process to involve auditing and external scrutiny, and this would make them anxious. Indeed, one informant who had also worked as an evaluator had experienced first hand how such distrust and lack of cooperation made the evaluation task far more difficult. The evaluator's style and approach, several informants asserted, was therefore of key importance. As one stated: "The evaluators cannot simply march in and ask their questions, they must create a good process." To facilitate participation, engagement, and willingness to learn is difficult if people experience (or even only anticipate) to being checked by someone from the outside.[90] Downplaying the significance of the evaluation report itself might then be a way to ensure cooperation, and thereby learning. The timing of a decentralised evaluation, which is often directly related to a decision-making process that involves the continuation or end of the programme, is one important reason why partners will often anticipate the evaluation team to function as auditors. If an evaluation process came *after* the decision was made, one informant suggested, then trust, openness, and learning would be easier to achieve. Yet this was disputed by another informant, who noted that "this would likely reduce the utility and therefore the value of the evaluation".

Finally, several informants questioned the very premise that external consultants in and of themselves ensure independence and integrity. Given the system of public procurement, where consultancy firms and research agencies compete for bids, the consultants are

[89] The same point is made by former World Bank economist Michael Bamberger (1991).
[90] This point parallels the argument made in much of the evaluation literature, cf. Chapter 2.

dependent on securing bids to secure future activities. While we have not studied this point empirically, the implication of this situation might be that consultants are wary of prejudicing good relations with the commissioning agencies, which may in turn affect their integrity and willingness to voice criticism of the evaluation units' own systems and routines.

Who learns from an evaluation process?

The principle of independence and distance in aid evaluation have several practical implications for learning. First of all: *Who learns?* The question of who learns, we will suggest, is directly related to the question: *Who writes?* Our informants agreed that those who learn the most from an evaluation process are the evaluation team members who conduct the work; evaluation managers who prepare the Terms of Reference and follow the process; and finally those who work on the programmes being evaluated and use evaluation as direct input in their decision-making. This, we suggest, may be tremed "sideways learning".[91] Being close to daily aid operations, the evaluation process, or preferably both, thus expands both the potential and experience of learning. Whether evaluations contributed to what one informant dubbed "big learning", meaning beyond the programme level in the organisation and society at large, was much harder to say, our informants thought.

The fact that the practical work of analysis and writing evaluation reports most often is done by actors outside the aid agencies themselves clearly serves the accountability purpose of evaluation. Yet this also means that substantial learning *remains* on the outside. Arguably, the most effective way of learning something new is through the active work of articulating questions, searching for answers, analysing data, discussing with others, and putting findings and conclusions into writing. Furthermore, given that evaluators have responsibility only for fulfilling the Terms of Reference for that individual evaluation assignment, they have no responsibility for ensuring that the report contributes to learning. This is entirely the responsibility of the aid agency and the Ministry. Yet unless they have been actively involved in the evaluation porcess, these users receive

[91] This builds on Ebrahims differentiation between upwards, sideways, and downwards accountability. See Ebrahim 2005 and Reeger et.al 2016.

only the final report – the finished text. This makes the document itself, down to its specific wording, extremely important: This is all that is left after the Terms of Reference have been completed and the evaluation process is over. Hence, one practical prerequisite for learning would be to ensure that the Terms of Reference explicitly ask learning-oriented questions and require extended stakeholder participation throughout the process. Still, as we discussed above, this is not sufficient if the concern for learning turns out not to be possible within the given budgetary and calendar constraints.

Another key challenge to learning is the sheer amount of people involved in an evaluation process: the evaluation managers, the team, the programme officers, the partner organisations, and other stakeholders both on the donor and recipient side. If one follows the approach of utilisation-focused evaluation, involving relevant actors and maintaining their interest is an important part of the work. This is further complicated by the constant circulation of staff within both the aid agencies and the Foreign Service: Oftentimes, one informant lamented, the officer responsible for a specific programme or sector upon the start of an evaluation cycle had left for a new position before the evaluation is finished. Finally, the sheer number of consultants involved is also very high; according to a recent study of Swedish aid evaluation, consultants normally partake in but a few evaluations for the same agency, even in a situation where Sida has a framework agreement with consultancy agencies precisely to ensure continuity.[92] In sum, these features lead to a lack of continuity and a fragmentation of knowledge production, which in turn poses major challenges to learning at both the individual and institutional level.

Finally, a critical question is the following: *Who reads*? Exploring this question empirically is beyond the scope of this study, but among our informants a main experience was that few have the time to read evaluation reports and absorb their content. This was especially pertinent for ministry staff who receive the centrally produced evaluation reports. As one informant exclaimed in frustration: "The Ministry's absorption capacity is zero!" No other informants used such strong words, but some asserted that while there were indeed systems in place for enabling such absorption, few had time to use

[92] EBA 2015: *Utvärdering av svenskt bistånd. En kartläggning.* EBA Report 2015:02. Sida has a framework agreement with three consultancy firms. Norad has no similar framework agreement for evaluation; on the contrary, there has been an increase in the amount of different firms granted evaluation assignments, as the tenders are advertised internationally.

them in practice.[93] The feeling of there being too little time is clearly related to how the use of staff resources are planned and prioritised.

The notion of "absorption" is indicative of the aid evaluation system resting on a one-directional definition of how evaluation reports may contribute to learning: A report is submitted to the agency/ministry which is then expected to "absorb" its main message and transform it into action. Clearly, as we have shown in this chapter, the practitioners themselves, as represented by our informants, are acutely aware of the shortcomings of this model and work hard to remedy them through informal contact with colleagues and by actively managing the evaluation consultants. Yet even these efforts do not ensure that the reports are used.

What does it take for a report to be used?

For evaluation managers, ensuring that the evaluation is used is critical. In both Sida and Norad, the evaluation units have commissioned evaluation reports specifically on this topic to establish whether and how evaluation reports are being used.[94] In decentralised evaluations, the users are clearly defined as the programme officers or donor representatives who are going to make decisions about the future of that specific evaluation aid programme. Most of Sida's evaluations are of this kind, while the centrally produced reports normally have Sida's top management as their main recipient. For Norad's Evaluation Department, the Ministry of Foreign Affairs is of key importance, as this is the main recipient of centrally produced evaluation reports.[95]

Our informants highlighted especially four aspects that in combination determine the degree of use (of both central and decentralised reports): Quality, credibility, timing, and the relevance of the recommendations.

Quality refers to the report itself, and is especially concerned the methodological rigour of the report. The quality was thus closely

[93] For further discussion and references on this point, cf. Cohen, E. 2013. *Evaluation and Learning in Rule of Law Assistance.* Research Report, Folke Bernadotte Academy.
[94] Norad 2012, Sida 2009.
[95] The Ministry of Climate and Environment is the recipient for reports key parts of Norway's climate-related aid portfolio, notably REDD+.

related to the *credibility* of the report itself, and by extension also the consultants and the evaluation managers. For evaluation managers, maintaining credibility is of especially high importance. If the intended users do not trust the staff or the report, then they are less likely to use it. Hence, it is of key importance for evaluation staff to ensure that no-one doubts the credibility and quality of their work. This is ensured by working thoroughly, systematically, and transparently. *Timing* is critical for ensuring that the report arrives at a time when it is in fact needed: This entails that there must be a need for the evaluation, and it must arrive at the point when a decision is to be made about the future of the specific programme. As one informant noted; "good timing is more important than a perfect report". Finally, the *recommendations* must be concrete and relevant. Yet this is, according to our informants, also the most challenging part of the evaluation report (see chapter 3 for an elaboration of this point). Indeed, "[w]riting recommendations is an art".[96]

In decentralised evaluations, staff sometimes experienced a need to downplay the importance of the evaluation report itself in order to build trust among partners and thereby increase participation and learning. In contrast, when the main intended users are decision-makers in the Foreign Ministry, the report itself becomes of utmost importance: This is the concrete outcome of the evaluation process that will at the handed over to the ministry in public. While relevant ministry staff may often have been involved as stakeholders during the process, the top-level follow-up process and potential public display makes this process much different than in decentralised evaluations. The evaluation managers' role thus continues to be crucial. As one informant noted, their job is to translate the findings of the evaluation report into something that policy makers may be able to use. We will return this point in Chapter 5.

Public communication of evaluation reports

In addition to the direct internal use of evaluations on the project, programme, and policy levels, the evaluation units in both Sweden and Norway make the evaluation reports publicly available in different

[96] *Humanitarian Action: Improving Monitoring to Enhance Accountability and Learning, Meta-evaluation,* ALNAP Annual Review 2003, Section M, p. 171.

ways: from including them in open databases to writing shorter documents (in Sida called "Evaluation Briefs") aimed at wider audiences, to organising public seminars in which new evaluation reports are presented and discussed. The wider communication of evaluation reports after they are finished thus spurs a new set of documents for new sets of readers. These public displays are often covered by aid-related publications in the two countries,[97] and also, occasionally, by national media. In Norway, the Evaluation Department has in recent years taken a more active, independent role in the public debate, by writing opinion pieces in national newspapers and partaking in live broadcast news debates.

While the outwards-oriented communication of evaluation findings clearly opens the aid administration to external scrutiny and as such helps to hold aid managers accountable to the public, it also highlights a potential contradiction between accountability and learning. Public debates about a published report may increase public criticism. This may force the aid administration to defend themselves in public, which in turn may make them defensive and avoid debates about their own sector. As such, the communication of evaluation findings is also critical for the potential use of and learning from evaluations. Less formalised fora that rather encourage informal exchanges and trust-building may better contribute to learning, while they clearly also compromise the public's insight in their domestic aid administration. Hence, the choice of communication and dissemination strategy in itself involves trade-offs between accountability and learning.

A question of quality?

Before we conclude this chapter, we will highlight one key issue which was raised by all our informants: the question of quality. Indeed, as shown in this chapter, our informants, when responding to our questions or to examples we introduce in this study, often pointed to the reports' quality as a key concern. Furthermore, several use the concept of quality to explain why certain reports or processes become successful or insignificant. If they did not fulfil the OECD-DAC's "Quality Standards for Development Evaluation", they were "not best practice", and hence less interesting to discuss. The quality argument

[97] Notably the monthly magazine *Bistandsaktuelt* in Norway, the website biståndsdebatten.se in Sweden, and the Nordic journal *Development Today*.

might be used against Terms of Reference documents, external consultants, evaluation reports, and evaluation processes at large. Conversely, a common response to our hypothesis of there being a contradiction between accountability and learning was that the dual purpose could be achieved through ensuring higher quality and better methodology, what one informant summarised as "methodological rigour".

Indeed, our conclusions in chapter 3 might be interpreted in the same direction: that the documents are simply not good enough and that the solution is to write better reports. Yet this would be to conclude too hastily. We will rather suggest that the continuous experience of difficult evaluation processes and low-quality reports points to an existing – perhaps even widening – gap between evaluation theory and practice, between ideal and reality. While evaluation guidelines and standards articulate clear and transparent protocols, the everyday life of aid evaluation involves highly complex and complicated processes, involving a number of different actors with diverging interests. The final text that is normally produced through this process is expected to serve multiple purposes for multiple audiences who may often have conflicting concerns. The ambition on behalf of aid evaluation may well be unrealistically high, both among the public, who expect to hear the objective verdict about an aid intervention's success or failure, and among evaluation practitioners, who continue to develop new methods and systems of aid evaluation.

Chapter conclusions

In this chapter, we have asked: Who learns from evaluations? This involves two very practical questions: Who writes? Who reads? It is necessary, we will argue, to be more realistic about the value of evaluation reports: What does one want to achieve by employing this tool? In short, the report makes visible to outsiders what happens inside an aid programme. This is, of course, of great democratic value, and it is necessary for maintaining public trust in aid. Yet this external scrutiny may come at the cost of building trust between donors, recipients, and evaluators, creating internal engagement, and thus fostering learning. As our informants highlighted, learning happens in the active *process* of doing the evaluation and during its subsequent follow-up, but not necessarily through receiving the finished *document* as such. We should thus conceive of "evaluation" as a verb rather than

a noun: it is the practical process and the hard work this involves, rather than the end product, that is important for learning.

The ambitions on behalf of aid evaluations may well be unrealistically high. Perhaps, we suggest, the experience of unclarity and lack of overview is not a result of low quality, but rather precisely what may be expected. Managing good evaluation processes would then be more concerned with finding ways to handle this situation rather than designing increasingly sophisticated methods for seeking to escape it.

Yet this line of argument highlights a deep dilemma of aid evaluation with unavoidable political implications: Why is it so important to keep producing reports that so few people read and use? Why is the concern for independent, un-biased evaluation more important than enabling internal learning processes? These questions will guide our discussions in the next chapter, where we move inside the evaluation systems and see evaluation more directly in relation to the existing mechanisms of accountability and performance management.

Inside the evaluation systems

Aid evaluation is always but one part of a larger context, what one of our informants aptly called "a power field of diverging concerns and interests". In this field, the concerns of accountability and learning meet the concerns of foreign policy and diplomacy; the ideology and principles of aid policy; and domestic public concerns. While we in the previous chapters investigated first evaluation reports in themselves and then the evaluation processes through which the reports are produced, we will in this chapter contextualise evaluation and explore the broader systems of results management and knowledge production in development aid of which evaluation is a distinct, but also integral part.

Aid is a risky business. And the question 'does aid work?' has never been – and can never be – satisfactorily answered. Those who criticise it can find ample evidence of failure; those who support it can cite numerous examples of success. In contrast to the situation described four decades ago in *The Politics of Foreign Aid*,[98] aid is today very much in the public eye, and is the subject of political contestation in both Sweden and Norway. But working within the aid administration differs clearly from other sectors "most strikingly in that the people for whose benefit they are supposed to work are not the same as those from whom their revenues are obtained".[99] This renders aid more vulnerable than other sectors since it does not respond to the needs of any particular section of the electorate (though there are domestic interest groups involved, such as NGOs and consultants). In this situation, aid evaluation reports take on special significance, serving not merely as a resource to help bureaucrats learn and design better projects and programmes, but also as a major basis for an aid agency's claim for political support.

Institutional set-up of aid evaluation

Both historically and today, Sweden and Norway have institutionalised their evaluation function in different ways. Figures 1

[98] White, J. 1974.
[99] Seabright, P. 2002. "Conflicts of objectives and task allocation in aid agencies", p. 34. In Martens, B. et.al. (eds.), *The Institutional Economics of Foreign Aid*. Cambridge University Press.

and 2 show the current institutional set-up of aid evaluation in the two countries. Two main differences between the two countries is the relation between central and decentralised evaluations and the specific Swedish experience with of the establishment of external evaluation agencies. These different arrangements have direct implications for the relative weight given to accountability and learning within the two evaluation systems. Indeed, evaluation scholars already in the 1990s placed the two countries in two different traditions of evaluation, arguing that in Norway, evaluation was more closely connected with the purpose of accountability, while it in Sweden was a more internally oriented exercise concerned with organisational change.[100]

In Sweden, Sida's system for both aid management and aid evaluation has always been largely decentralised, while the size and autonomy of the central evaluation unit has shifted during different decades. This has entailed much independence for programme officers in designing both aid programmes and evaluations. Both the current unit and its predecessors has made methodological support to decentralised evaluations one of their key tasks. Depending on their size and mandate, Sida's central evaluation unit has also previously engaged in other key forms of evaluation work (cf. Appendix 3), but the bulk of evaluation reports have been managed by programme officers across the organisation. In contrast, Norway's aid evaluation has remained a centralised activity with a sharp distinction between centrally managed evaluations and locally managed reviews. Only recently have Norad's evaluation department started using the term "decentralised evaluations" about reports commissioned by programme staff. The central unit's main concern has thus been to facilitate comprehensive evaluations of whole sectors, countries, themes, and initiatives within the Norwegian aid portfolio.

In both countries, the central evaluation unit has undergone multiple major reorganisations.[101] In Norway, evaluation was first established as a distinct department directly under Norad's Director General in 1977. The evaluation function was moved into the (no longer existing) Ministry of Development Cooperation in 1984, where it ceased to be a distinct department, and then further integrated into the Ministry of Foreign Affairs in 1990. In 2004, the evaluation

[100] Cf. Cracknell 1996, Rist 1991 (in Johnson 1991).
[101] Cf. Appendix 3 for a short historical overview of the changes in the two countries' evaluation systems.

function was returned to Norad, where it again became a distinct department and gained more resources and a distinct mandate and now reports directly to the Ministries' Secretary Generals. These changes in the location of aid evaluation reflect the key tension in aid evaluation between distance and independence on one side, and relevance and use on the other: Being separated from power and practice may give more autonomy, while being closer may give more influence. Yet in the Norwegian experience, being closer to the responsible Minister did not automatically translate into more attention, resources, and influence. The political attention still depended on the Minister taking a specific interest in strengthening aid evaluation.

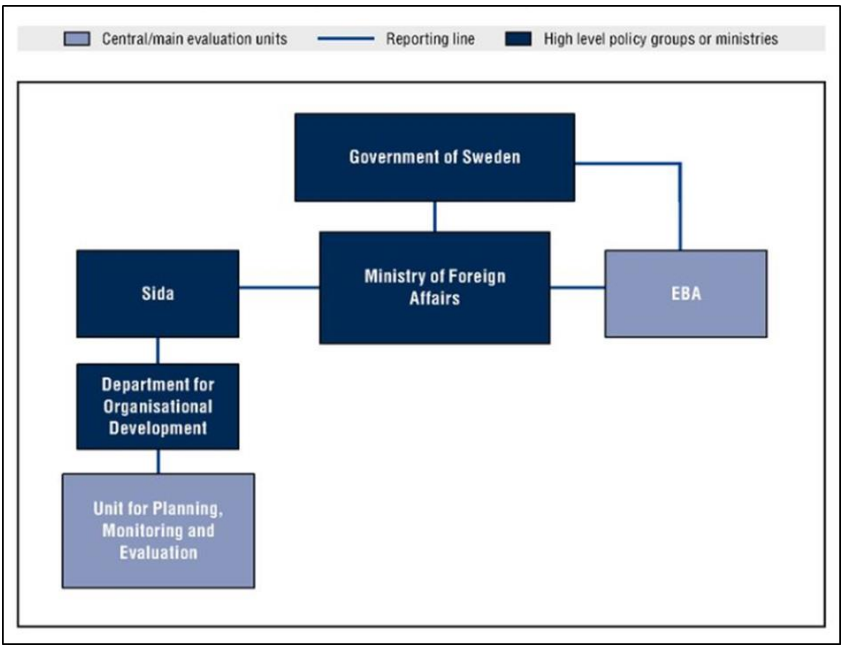
The Swedish evaluation function has also experience notable institutional shifts, both in terms of internal organisation in Sida and the Ministry's attention. Aid evaluation was established as a distinct unit within SIDA in 1971. Following a reorganisation into "new Sida" in 1995, the unit was expanded into a semi-autonomous secretariat (UTV) directly under Sida's director general. UTV included both the functions of evaluation and internal audit, and the first two heads and most of the secretariat's staff were recruited from the outside. In 2011, the evaluation secretariat was included in Sida's Department of Organisational Development. During the following years, the central evaluation function experienced a gradual reduction in budgets and staff, before it recently again has been strengthened (we discuss this in more detail in a later section of this chapter). Sida has thus operationalised the concerns for autonomy and integration in very different ways during the past decades.

A key feature of the Swedish evaluation system is the repeated efforts from Parliament and the Ministry for Foreign Affairs at establishing independent agencies tasked with external analysis and evaluation of Swedish development aid: SASDA, EGDI, SADEV, and currently EBA.[102] These agencies have been established to provide the Ministry with independent assessments of Swedish development aid, underlining the separation between Sida's learning-oriented, decentralised evaluation system and a perceived need for external, unbiased evaluations of the field of aid, including the work of Sida

[102] Secretariat for Analysis of Swedish Development Assistance (SASDA), 1993-1994; the Expert Group on Development Issues (EGDI), 1998-2003; the Swedish Agency for Development Evaluation (2006-2013); the Expert Group on Aid Studies (EBA), 2013-present. Cf. appendix 3 for a brief description of the four.

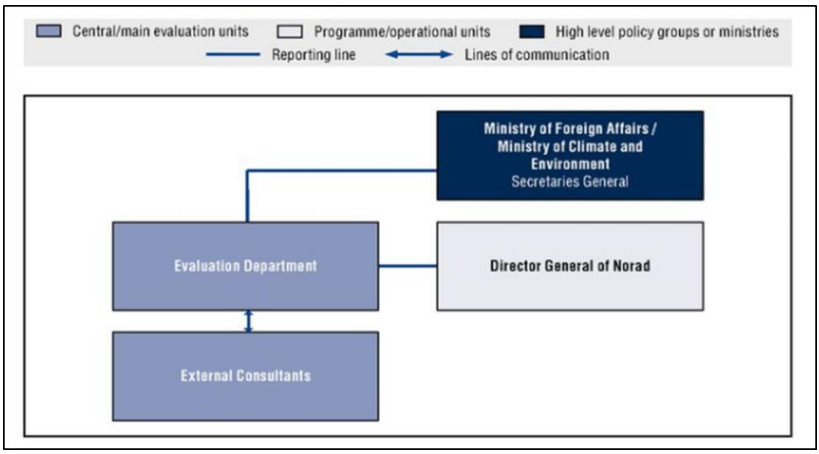
itself. A proper analysis of these four agencies' activities and their relation to Sida's own evaluation systems has unfortunately been beyond the scope of this study, yet our interviews indicate that there are indeed a most interesting interplay between the external and internal evaluation functions. One of our informants noted that one might see "a dynamic link between the existence of an external evaluation function and how Sida has conceptualized and resourced its central evaluation function." As an example, another informant explained that the establishment of SADEV made it possible to take on a more learning-oriented approach. A second important point is that, according to our informants, the experience with external evaluation agencies has clearly been mixed. Independence thus does not automatically ensure better control, objective critique, and higher accountability, only because they are positioned outside of the aid agency.

Figure 1. Organisational structure and reporting lines in Sweden's evaluation system [103]



[103] Retrieved from OECD 2016, p. 201. © OECD 2016. Republished with permission of OECD, from *Evaluation Systems in Development Co-operation: 2016 Review*, permission conveyed through Copyright Clearance Center, Inc. (Note: This chart is slightly inaccurate as it places the MFA outside, and not as a part of, the Swedish government.)

Figure 2. Organisational structure and reporting lines in Norway's evaluation system [104]



"Big learning": How to synthesise evaluation findings?

In our analysis of evaluation processes in chapter 4, we showed how staff often experience that learning may happen at the programme level – for the people involved in a specific evaluation of a specific programme (notably the involved consultants, programme officers, and evaluation managers) – while it is far more challenging to move the insights beyond this level and achieve learning on a larger organisational scale, what one informant dubbed "big learning".

Evaluation staff have, both historically and today, developed a number of tools for synthesising and facilitating learning (see Box 3). Two such tools are annual reports and newsletters. In both Sweden and Norway, the central evaluation units publish annual reports that summarise the past year's evaluation reports and often highlight a set of common themes, key findings, or main lessons. Sida's former evaluation secretariat prepared short newsletters several times a year in

[104] Retrieved from OECD 2016, p. 176. © OECD 2016. Republished with permission of OECD, from *Evaluation Systems in Development Co-operation: 2016 Review*, permission conveyed through Copyright Clearance Center, Inc.

order to reach a wider audience with summaries of recent evaluation reports.[105]

Box 3: Forms of evaluation synthesis and communication

Format	Description
Annual report (Sida)	Published independently by Sida's evaluation function. Sida's 2015 report contains a foreword by the evaluation director and compiles summaries of the past year's decentralised and strategic evaluations.
Annual report (Norad)	Published independently by Norad's Evaluation Department. Norad's 2015 report includes a foreword by the Evaluation Director, a short analysis of main findings and main conclusions of the past year's centrally commissioned reports, and summaries of each report commissioned or co-funded by the department.
Newsletter	Published multiple times a year during 1997-2010 by Sida's evaluation secretariat (UTV). Compiled easy-to read summaries of recent Sida-commissioned evaluations. Sent to a wide list of interested subscribers.
Evaluation brief	Currently used by Sida's evaluation unit to summarise and disseminate findings from new evaluation reports. Available on Sida's evaluation website.
Country evaluation brief	New format introduced in 2016 by Norad's evaluation department. Short presentations of Norway's main aid recipients that include descriptions of key contextual issues, the Norwegian aid portfolio, and findings from a selection of relevant evaluation reports. Commissioned by the evaluation department and conducted by independent researchers.
Evaluation synthesis	Report compiling findings from several evaluation reports. May also attempt to analyse the aggregate findings to identify more general recommendations or lessons learnt.
Meta-evaluation	Evaluation of evaluations.

[105] This was initiated by Sida's former Secretariat of Evaluation and Internal Audit (UTV). The newsletters are available from Sida's publication database.

Both annual reports and newsletters are important examples of efforts made by evaluation staff to translate individual evaluation findings into a more popularised form with less technical language, shorter text, and a broader audience. The central evaluation units have also commissioned synthesis reports, which summarise findings from a set of evaluation reports, and meta-evaluations, which analyse evaluation reports themselves according to specific criteria.[106] Yet despite these longstanding and ongoing efforts at generating more general learning from evaluations, evaluation staff continue to perceive broader organisational learning as a considerable challenge. Several of our informants stated that producing synthesis reports is particularly demanding: They are not only technically challenging and resource-demanding, but also often difficult to use in practice.

The synthesising of findings and "lessons learned" from individual evaluation reports also has practical challenges: What, precisely, may be generalisable from a given programme? What is relevant locally, but not in other settings? Is it possible to distinguish lessons that may travel beyond the specific programme, without being so general that they stop being relevant? As we argued in chapter 3 and 4, the articulation of recommendations is a most difficult task, and, we will suggest, identifying generalisable lessons is even harder. Is this just a question of priority – that donors are not willing to provide the resources necessary to enable synthesising and learning from evaluation? Or is the problem more fundamental – that we are expecting too much from the tool of evaluation?[107]

The experience of struggling with synthesising knowledge is not unique to our informants; as the researchers Casper Bruun Jensen and Brit Ross Winthereik showed in their analysis of Danish development aid, the sheer magnitude of available knowledge within development aid amounts to a considerable challenge.[108] Indeed, in a historical perspective, the aid field has shifted from having the problem of *no data* to the problem of *big data.* While this does make more knowledge

[106] One example of this is the report *Lessons and Reflections from 84 Sida Decentralised Evaluations 2013 – a Synthesis Review*, Sida Studies in Evaluation 2014:1.
[107] In recent years, so-called "systematic reviews" of aid evaluation, that mimic the methods of medical science, are promoted as a new tool for more effectively and accurately synthesising evaluation findings. Yet these are even more resource-demanding, given that they must be based on a number of impact evaluations (of a certain type), that in turn are very expensive and methodologically applicable only for a limited range of aid projects.
[108] Jensen and Winthereik 2012, p. 127, and Power 2007.

more available to more people, it also involves an increasing overload of available information. In principle, information infrastructures such as open databases make large amounts of documents easily accessible, yet organising and navigating these databases is no easy task and requires much work. In this way, digitalisation involves a proliferation of documents that both enables and obstructs transparency.

Indeed, as Jensen and Winthereik argues with reference to Michael Power's research, the sheer amount of accumulated documentation from monitoring, evaluation, and audit may cause not only an "audit explosion", but also an "audit implosion": An information infrastructure that sinks under its own weight, thus in practice becoming more inefficient and gaining less overview than it would have without the sophisticated documentation systems.[109] Paradoxically, while more information is available through monitoring and evaluation than ever before, the experience remains that there is a lack of knowledge, overview, and learning.[110]

Formalised routines for follow-up of evaluation

One important way in which evaluation systems seek to enable learning from evaluations, is through formalised systems of follow-up and use. Both Sida and Norad have systems in place for so-called "management response", which entail formalised routines for follow-up of evaluation reports. While we have not analysed these differences in detail, we will in the following point to few indicative findings emerging from our interviews and broader mapping of evaluation documents and routines.

In Sida, there are two paralell systems for management response, one for decentralised evaluations and one for strategic evaluations produced by the central evaluation unit. The former is integrated into Sida's project management system (Trac), while for the latter, the management reponse is decided by the Director General. This response to strategic evaluations is a public document that is made available in Sida's database together with the evaluation report. The follow-up of the evaluation report and the management response is

[109] Both Norad's and Sida's publication databases and biståndsdebatten.se are examples of such open databases. For research literature on information infrastructures, cf. notably Edwards 2012 and Bowker 2009.
[110] Cf. Reinertsen 2016 and 2017 (forthcoming).

then handled by Sida's Chief Controller. When this system was established in 1999, it was with the explicit aim of enhancing learning from evaluations, but in an evaluation report from 2006, the system was criticised for not achieving this purpose.[111] Several of our informants raised the same concerns and noted that one problem with the response system was that it was based on a check-list form of follow-up that "conflated learning into action points," as one informant stated.

In Norway, the Evaluation Department, upon publishing a new evaluation report, writes a so-called "transmission note" that summarises the evaluation and its recommendations. Also here, the recommendations are presented in the format of a table of concrete action points. This document goes directly to the Director General of the Ministry of Foreign Affairs, who is responsible for the further follow-up process.[112] The transmission note is thus the unit's main tool for communicating recommendations to the Ministry. After six weeks, the responsible unit within the Ministry must issue a formal response to the transmission note. Within one year, the department must issue a report detailing the progress of its follow-up work. The Evaluation Department publishes this string of documents on their public website and also lists the status of the ongoing follow-up processes in their Annual Report.[113]

While Sida's and Norad's systems for management response are different, they both have the same seemingly paradoxical feature of formalising learning processes by transforming them into checklists.

---

[111] Sida's Secretariat of Evaluation and Internal Audit (UTV) commissioned an evaluation of the management response system in 2006. It states that: "Sida's management response system was introduced in 1999 to promote learning and enhance Sida's effectiveness. This study analyses the system's characteristics and basic assumptions, as well as how it works in practice. It also assesses the systems relevance and identifies three options for the future. Three main conclusions are drawn. First, that the assumptions of the MRE system are reasonable and consistent to attain the desired outcome of better documentation and structure, but not with the intention of organizational learning. Secondly, it is concluded that the system made a limited contribution to learning as implementation has been slow and uneven. Thirdly, it is said that the system does not enhance partnership, dialogue and ownership." *Sida's Management Response System*. Sida Studies in Evaluation 06/01.

[112] If the evaluation report concerns Norwegian climate-related aid programs, notably the major program of rainforest conservation (REDD+), the Evaluation Department reports to the Secretary General of the Ministry of Climate and Environment, who currently handles this part of the Norwegian aid portfolio.

[113] The process is further described on the Evaluation Department's website, while the specific follow-up documents are published on the website of the corresponding evaluation report. https://www.norad.no/en/front/evaluation/what-is-evaluation/follow-up-of-evaluations/ (retrieved 02.11.2016).

The public display combined with the check-list format responds to the imperative of homewards accountability as it clearly contributes to increase the external insight into the follow-up of evaluations. Yet the set-up does not include strong mechanisms for sanctioning non-compliance. Norad's Evaluation Department has no authority beyond making it publicly visible online should the Ministry fail to submit its follow-report on time or refrain from implementing the recommentations from the transmission note. Similarly, Sida's Chief Controller has few instruments for sanctioning whether the follow-up plan is being followed. At the same time, this may also be seen as a way to counter-balance potentially haphazard recommendations in evaluation reports. Given the challenges of crafting good recommendations, as we discussed in chapters 3 and 4, this set-up also ensures that reports are not given too much power over internal affairs.

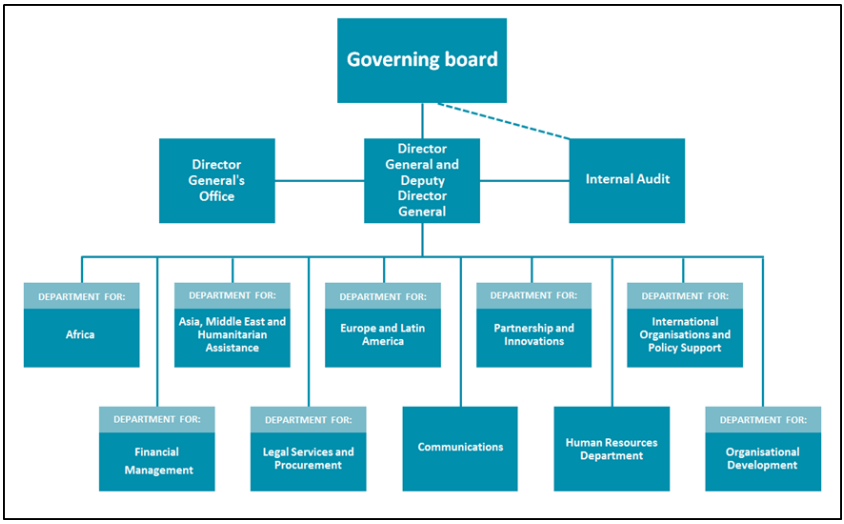
Indeed, views will differ, depending on the actors' institutional standpoint, on the degree of authority that should be assigned to evaluation reports and their recommendations. One interpretation of management response systems is that while they clearly ensure that all evaluations are formally handled by the wider organisation, they do not necessarily generate substantial change. Change will still depend on political leadership, top management, and individual staff members actively reading evaluation reports and taking action based on them - in a context that might be influenced by multiple other concerns, interests, and sources of knowledge. Hence, individual interest and engagement is critical, but so is the specific context into which the evaluation report arrives. Evaluation managers at the central units may, as we showed in chapter 4, seek to make their evaluations as timely and relevant as possible, and may promote stronger systems of management response. Yet the critical role of both individual engagement and the decision-making context may explain evaluation staff's persistent experience that, as one of our informants exclaimed, "evaluation uptake is a mystery".

The broader systems of monitoring and accountability

While evaluation has become a distinct field of expertise within international development, as we discussed in chapter 4, it is important to acknowledge that the field is historically and methodologically integrated with the wider management systems of

aid planning and monitoring.[114] In both countries, evaluation staff have historically been much involved in developing methods and systems for both evaluation as such and results management more generally.[115] Today, the relation between monitoring and evaluation is institutionalised differently in the two countries: In Sida, the functions of monitoring and evaluation are integrated in the Unit of Planning, Monitoring and Evaluation under the Department of Organisational Development. In Norad, evaluation is organisationally separate from performance monitoring and institutionalised in, respectively, the Evaluation Department (EVAL) and the Department of Quality Assurance (AMOR). See figures 3 and 4 for organisational charts of both institutions.

Figure 3: Organisational map of Sida. The Unit for Planning, Monitoring and Evaluation is under the Department of Organisational Development [116]

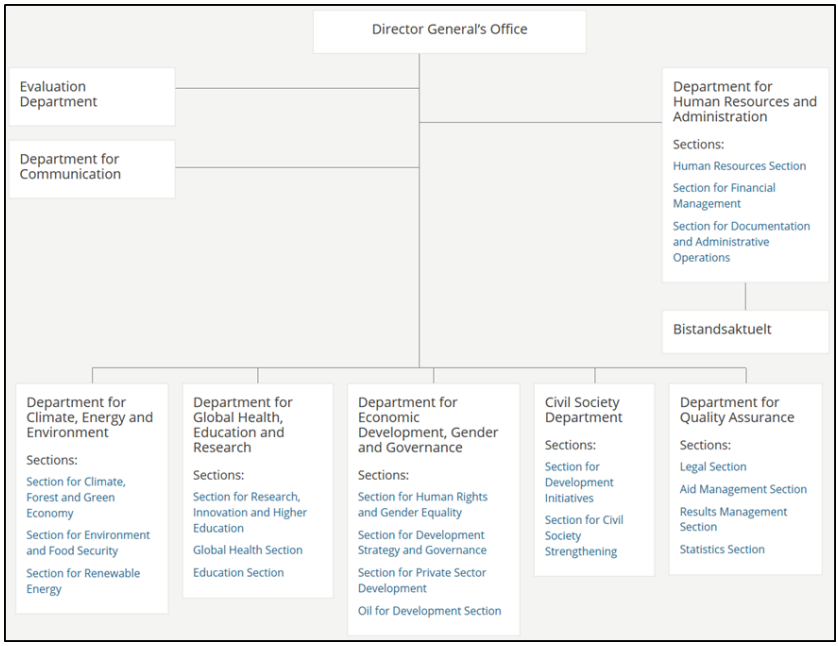


[114] Evaluation was initially developed within public sector administration in the early 1960s, notably in the United States, as an integrated part of planning and implementation of large-scale public programs within education and welfare during the 1960s. Cf. O'Connor 2001 for a detailed account of how evaluation was institutionalised within the US government, notably president Lyndon B. Johnson's major programme "War on Poverty", from 1964 onwards.

[115] Cf. Appendix 3 for a brief historical overview of both countries. For detailed historical analyses, cf. Vähämäki (2017) for Sweden and Reinertsen (2016) for Norway.

[116] http://www.sida.se/globalassets/global/about-sida/organigram_2016_eng.png (retrieved 03.01.2017).

Figure 4: Organisational map of Norad [117]



This interrelation between monitoring and evaluation also has a practical side: For an evaluation to be even possible, the evaluation team needs to access documentation of the programme's past activities. Under the current reporting routines, programme officers in both countries are obliged to submit regular reports throughout the life of an aid programme. However, external evaluations have often found that such documentation is lacking, and have recommended intensified reporting systems to remedy the situation. In Norway, this has been a consistent point of critique by the Evaluation Department, who both in the early 1980s and today argue that the lack of available documentation makes it difficult to know the effects of Norwegian aid, and, moreover, that this lack of documentation impedes learning.[118] In contrast, programme officers are reported to claim that the reporting systems in themselves may amount to a considerable

[117] https://www.norad.no/en/front/about-norad/organisation-chart/ (retrieved 03.01.2017).
[118] Cf. Norad 2013, Norad 2014, Norad 2016, Norad 2017. In these publications, either an external evaluation team or the Evaluation Department itself argues that the lack of documentation makes it difficult to conclude about the effects of Norwegian aid. This ecchoes the Evaluation Department of the early 1980s, who was concerned about a lack of knowledge about Norwegian aid and therefore introduced the Logical Framework Assessment and enhanced monitoring routines (cf. Reinertsen 2016 for this history).

burden. In practice, staff experience a "time squeeze" which leaves almost everyone dissatisfied.[119] Indeed, in past analyses of learning in development aid, the lack of time is singled out as a key barrier to learning.[120]

Institutionally, the evaluation functions of both Sida and Norad are distinctly separate from both internal audit and external audit.[121] In both countries, the management of development aid is also evaluated by the national Offices of the Auditor General, who routinely investigate the governmental ministries' use of parliamentary budget allocations.[122] This clearly separates the formal accountability function from the internal learning function. Yet we will suggest that the distinction between these two is not necessarily so clear-cut. The Office of the Auditor General (in both countries) not only undertakes financial audits, but has also during the past decade developed so-called "performance audits" that methodologically go beyond financial auditing to also assess whether programmes and interventions meet the stated goals and targets of Parliament. Increasingly, both the methods and professional networks of aid evaluation and performance audits overlap. Hence, aid evaluation is always a part of a larger institutional landscape that in combination should ideally enable both accountability and learning to take place. The way this landscape is organised affects what kinds of evaluations are performed. It clearly matters which institution, office, and managers are commissioning, conducting, and communicating the evaluation, as this affects people's expectations and reactions to the evaluation process, the evaluation team, and the evaluation report.

[119] Cf. Norad 2014, in which the evaluation team interviewed staff about the practical barriers to results assessment.
[120] Krohwinkel-Karlsson 2008, cf. especially the recommendations.
[121] In Sida, the Unit of Internal Audit is organisationally independent and reports directly to the Director General and Sida's governing board. In internal audit is overseen by the Department of Methods and Results (Avdeling for metode og resultater, AMOR); in the Ministry of Foreign Affairs, by the Unit of Internal Control (Intern kontrollenhet).
[122] The office is called "Riksrevisjonen" in Norwegian and "Riksrevisionen" in Swedish. In Sweden, evaluations are occasionally also performed by the Swedish Agency for Public Management ("Statskontoret").

The importance of political context and management support

While the drive for organisational learning comes mainly from within the aid community, the demands for accountability come from both internal and external actors. Historically, evaluation officers have strongly promoted the issue of better documentation and monitoring and actively contributed to develop methods and routines for this.[123] During the past decades, political and management support for evaluation has been shifting several times in both countries, ranging from strong support to indifference to outright neglect (cf. appendix 3). These shifts have obviously directly affected the evaluation efforts. High political interest and priority have meant increasing budgets, more staff, more independence, and direct lines to top management. Low political interest and priority has meant the opposite: reduced budgets, fewer staff members, less internal authority, and no direct line to top management.

The methodological developments of both aid evaluation and governmental audits are directly related to stronger public demands for more transparency in aid funding and demands for better documentation of aid results. As one of our informants stated; there are increasing demands from Parliament, the media, and the public at large to "know where the money goes". While this concern has been strong also in previous decades, it has gained more political prominence during recent years. In Norway, a main shift has been the new principle by the current government that "evaluations should have financial consequences".[124] While the practical operationalisation of this is still unclear, it signals a threat of reduced budgets should an intervention experience a negative evaluation. Given our analyses above of the implication of choosing an accountability purpose in an evaluation process, this political position clearly favours accountability and control over learning and openness – perhaps unwittingly.

In Sweden, a shift of government in 2006 both entailed a stronger attention to results management and stricter budget control. Sida underwent several major reorganisations that involved major budget cuts and a 20% reduction of staff. During these years, Sida also

[123] Reinertsen 2016, Vähämäki 2017. Cf. also Appendix 3.
[124] *The Sundvollen Declaration* (The political platform of the current government, signed October 7, 2013). https://www.regjeringen.no/no/dokumenter/politisk-plattform/id743014/#utenriks_bistand (last retrieved November 2, 2016).

implemented a new project management system (called "Trac"). From 2012, these initiatives were integrated into the so-called "Results Agenda" of the Swedish government.[125] During these processes, the evaluation function was also reorganised, as mentioned above: From having been a semi-autonomous unit (the Secretariat of Evaluation and Internal Audit, UTV) that reported directly to Sida's Director General, UTV was subsumed under the Department of Organisational Development. This shift was encouraged by the evaluation unit itself, which considered it an advantage to work closer to the overall organisation. Hence, at a time when Sida at large experienced an shift towards stricter accountability and an external agency (SADEV) was established to undertake external evaluations of Swedish aid, Sida's central evaluation unit deliberately sought integration and organisational learning. During 2008-2011, the central unit again expanded its methodological ambitions, took part in several major joint evaluations, and experimented with new learning-oriented evaluation processes (including the utilisation-focused evaluation *Supporting Child Rights* during 2010-2011, which we discussed in previous chapters).

Since 2011, the evaluation unit has been further integrated in the department and become part of the Unit for Planning, Monitoring and Evaluation. In the process, budgetary and staff resources that had previously been reserved for evaluation were subsumed under the unit's overall budget and staff, with the result that both funding and staff were gradually diverted to other purposes. With eventually only 1-2 staff members fully dedicated to work with evaluation, the unit made the production of strategic evaluations the main priority. Several of our informants voiced deep concerns about what they considered a "demotion of evaluation within Sida", which they considered to have caused an erosion of Sida's former evaluation expertise and learning orientation in general, as systems, routines, and initiatives were discontinued.

Following a new change of government in 2014, the Results Agenda has been eased. The evaluation function is also experiencing more political and management support. In 2015, the Swedish parliament explicitly called for enhanced evaluation efforts, and this was included in the MFA's annual letter of appropriation to Sida.

[125] Cf. Vähämäki 2015, 2017 for analyses of the "Results Agenda". See Appendix 3 for a short summary.

During 2016, the evaluation function has experienced increasing budgets, and as of January 2017, staff resources are back to their previous size (5-6 staff members reserved for evaluation). As such, the Swedish experience again illustrates both how political interest and management support directly affect the authority, ambitions, and orientation of the evaluation function, including the weight it may give to either accountability or learning.

Donor orientation or recipient orientation?

One key concern emerging from our analysis, which was also voiced by several of our informants, is the question of *who we evaluate for*. This is directly related to the issues of accountability and learning: To whom are aid programmes accountable? Who should learn from evaluations? In this matter, our informants' views varied greatly. One of our informants frankly stated: "We evaluate for ourselves." Another was especially critical of what he perceived to be a lack of a recipients' perspective in aid evaluation. A third informant did not share this critique; with reference to the Norwegian system, he made a clear distinction between the partners and recipients "out there" (to whom decentralised evaluations would attend) and the users "here at home", and was thus mainly concerned with ensuring that evaluation was used and contributed to change and learning among policy makers and donor staff. A fourth informant directly opposed the notion that "we evaluate for ourselves", and highlighted that the evaluation reports published in Sida's publication series were only complementing the much larger amount of evaluations undertaken by the partners who were implementing the aid programs. At this level, it was the partners themselves and not Sida who should learn from evaluations.

Our informants' diverging responses to this question is most interesting, as they highlight how their geographical, institutional, and also historical positions have implications for their views. As we have demonstrated in this and previous chapters, there is ample evidence that both evaluation reports and evaluation processes in practice enable what Alnoor Ebrahim terms *upwards accountability*, i.e. "homewards" to the donor countries. Yet *downwards accountability* to the partner organisations, aid intermediaries, and end beneficiaries

seems only to a little extent to be a part of the donors' own evaluation systems.[126]

This is a paradoxical situation, given that development aid is obviously intended to assist aid recipients. Indeed, in the early 1990s, both Sweden and Norway introduced major reforms to promote "recipient orientation", often also termed "recipient responsibility".[127] In Norway, a key part of this reform was the introduction of new forms of project reporting: The recipient was responsible for managing the programme and to report on results, while the donor was responsible for the follow-up of the recipients, i.e. to ensure that the recipient could account for funds spent and also document that the expected results had been delivered as planned. Yet while the purpose of this reform was to transfer responsibility to the recipients and thus reduce the donor's role, this in practice entailed a different form of donor dominance: stricter systems for project management enhanced accountability and transparency at home while reducing the recipients' autonomy.[128]

The donor community is very much aware of the accumulated negative effects that their individual domestic systems of project monitoring and evaluation have upon aid recipients. Both the Paris Declaration for Aid Effectiveness (2005) and the Accra Agenda for Action (2008) include calls for harmonisation of monitoring and evaluation between donors, in order to ease the administrative burden of having to respond to multiple different systems and routines. Indeed, according to a review prepared for OECD-DAC's Network on Development Evaluation (EvalNet) in 2010, "mutual accountability" was then emerging as "an overall trend". The principle was included in EvalNet's Quality Standards for Development Evaluation from 2010 and further confirmed in the high-level meetings in the Busan Partnership for Effective Development in 2011. Yet according to EvalNet's comprehensive follow-up report from 2016, "there is evidence that the principles of ownership and mutual accountability have yet to result in a high level of partner

[126] Reeger et.al 2016, building on Ebrahim (2005). Cf. chapter 2 for a brief description of the concepts.
[127] For Sweden, see Odén 2005, Vähämäki 2015, 2017. For Norway, see Liland and Kjerland 2003, Reinertsen 2016.
[128] Cf. Reinertsen 2016 for the Norwegian case. See Rottenburg 1990 for a similar analysis, based on data from the German Development Bank in the 1990s.

participation".[129] In practice, harmonisation remains difficult and donors are undertaking fewer joint evaluations than before. This may point to a practical contradiction between the donors' concern for accountability towards their tax payers and the effort at enabling better aid efficiency for the recipients.

There is a second important dimension to the paradox of recipient responsibility: It means that the donors are less present in the field. For example, in Norway, the embassies are responsible for handling the aid portfolio, while Norad staff are based in the Oslo headquarters.[130] Increasingly, we suggest, aid work is becoming paperwork.[131] This brings us back to the question: *Who learns and how?* Reduced personal contact and increased written documentation is clearly a result of specific instructions given from management following major reforms in both countries. This determines how staff members spend their time and the priorities they make in their daily work. As one of our informants noted: "There is a greater distance now between the aid world and the real world." This raises the question: Which site is more important – the donor country or the recipient country? While most would agree that aiding the recipient is the most important part of an aid relationship, the specific ways in which the donors' management systems are set up may in effect promote accountability homewards (to the donor countries) rather than to the end beneficiaries.

Chapter conclusions

In this chapter, we have sought to broaden the scope of our analysis to how the context of aid evaluation has implications for the relative weight given to accountability and learning. While evaluations are but one of several inputs into decision-making, they are also an integrated part of multiple mechanisms in place to ensure accountability: planning, monitoring, auditing, and management response.

[129] OECD 2016, *Evaluation Systems in Development Co-operation*, p. 21. The first quote refers to the report *Evaluation in Development Systems* from 2010, to which the 2016 report was a follow-up. Both reports were commissioned and published by the OECD-DAC Network on Development Evaluation (EvalNet).
[130] This was a key feature of the major reorganisation of Norwegian development aid in 2003. Cf. Norad 2014 for the recommendation that Norad staff should have more missions abroad, as used to be the norm. Cf. Simensen et.al 2003 for historical analyses of these changes.
[131] Cf. Reinertsen 2016 for an elaboration of this argument.

Furthermore, the systems and tools available to ensure learning are challenging to use: While evaluation staff works hard to produce annual reports, newsletters, synthesis reports, and public databases, "big learning" remains elusive.

Based on our discussions in this chapter, we suggest that there clearly are tensions – and sometimes even contradictions – between accountability and learning. These tensions and contradictions are clearly too fundamental for individual evaluation managers to resolve in their daily work. As we showed in chapter 4, evaluation staff work hard to reconcile the two purposes in practice. The evaluation literature, as discussed in chapter 2, has long identified these challenges and suggested ways to resolve them. Our analysis in this chapter suggests that the challenges are still unresolved and potentially also exacerbated by built-in features of the broader systems of aid evaluation and management: The expectations of accountability and the practical arrangements that seek to enable this also entail a reduced potential for learning – for decision makers, programme managers, the wider public, and the end beneficiaries, whose concerns are only to a limited extent included in the current evaluation systems of Sida and Norad.

Conclusion

Learning is a key purpose of aid evaluation. So why do aid organisations not learn more from their own experiences? More specifically, why do they not learn more from their own evaluations? For more than 30 years these questions have been asked by the public, by politicians, by aid staff, and by evaluation professionals. Yet learning is but one part of the well-established "dual purpose" of aid evaluation: The other key purpose is accountability. In this study, we have investigated how these two purposes are often difficult to reconcile in practice.

Main argument

Our main conclusion is that *the dual purpose of accountability and learning in practice causes difficult trade-offs*. Our integrated analysis of evaluation texts, evaluation processes, and evaluation systems shows how tensions, and sometimes direct contradictions, between accountability and learning arise. In the following, we present our main findings from each level. Key questions guiding our analysis have been: Who writes and reads evaluation reports? How are they produced, circulated, and used? Who learns from evaluations, and how? How do reports, staff, and systems negotiate between the diverging concerns of accountability and learning? And how has this varied over time and between Sweden and Norway?

The evaluation text

Our rhetorical analysis of a sample of evaluation reports shows that while they may clearly contribute to accountability, they to a much lesser extent contribute to learning. This finding is consistent over time and between the two countries. Although the reports at first sight *look* different than they did 40 years ago, they have changed rather little in terms of structure and content. While several sub-genres of evaluation reports exist, the main report genre is generally well-established and combines the three classic rhetorical elements: to establish what happened, to allocate praise or blame, and to propose what to do.

In our sample of 20 evaluation reports, the first and second rhetorical elements (establish what happened and allocate praise or

blame) are largely covered through description and analysis. This contributes to fulfil the accountability purpose of evaluation. The third element (propose what to do) is covered through the mandatory sections of "recommendations" and "lessons learned". Yet these sections are most often only loosely based on the preceding analysis. In most of the reports we studied, the recommendations disregard critical contextual factors even when the importance of context is explicitly noted in earlier sections of the same report. This further deepens the disconnection between description and recommendations, which greatly impedes the potential learning from evaluations. While this could mean that the reports are simply of low quality, we conclude that improving the quality is an insufficient solution; it is also necessary to consider how the quality is contingent on processes and structures outside the report itself, notably by how the Terms of Reference (ToR) are formulated by those commissioning the evaluation report and by the resources made available for aid evaluation.

The evaluation process

Both Sida and Norad have well-established formalised routines for how to plan an evaluation, prepare the Terms of Reference (ToR), procure an evaluation team, lead the evaluation process, and follow up the published report. Already at the starting point in the evaluation process, key premises are established for whether an evaluation will contribute primarily to accountability or learning. The two purposes involve asking different sets of questions and applying diverging methods. Furthermore, the formal routines are complemented by informal practices. Building and sustaining internal engagement for the evaluation is critical to ensure cooperation, interest, trust, and, ultimately, learning and use. But this must constantly be balanced against the accountability principles of critical distance and independence, as too much internal involvement may reduce the external trust in the evaluation process.

This situation poses important dilemmas: Should the evaluation team function as auditors or process facilitators? Should they write their report mainly for external control or internal change? Should they prioritise internal or external trust? Transparent processes and methodological rigour may enable some reconciliation between these diverging concerns, but they cannot completely avoid the trade-offs. The role assigned to and taken by the external consultants directly affects the learning potential. If they take on an exaggerated role as

critics, they may end up conducting their work in an unpedagogical manner that makes people defensive, which in turn means that learning opportunities may be lost. Furthermore, their recommendations are often perceived to be inappropriate; they could be too specific, or too general, or too ambitious. Yet the most fundamental problem with using external consultants is that those who learn the most in the process have no responsibility for applying the lessons. This relates to the simple question of *who writes evaluation reports*. The fact that the practical work of analysis and writing is mainly done outside the aid agencies themselves clearly serves the accountability purpose of evaluation, yet it also means that important learning disappears from the aid agencies.

Correspondingly, asking *who reads evaluation reports* is illuminating. Feeding lessons learned back into the organisation by means of the evaluation reports and related efforts at synthesis and communication remains a considerable challenge for the evaluation staff. Their main experience is that few have the time to read evaluation reports and absorb their content. Our analysis prompts fundamental questions: Why is it so important to keep producing reports that few will read? Why is the procurement of external consultants more important than enabling internal learning processes? Answers to these questions relate to the wider context in which aid evaluations take place.

The evaluation system

Aid evaluation is always but one part of a larger context, what one of our informants aptly called "a power field of diverging concerns and interests". Sweden and Norway have repeatedly re-organised their aid evaluation activities during the past 40 years, choosing different ways of balancing the concerns for integration/distance, involvement/control, and accountability/learning. Again, as in the case of evaluation reports and evaluation processes, there exists no perfect solution; rather, the balancing act involves making pragmatic choices between important concerns that in effect involve difficult trade-offs. Given that evaluation reports make visible to outsiders what happens inside the world of aid, they are of obvious democratic value and a necessary means for maintaining public trust in aid. But when accountability is too narrowly defined to mean merely the reporting of documented results, it may clearly come at the cost of learning.

Two main comparative features of the Swedish and Norwegian evaluation systems stand out: Firstly, in Sweden, the evaluation system is largely decentralised, which means that programme-based evaluations are also considered a key part of the evaluation system. The central unit has produced strategic evaluations and assisted in decentralised evaluations. In contrast, in Norway, there is a clear separation between the centrally produced evaluations and decentralised evaluations, which untill recently was termed programme reviews (and still is in Norwegian). Second, there are notable in differences in how the two countries have chosen to institutionalise the two concerns of integration and autonomy. The Norwegian evaluation unit historically has moved from Norad into the Ministry of Foreign Affairs and back to Norad, and in the process shifted from semi-autonomy to an integrated model and back to semi-autonomy. Sida's central evaluation unit has also experienced clear shifts – from being first its own unit within the wider organisation, then expanded into a strong semi-autonomous unit before again being integrated into the wider organisation. Yet a key feature of the Swedish model that complements Sida's own evalaution work has been the repeated establishments by Parliament and the MFA of external agencies (SASDA, EGDI, SADEV, and EBA) that were also tasked with aid evaluation.

The choice of evaluation system clearly has implications for how and where evaluation may contribute to either learning, accountability, or both. As such, they are manifestations of different ways of answering the key questions of "accountability for whom" and "learning for whom": Should the accountability chain "homewards" be given more weight than the accountability towards aid recipients, aid intermediaries, and end beneficiaries? May learning be acknowledged to mean project-level learning based on inclusive evaluation processes, or is this insufficient from a donor perspective? How donor countries choose to handle these important questions in turn directly affects what role evaluation may play.

Who learns from evaluations?

Given that mainly external actors write evaluation reports and few people read them, *who learns from evaluation reports*, or more broadly, from evaluation processes? Our analysis shows that learning may well happen at the programme level, notably for the external consultants, evaluation managers, and programme officers partaking in specific evaluation processes. All our informants emphasised this point,

whether they were concerned with decentralised or centralised evaluations. What we may term "sideways learning", for actors involved in specific evaluation processes, is thus reported to be common. Yet also in these practical evaluation processes we repeatedly encountered examples of how learning might be limited by the tension between accountability and learning.

The notion of "sideways learning" mainly involves those working for the donor agencies, whether as evaluation staff, programme staff, policy staff, or external consultants. The role of partner organisations, aid mediaries, and end beneficiaries is yet another set of relevant groups one step removed from the donors. Is evaluation supposed to be *about* them, *with* them, or even *by* them? In effect, according to some informants, recipients and beneficiaries were at best included as stakeholders, but rarerly made active partners in the evaluation process itself. Other informants disagreed with this understanding and pointed beyond the evaluation systems of Sida and Norad, noting that one also needed to take into account the partner organisations' own evaluation systems, which were designed to enable learning not for the donors, but for the recipients, here meaning the partners and implementers themselves.

The end beneficiaries of aid thus hold only a limited role in the donors' own evaluation systems. As one of our informants stated: "We evaluate for ourselves." Yet the difference between learning on-site and learning at home is considerable: It is most challenging to generalise and synthesise findings from evaluations and achieve learning on a larger organisational scale, what one of our informants called "big learning". While the central evaluation units have sought to enable this in multiple ways during several decades – through the means of annual reports, newsletters, synthesis reports, public databases, and follow-up plans – "big learning" remains elusive.

Finally, a fundamental problem is that the aid system on all levels displays exaggerated expectations of what aid evaluation may accomplish. The expansive growth of evaluation reports and other available documentation and information makes many assume that increased knowledge and learning will automatically follow. Yet this linear learning model does not match the practical experiences in the field: The current situation of "big aid data" does not remedy the widespread experience that we know too little and learn too little. This problem is only deepened by intensified calls for transparency, accountability, audit, and control, which, while serving critical

democratic functions, are currently operationalised in ways that do not necessarily harmonize well with the ambition to learn. One learns not least by making mistakes, and one must expect aid work to involve making *many* mistakes. A realistic approach would thus entail high tolerance for error. In reality, the expectations to aid are much stricter than this. If an aid effort fails to achieve its goals, if funds fall to corruption, or if the impacts are not what one had planned, the media, and – in some cases – politicians are quick to make a scandal of it, while the aid administration is forced to defend itself in public. This may deepen distrust both externally (to the institution of aid) and internally (to the institution of aid evaluation).

To conclude, what we have described above are all expressions, on different levels, of a persistent tension between accountability and learning in aid evaluation that cause difficult trade-offs. In practice, the main result of this is a prioritisation of the former at the expense of the latter. To put it simply: Learning is crowded out by accountability.

Key recommendations

1. We must talk openly about the trade-offs between accountability and learning.
2. We must adjust our expectations to both aid interventions and aid evaluations.

The term "we" here points to everyone involved in doing and discussing aid evalution: from evaluation managers, aid practioners and policy-makers to researchers and the wider public. Following these recommendations would, we suggest, require that both those involved in aid and those discussing it on the outside must acknowledge that regardless of their own position on the topic discussed in this report, a set of choices will have to be made. The following list is not exhaustive, but it captures the most important choices that are now often made without explicit discussion of their implications.

Choice 1: Does the evaluation process need an evaluation report, and if so, what kind? Too many evaluation reports are hardly read. One should therefore always answer the question of whether a report is needed, and if it is, what purpose it should fulfil and how it thus

should be produced. This includes determining its intended readers and users, which in turn should inform who the writers should be. If the purpose is external accountability, then a consise report mapping existing activities and outcomes may be sufficient. If the purpose is internal learning, then a published, publicly available report may be counter-productive.

Choice 2: Does the evaluation process benefit from an external evaluation team? The use of external consultants should be weighed against their cost, and their added value should be explicitly justified. The role of consultants is directly related to the purpose of the evaluation. If the purpose is accountability, then a limited audit mission might be most beneficial. If the purpose is learning, then the team may rather function as facilitators of the evaluation process, providing a neutral outsider perspective. Internal participants and external stakeholders must be actively included throughout the process, at the minimum through a self-evaluation that is granted equal weight as the external evaluation.

Choice 3: Should the evaluation report include recommendations? Recommendations are commonly produced by the evaluation team as part of the evaluation assignment. The articulation of recommendations is often the weakest point of the evaluation process, yet it is also the most important one. This is where the mapping and analyses produced through the evaluation process may be translated into potential action. It is not a given that the evaluation team are best equipped to articulate recommendations. Other models may be more useful: The team could instead suggest a set of scenarios from which the involved programme staff and policy makers may choose, after being well-informed of the potential trade-offs thus involved. Recommendations may be articulated by them, possibly in a process facilitated by the evaluation team. Or the intended users of an evaluation may have the responsibility, upon receiving the report, to articulate recommendations to which they in turn will be held accountable.

The three choices above are practical manifestations of our overall recommendations, and they pertain mainly to evaluation processes and concerns within the evaluation community. At the same time, our recommendations also connect to more fundamental questions about the legitimacy of development aid at large, and the expectations of external actors – policy-makers, commentators, the public – of what

aid evaluation should be and what it should achieve. Our final choice addresses this more fundamental issue.

Choice 4. Should accountability systems be given the current high priority by donors, even when they come at the expense of internal learning? There is an obvious, democratic need for systems of monitoring and evaluation of aid, because they promote accountability and transparency to taxpayers. There is, however, in theory no limit to how comprehensive such accountability systems can be; and they have become steadily more demanding over time. There should thus be a debate, both within and outside the aid community, about the choice between enhancing the accountability-focused evaluation systems and allowing a greater emphasis on learning. Those calling for more comprehensive systems of control and stronger evidence of success should thus acknowledge the actual cost of their demands in terms of increased budgetary expenses, administrative work, and organisational stress, and reduced learning potential.

Appendix 1: Theoretical framework and methodology

Theoretical framework

The research team brings together three strands of scholarship that are all concerned with the question of how knowledge is produced and used: Political Economy, Science and Technology Studies (STS), and Rhetoric. While Political Economy has had a long engagement with aid as a field of study, this is fairly new within STS and hardly existing within Rhetoric. In combination, we suggest, they provide a powerful tool for analysing learning in aid on multiple levels, from evaluation documents and evaluation processes to evaluation systems and their political context.

Political Economy

Within political economy, there is a longstanding tradition of critically examining the power and authority of aid expertise. This is directly related to the question of accountability, or in our case: To whom, precisely, are Sida and Norad accountable? And how do evaluation reports respond to this imperative? The domestic political context in Sweden and Norway has changed in recent years, with increasing criticism of aid effectiveness, from sources ranging from well-informed practitioners to "scandal-seeking" media. Associated with this have been increasing demands for accountability – not so much to the "recipients" of aid as to the taxpayer. Rottenburg (2000:147) notes how the need for "juridical" accountability is necessarily translated, in development aid, into a requirement of the correct application of procedures, and that risk-taking and innovation are hampered by the demands of what Power (1994) labelled "the audit explosion."

The political economy perspective is useful for better understanding the context within which evaluation reports are prepared. This involves a critical analysis of the (changing) power and authority of all the different actors involved: those for whom reports are written and those by whom they are written. In the former group, for example, the Auditor General enjoys authority as the formal representative of the

Norwegian taxpayer, and has the power to influence Norad's budget; by contrast, the media lack formal authority but can influence public opinion – in general or about specific aid activities. The authority of those who produce evaluation reports derives from their technical expertise and objectivity, and also perhaps their ability to bring the reports to the attention of those who have power to change aid policy and practice. The analysis draws on theoretical literature relating to power and knowledge in the development and international relations field (e.g. Haas 1992, Cox 1997, Bøås and McNeill, 2004), and on studies of the politics of evidence, the audit society, and the power of numbers (e.g. Power 1994, Mosse 2005; 2011, Eyben 2013). There is, of course, a voluminous literature on the political economy of aid, some of which is of relevance to the more specific focus of this study.

Science and Technology Studies (STS)

The relation between knowledge and politics is also the foundation of the interdisciplinary field of Science and Technology Studies (STS), which holds that the two must be understood as co-constructed, inseparable, and interdependent (Asdal 2012, Jasanoff 2004, Latour 1999). Hence, knowledge and politics, or science and society, must always be analysed in combination. Studying empirically the relation between knowledge, science and expertise and the wider society, STS has increasingly turned to aid policy and practice as a field of study (Brodén 2013; Jensen and Winthereik 2013; Mosse 2005; Reinertsen 2016; Riles 2000; Rottenburg 2000, 2009).

Following this analytical framework, the aid administration is understood as organized around the particular tools, routines, offices, and documents of planning and governing (Asdal 2008, 2011; Hull 2012; Kafka 2009; Latour 2010; Reinertsen 2016, 2017; Riles 2006). Documents and office routines structure, organize, intervene in, and enable politics, and hence serve as an indispensable foundation of the state itself. Still, this research argues, documents and offices are not in and of themselves powerful, although being carriers of, and themselves producing, knowledge and expertise; they may in fact also produce "non-authority" (Asdal 2011:2). To that end, the research shows the limitations of the audit society thesis, and suggests that the problem is not necessarily always the power of experts, but perhaps rather their lack of power. This potential disconnection between expert advice and political/public action is also a main concern in sociological STS

studies of socio-technical controversies (such as climate change mitigation or the disposal of nuclear waste), in which researchers aim to bridge the gap between science and politics by experimenting with different kinds of expert interventions, stakeholder engagement, and science communication (e.g. Callon et al. 2009, Hilgartner 2000).

Rhetoric

The third theoretical entry point is rhetoric, which typically concerns the study of texts, including their production, dissemination, and consumption. Relevant questions for this type of study are: *Who writes what, how, and for whom?* And, at an opposite end: *Who reads what, how, and for what purpose?* Furthermore, we incorporate rhetorical scholar Carolyn Miller's notion that the production and consumption of text should be studied as a form of "social action" (Miller 1984). We thus aim to understand not just what particular texts are, but also what they do. This takes us to text historian Robert Darnton's widely used "communications circuit," which he describes as a "general model for analysing the way [texts] come into being and spread through society" (Darnton 1982:67). From this starting point, text historians typically reject narrow notions of "rhetorical impact" and replace them with a broader, more holistic, notion of the forces at work in communication processes.

Correspondingly, we study the evaluation report as an historical event and investigate the entire complex of imperatives that are at work on the production and consumption of such reports. While this approach leans on established methods in rhetoric and text history, it is novel, we suggest, when applied to the topic of aid evaluation. Yet, as Miller has argued, studying reports, lectures, white papers, and other "homely discourse" is "not to trivialize the study of genres; it is to take seriously the rhetoric in which we are immersed and the situations in which we find ourselves" (Miller 1984:155). Following from this, evaluation reports in themselves become key objects of research along with all the other documents relating to them. Using methods from rhetorical criticism and argumentation theory, we investigate the rhetorical strategies employed in these texts, and ask questions about their techniques of framing the issue, as well as about their argumentative structures.

Hypothesis and research questions

Based on our theoretical framework and our past academic work, we developed a strong hypothesis to guide our analysis: *The demand for accountability itself impedes learning. Put strongly, the two are incompatible.* In developing this hypothesis, we distinguished between three levels of analysis:

1. The problem lies in the *evaluation text*: Designed for multiple audiences, both internal and external, it is expected to achieve the two largely contradictory goals of accountability and learning. By analysing the evaluation reports as pieces of text, we ask: Might the problem of learning be solved by writing evaluation reports differently?
2. The problem is the *evaluation process:* The process of commissioning, writing and spreading evaluation reports does not encourage the relevant audiences to use and learn from them. By analysing evaluation processes as examples of knowledge production, we ask: Might the problem be solved by changing the way that evaluation processes are designed?
3. The problem is the *evaluation system* that appears to emphasise accountability as the primary issue, at the expense of learning. By analysing how evaluations are part of a broader political context, we ask: Might the problem be solved by reducing the audit demands and granting more time and space for trial-and-error, innovation, experimentation, and risk-taking within the aid administration?

In order to operationalise these hypotheses, we articulated the following research questions:

- How has learning and accountability been conceptualised and institutionalized historically within Swedish and Norwegian aid evaluation?
- How do the current evaluation units of Sida and Norad deal with the potential contradiction between learning and accountability in aid evaluation?
- How may aid evaluation be organised differently so as to better facilitate learning?

Data and methods

Our aim has been to explore a broad empirical base while at the same time pursuing the specific problem articulated by our hypothesis. Given this, our investigation includes the following elements:

1. A review of relevant international literature that addresses both accountability and learning (including academic research, practice-based research, and publications from aid agencies).
2. A historical mapping of how aid evaluation has been organised in Sweden and Norway.
3. A mapping of available evaluation documents in the databases of Sida, Norad, and biståndsdebatten.se.
4. Selection of 20 evaluation reports for in-depth rhetorical analysis.
5. In-depth interviews with key senior evaluation staff.

Review of existing literature

We undertook systematic searches in key journals and also employed the snowballing method to access as much relevant literature as possible. The review includes both academic research, articles in expert journals (notably *Evaluation* and *American Journal of Evaluation*), and publications from key aid agencies (notably DFID, the World Bank, and OECD-DAC), and relevant publications from Sida and Norad (both evaluation manuals, guidelines, and reports).

The review produced a higher number of publications than expected, several of which directly addressed challenges pertaining to the dual purposes of accountability and learning in evaluation (both in aid and other sectors). The publications took up a number of different positions vis-à-vis our hypothesis. We discerned four main categories of positions. This we in turn helped refine our own hypothesis. (The literature review is presented in chapter 2.)

Mapping of historical changes in aid evaluation in Sweden and Norway

This included a brief overview of the shifting organisational landscape of aid evaluation in both countries, including the mandates, guidelines,

and handbooks developed and how learning and accountability have been handled. This overview serves to highlight how the balance between learning and accountability has been operationalised in different ways.

This historical approach is a key part of our analysis. By comparing across time, we may assess the current situation with a more critical eye. Often, current practice is perceived as natural, rational, and necessary, while past solutions are assumed to be old and obsolete. Yet the most recent is not necessarily the best. Similarly, if something is not working, people tend to expect that they have just not yet found the right solution, and that the means for doing so is to continue moving in the same direction as they are already doing, away from the past. Yet all models may have certain benefits, but they will also have trade-offs. Hence, it is not a given that what was done in the 1990s is now outdated and irrelevant. (The historical mapping is presented in appendix 3 and discussed in chapter 5.)

Selection and analysis of evaluation reports

After going through the three publication databases (Sida's, Norad's, and bistandsdebatten.se), we chose in total 20 reports for an in-depth rhetorical analysis. The majority of the reports were not selected because they were important or significant in themselves. Quite the contrary: We wanted a sample of ordinary reports. We therefore gave most weight to the following criteria: In combination, they should cover the entire historical period with an emphasis on the recent decade; they should cover both countries with a majority from Sweden; they should be commissioned by Sida/Norad; and they should cover aid sectors that have been relevant and significant through the historical period (health in Sweden, natural resources in Norway). The last point was important in order to enable the historical comparison. In addition, we conferred with our reference group and informants to adjust the sample with reports they considered especially interesting. Our ambition was not to make a representative selection of evaluation reports; rather, we were looking for commonalities and genre traits of the evaluation reports *as such*, and how the genre changed (or not) over time. The genre analysis involved the following main questions: How are the reports structured? How do they build their analysis? How are recommendations articulated? How do the reports conceptualise

accountability and learning (if at all)? How do the reports discuss (if at all) what may be learnt from it?

Interviews with senior evaluation staff

Given that our aim was to analyse evaluation both as text, process, and the wider context, we complemented the historical and rhetorical analyses with in-depth interviews with six key senior evaluation staff members. In combination, they covered the whole historical period in both countries. The interviews were semi-structured according to an interview guide distributed in advance. The interviews had three main parts: First, we employed methods from oral history to document the interviewees' own professional trajectory. This provided valuable insight into the historical changes in the field. Second, we invited the interviewees to reflect on the practical work of aid evaluation (commissioning, writing, using, and synthesising evaluation reports). Third, we asked more specifically about the dual purposes of accountability and learning, and invited their reactions to our hypothesis. The interviews were recorded, partly transcribed, and stored according to the ethical guidelines of academic research. We have anonymised all quotes in order to allow our interviewees to speak freely.

Discussions with the Reference Group

This study has benefited greatly from discussions with EBA's staff and Reference Group at several stages during the working process: while preparing the project proposal; in developing the study's practical methodology and design; and discussing early and final drafts. The Reference Group combined a strong and long-standing expertise and experience in aid evaluation, including former heads of evaluation in Sida, Norad, and DFID, and was thus of great value for our study. While we initially considered organizing workshops with evaluation staff during the study period to test our hypothesis, collect data, and verify preliminary conclusions, we instead used the Reference Group to achieve these same objectives. The individual comments from group members to previous drafts and the joint meeting discussions greatly enhanced the quality of our analysis and clarity of our argument, but also contributed valuable data in the form of responses to our claims.

We have incorporated some of these responses as quotes. When doing so, we have for the sake of anonymity not distinguished between quotes from informants and reference group members.

Limitations

The main empirical limitation of our study is that we had to delineate our data collection to the central evaluation units. Both partner organisations and external agencies were thus beyond the scope of our study. Furthermore, in order to better understand how evaluation reports foster learning (or not) in practice, we would have had to interview actors in policy and management positions who are in charge of following up on the evaluation reports' specific recommendations. While this is clearly of great interest, it was outside the scope of a study of our limited size.

As regards our methodology, it might be argued that articulating strong hypotheses runs counter to employing an explorative approach. Yet we will maintain that this has been a most productive combination for such a short study. We could have phrased the hypothesis as a more standard research question ("Is there a contradiction between accountability and learning?"), yet we believe that this would be too open-ended, given that both our own and others' past research clearly indicates that there is indeed a tension, if not contradiction. Furthermore, the hypothesis gave a clear direction and force to our study, especially in our interviews, where it served as a concrete starting-point for discussion that sparked highly interesting and valuable responses. We furthermore conducted our data collection and practical analyses in the same hermeneutical manner as we would normally favour, exploring the empirical material on its own terms and letting this guide how we ultimately structured and concluded the study.

Appendix 2: Analysed evaluation documents

Evaluation manuals and guidelines

Sweden

Sida 1976. *Resultatvärdering, några råd och anvisningar («Metodhandboken»).*

Sida 1985. *Metodhandboken. Metoder för beredning, genomförande och utvärdering av biståndsinnsatser.*

Sida 1994. *Evaluation manual for Sida.*

Sida 1999. *Managing and Conducting Evaluations – Design study for a Sida evaluation manual.*

Sida 1999. *Sida Evaluation Policy.*

Sida 2004. *Looking Back, Moving Forward. Sida Evaluation Manual.*

Sida 2007. *Looking Back, Moving Forward. Sida Evaluation Manual.* 2. edition.

Norway

Norad 1980. *Håndbok for evalueringsspørsmål.*

Utenriksdepartmentet 1992. *Evaluering og resultatvurdering i bistanden. Håndbok for utøvere og beslutningstakere.*

Norad 2016. *Guidelines for the evaluation process and for preparing reports for the Evaluation Department.*

Evaluation reports

Sweden

Sida 2011. *Evaluation of Swedish Health Sector Programme Support in Uganda 2000-2010.* Sida Review 2011:4.

Sida 2008. *Phasing-out Swedish Health Support in Luanda, Angola. A Study of the Evolution of Reproductive and Child Health Services, 2006-2007.* Sida Evaluation 2008:03.

Sida 2007. *Healthy Support? Sida's Support to the Health Sector in Angola 1977-2006.* Sida Evaluation 07/50.

Sida 2006. *Health through Sanitation and Water Programme (HESAWA), Tanzania - Ex-post (Retrospective) Evaluation Study*. Sida Evaluation 06/36.

Sida 2006. *Health Cooperation at the Crossroads: More of the same - or making a difference. Vietnam-Sweden Health Cooperation on Health Policy and Systems Development 2001-2005.* Sida Evaluation 06/02.
Sida 2001. *Tackling Turmoil of Transition. An evaluation of lessons from Vietnam-Sweden Health Cooperation 1994 to 2000.* Sida Evaluation 01/03.
Sida 2000. *Butajira Rural Health Project - An evaluation of a demographic surveillance site.* Sida Evaluation 00/11.
Sida 1993. *Health Through Sanitation and Water - A study from a village perspective.*
Sida 1992. *Maintaining Health - An Evaluation of the Maintenance Project for Rural Health Facilities in Kenya.*
Sida 1992. *Doi Moi and Health - Evaluation of the Health Sector Co-operation Programme between Vietnam and Sweden.*
Sida 1986. *From Hospitals to Health Centers - A Joint Evaluation of Swedish Assistance to Health Sector Development in Kenya 1969- 1985*, parts II and III.
Sida 1974. *Hälsocentraler på landsbygden i Tanzania.* Resultatutvärdering 1.

Norway

Norad 2013. *Facing the Resource Curse: Norway's Oil for Development Programme.* Evaluation Report 6/2012.
Norad 2008. *Evaluation of Norwegian Development Co-operation in the Fisheries Sector*. Evaluation Report 6/2008.
Norad 2007. *Evaluation of Norwegian Power-Related Assistance.* Evaluation Report 2/2007.
Ministry of Foreign Affairs 1990. *General Report on Norwegian Assistance to the Energy Sector of Mozambique.* Evaluation Report 4.90.
Ministry of Foreign Affairs 1990. *Mini-hydropower Plants in Lesotho.* Evaluation Report 1.90.
Ministry of Development Aid 1985. *Lake Turkana Fisheries Development Project.* Evaluation Report 5.85.

Joint evaluations Norway/Sweden

Sida 2011. *Supporting Child Rights. Synthesis of Lessons Learned in Four Countries.* Joint Evaluation 2011/1.
Norad 1988. *Evaluation of the effectiveness of technical assistance personnel.* Evaluation Report 5/88.

Publications on aid evaluation, accountability, and learning

ALNAP Annual Review 2003. *Humanitarian Action: Improving Monitoring to Enhance Accountability and Learning, Meta-evaluation.*

Asian Development Bank 2014. *Evaluation for better results. Independent Evaluation at the Asian Development Bank 10 years.*

Auditor General of Norway 2004. *Riksrevisjonens undersøkelse av effektiviteten av norsk bistand til Mosambik*. Document no. 3:6. Oslo: Riksrevisjonen.

Auditor General of Norway 2011. *Riksrevisjonens undersøkelse av resultatorienteringen i norsk bistand*" Document 3:4. Oslo: Riksrevisjonen.

Carlsson, J. et.al 1999. *Are Evaluations Useful? Cases from Swedish Development Cooperation*. Sida Studies in Evaluation 99/1.

Carlsson, J. & L. Wohlgemuth (eds.) 2001. *Learning in Development Cooperation.* Almqvist & Wiksell International.

Center for Global Development 2009. *When Will We Ever Learn? Improving Lives through Impact Evaluation*. Report of the Evaluation Gap Working Group.

Christopolos, I. et.al 2013. *Swedish Development Cooperation in Transition? Lessons and Reflections from 71 Sida Decentralised Evaluations (April 2011-April 2013). Final Report.* Sida Studies in Evaluation 2013:1.

Christopolos, I. et.al 2014. *Lessons and Reflections from 84 Sida Decentralised Evaluations 2013: a Synthesis Review.* Sida Studies in Evaluation 2014:1.

EBA 2015. *Utvärdering av svenskt bistånd - en kartlägning.* Rapport 02-2015. Expertgruppen for bistandsanalys (EBA).

EuropeAid 2014. *Assessing the uptake of strategic evaluations in EU development cooperation.*

European Evaluation Society 2016. "Forum: Is there a trade-off between accountability and learning in evaluation?", in the newsletter *Evaluation Connections*, February 2016 edition.

Finansdepartementet 2015. *Reglement for økonomistyring i staten.* (Revidert utgave 5.11.2015) https://www.regjeringen.no/globalassets/upload/FIN/Vedlegg/okstyring/Reglement_for_okonomistyring_i_staten.pdf

Forss, K., B. Cracknell & K. Samset 1994. "Can evaluation help an organization to learn?", *Evaluation Review* 18 (5): 574-91.

Forss, K., E. Vedung, S.E. Kruse, A. Mwaiselage and A. Nilsdotter 2008. *Are Sida Evaluations Good Enough? An Assessment of 34 Evaluation Reports.* Sida Studies in Evaluation 2008:1.

Furubo, J.-E. 2003. "The Role of Evaluations in Political and Administrative Learning and the Role of Learning in Evaluation Praxis", *OECD Journal on Budgeting,* Vol. 3/3.

Heider, C. 2016. "Facing Off: Accountability and Learning – the Next Big Dichotomy in Evaluation?", blog post, March 22, 2016. Available at https://ieg.worldbankgroup.org/blog/facing-accountability-and-learning-next-big-dichotomy-evaluation (retrieved 17.12.2016).

International Labor Organization 2005 (revised 2010). "Evaluation policy", available at http://www.ilo.org/eval/Evaluationpolicy/lang--en/index.htm (lretrieved 09.01.2017).

Independent Office of Evaluation of IFAD 2015. *Evaluation manual*. International Fund for Agricultural Development.

Independent Commission on Aid Impact (ICAI) 2014. *How DFID Learns.*

Independent Evaluation Group 2016. *Behind the Mirror. A Report on Self-Evaluation Systems of the World Bank Group*. Washington DC: World Bank Group.

Jones, H. & E. Mendizaba 2010. *Strengthening learning from research and evaluation: going with the grain.* Final report. Rapid Research and Policy in Development. Overseas Development Institute.

Krohwinkel-Karlsson, Anna 2007. *Knowledge and Learning in Aid Organizations – A literature review with suggestions for further studies.* SADEV working paper 2007:1.

Krohwinkel-Karlsson, Anna 2008. *Lär sig Sida mer än förr? En jämförande studie av attityderna till lärande inom Sida idag och för tjugo årsedan,* SADEV Report 2008:1.

Ministry of Foreign Affairs 1993. *Internal Learning from Evaluations and Reviews*. Evaluation report 1/93. Oslo: MFA.

Norad 2013. *Use of Evaluations in the Norwegian Development Cooperation System*. Evaluation Report 8/2012. Oslo: Norad.

Norad 2014. *Can We Demonstrate the Difference that Norwegian Aid Makes?* Evaluation Report 1/2014. Oslo: Norad.

Norad 2016. *Kan lærdommer fornye utviklingspolitikken?* Evalueringsavdelingen: Årsrapport 2015/16.

Norad 2017. The Quality of Reviews and Decentralised Evaluations in Norwegian Development Cooperation. Evaluation Department Report 1/2017. Oslo: Norad.
OECD 1986. *Methods and Procedures in Aid Evaluation.*
OECD 1991. *Principles for Evaluation of Development Assistance.*
OECD 2001. *Evaluation Feedback for Effective Learning and Accountability.* Report series "Evaluation and aid effectiveness", no: 5. DAC working party on aid evaluation.
OECD 2010. *Quality Standards of Aid Evaluation.*
OECD 2013. *The DAC Network on Development Evaluation – 30 years of strengthening learning in development.*
OECD 2016. *Evaluation Systems in Development Co-operation. 2016 Review.* Paris: OECD Publishing.
Ostrom, E. et.al. 2001. *Aid, Incentives, and Sustainability: An Institutional Analysis of Development Cooperation.* Sida Studies in Evaluation 02/01.
Pasteur, K. 2004. *Learning for development: A literature review.* Lessons for Change in Policy and Organisations no. 6. Brighton: Institute of Development Studies.
Riksrevisionsverket 1988. *Lär sig SIDA? En granskaning av SIDAs förmåga att lära sig av erfarenheterna.* Stockholm: Riksrevisionsverket.
Sandström, Sven 1995. "Chapter 2: Evaluation and Learning: Keys to Improving Development Impact", *Evaluation and Development. Proceedings of the 1994 World Bank Conference.*
SADEV 2008. *Reaping the Fruits of Evaluation? An evaluation of management response systems within aid organisations.* SADEV REPORT 2008:7.
Sida 1997. *Using the evaluation tool – a survey of conventional wisdom and common practice at Sida.*
Sida 1999. *Are evaluations useful? Cases from Swedish Development Cooperation.*
Sida 2005. *Sida lär. Sidas syn på lärande.* Avdelningen för personal- och organisationsutveckling.
SASDA 1994. *Improving Monitoring and Evaluation in Swedish Development Assistance.* SASDA working paper no. 4.
SASDA 1994. *Studier av bistånd – Slutrapport från kommittén för analys av utvecklingssamarbete*. SASDA rapport no 8.
Serrat 2010. *Learning from evaluation*. Washington, DC: Asian Development Bank.

Sundvollen Declaration 2013 (The political platform of the Norwegian Solberg government, signed October 7, 2013) https://www.regjeringen.no/no/dokumenter/politisk-plattform/id743014/#utenriks_bistand (last retrieved November 2, 2016).
World Bank 2011. *Self-Evaluation of the Independent Evaluation Group*. Washington, DC: World Bank Group.

Appendix 3: Historical overview of evaluation systems

Sweden

Offices within Sida

1971-1995: The Unit of Results Assessment. SIDA initiated its first evaluation program in 1971 and established a discrete unit to handle the build-up of its evaluation system. In 1974, the program was evaluated and renewed. During these first years, the unit sought to build evaluation capacity both within SIDA and in the recipient countries. A key part of this work in SIDA was to assist programme officers in managing decentralised evaluations. The unit commissioned evaluation reports, published a yearbook in Swedish for a popular audience, and participated in the preparation of SIDA's first methods handbook for evaluation and results assessment (*Metodhandboken*, published in 1976, revised in 1985 and 1988). A third evaluation program was approved in 1988. Between 1988 and 1995, the unit was reorganised into a group under Sida's Planning Secretariat. In 1994, the evaluation group published a new handbook, now titled "*the Evaluation Handbook*".

1995-2008: Secretariat of evaluation and internal audit (UTV). Following a major reorganisation of Swedish development aid, in which several different agencies were combined in "New Sida", the evaluation function was expanded and placed in a new semi-autonomous secretariat (together with the internal audit function) that reported directly to Sida's board of directors. UTV prepared annual evaluation plans, commissioned strategic-, thematic-, country-, and meta-evaluations, published the *Sida Series in Evaluation* and annual reports summarising all Sida's evaluation activities, issued newsletters presenting main findings from recent evaluation reports for a wider audience, and provided methodological support to decentralised evaluations, and engaged in extensive international cooperation, especially through OECD-DAC's evaluation network and joint evaluations with other donors. In 1999, UTV revised Sida's

Evaluation Handbook and prepared a new manual in 2004 titled *Looking Back, Moving Forward.* The revised version from 2007 is currently in use.

2008-2011: Multiple reorganisations of the evaluation function. Following a change in political leadership in 2006 and the appointments of new Director Generals in 2007 and 2010, Sida underwent two major reorganisations in 2008 and 2011, that entailed a 20% reduction in staff. As part of the reorganisations, the evaluation function also underwent multiple changes. First, the internal audit function was moved out of UTV. UTV remained a semi-autonomous secretariat until 2011, when it became a unit under the Department of Organisational Development, following a proposal from the head of the evaluation function. In this new position, UTV sought to integrate its work more firmly into Sida at large: supporting programme officers in undertaking decentralised evaluations, facilitating evaluation as part of learning processes, developing better routines and systems for integrating evaluation in planning and monitoring, following up international initiatives, and taking part in joint evaluations with other donors.

2011-present: Evaluation function, Unit of Planning, Monitoring and Evaluation (PME), the Department of Organisational Development. After the reorganisation in 2011, UTV was gradually disbanded as a distinct organisational entity and the evaluation function was integrated into the Unit of Planning, Monitoring and Evaluation. Budgetary resources and staff members were reallocated to other tasks within the Department of Organisational Development, in effect drastically reducing Sida's evaluation capacity and eroding the systems and routines that had previously been established. While decentralised evaluations proceeded as before across Sida's organisation, only 1-2 staff members worked exclusively with evaluation at Sida's headquarters by 2015. The staff continues to publish the *Sida Studies in Evaluation*, which now mainly consists of decentralised evaluations, and prepares an annual report synthesizing the main findings from the past year's reports. During 2015, evaluation gained a higher priority within Sida, leading to increased budgets and more staff (by December 2016, 5-6 full-time staff members are dedicated to evaluation). In 2016, aid evaluation was singled out as a priority in Parliament's allocation letter to Sida. In

2017, the evaluation staff plans to expand its work to include closer follow-up and methodological support to decentralised evaluations and evaluation consultants, and also experimentation with new and shorter evaluation formats.

External agencies

1993-1994: Secretariat for Analysis of Swedish Development Assistance (SASDA). In 1993, the Swedish government established the Secretariat for Analysis of Swedish Development Assistance (SASDA), an independent commission "appointed with the task of analysing the results and effectiveness of Swedish development aid".[132] In its final report, SASDA recommended establishing a separate evaluation secretariat and link evaluation more directly to planning and monitoring, which they also recommended to expand.

1988-2003: the Expert Group on Development Issues (EGDI). In 1988, this new independent unit was established under the Ministry of Foreign Affairs. EGDI commissioned studies by external researchers on current issues of development policy and strategy, among them the volume *Learning in Development Cooperation* (2001), edited by J. Carlsson and L. Wohlgemuth.

2006-2013: Swedish Agency for Development Evaluation (SADEV). SADEV was established as an independent agency specifically devoted to the topic of aid evaluation. SADEV was expected to produce their evaluations in-house. They prepared both evaluation reports (thematic and strategic) and analytical studies of Swedish aid management, including on the question of learning and use of aid evaluations. While SADEV was funded by a special allocation from parliament and by definition independent, the Ministry of Foreign Affairs increasingly commissioned specific assignments which in effect restricted its independence. Following a critical review by the Swedish Agency for Public Management (Statskontoret), SADEV was closed in 2013.

[132] SASDA 1994. Quote from SASDA's mandate, retrieved from SASDA working paper nr 4, 1994, preword.

2013-present: Expert Group on Aid Studies (EBA). EBA is a government committee under the Ministry of Foreign Affairs with a mandate to evaluate and analyse Sweden's international development assistance. EBA both conducts, commissions, and funds studies on issues with relevance for the Swedish development sector, among them, this present study.

Norway

1977-1983: Office of evaluation and research (Evalfo), NORAD. Evalfo was established in 1977 as a semi-autonomous office with 3-4 staff members, reporting directly to the Director General. The office outlined an upscaling of NORAD's evaluation efforts, which included both establishing a more comprehensive evaluation system and related changes in aid planning and monitoring. In 1981, NORAD's Director General approved a distinct evaluation mandate and an evaluation handbook, both prepared by Evalfo. This included preparing an annual evaluation program, establishing work routines and formal systems for evaluation, commissioning reports from external consultants and making the reports available for the public. Evalfo experienced considerable management support and a high degree of independence.

1984-1990: Evaluation function, 2. Planning Office, Ministry of Development Cooperation (MDC). Following a reorganisation of the Norwegian aid administration, a separate Ministry for Development Cooperation (MDC) was established in 1984, gaining resources and portfolios from both NORAD and the Ministry of Foreign Affairs. As part of this, Evalfo moved from NORAD into the MDC and made part of the 2. Planning Office. While being organisationally closer to the functions of aid planning, strategy, and policy, evaluation received less support and interest from the new NORAD and Ministry leadership. Staff continued to prepare annual evaluation programs, further developed the evaluation routines and systems, and contributed to institutionalise new internal systems of planning and monitoring in Norad, yet they also experienced a budget stagnation and published fewer evaluation reports per year. One main priority was to commission so-called *Country Studies and Norwegian*

Aid Reviews of Norway's main partner countries (ten in total, the last one in 1990). Staff was also active in international evaluation networks, notably in OECD-DAC's expert group on aid evaluation and its work to harmonise evaluation standards and systems.

1990-1997: Evaluation function, 2. Planning Office, Ministry of Foreign Affairs. In 1990, the Ministry of Development Cooperation (MDC) was dissolved and integrated in the Ministry of Foreign Affairs. The evaluation function moved with the 2. Planning Office into the MFA. While still experiencing little interest among the political leadership, the already established evaluation routines and systems continued as before, as did the international network activities (see above). A key priority of the staff during 1990-91 was the preparation of a new and expanded evaluation handbook, *Evaluation and results assessments in aid: Handbook for practitioners and decision makers,* published in 1992. The handbook was later translated into several languages and distributed widely throughout the OECD-DAC.

1997-2003: Unit of Planning and Evaluation (PEV), MFA. Following a shift in government in 1997, the new Minister of Development showed a new and increased interest in evaluation. The evaluation function gained increased budgets, more staff, and higher independence. During 1998-2000, the function was reorganised and made part of the new Unit of planning and evaluation (PEV). This caused an expansion of evaluation activities, including more comprehensive evaluation programs, more and larger evaluation assignments. The minister was particularly interested in fostering learning from evaluations throughout the aid administration. The unit also initiated institutional cooperation with World Bank's Evaluation Group (OEG) on evaluation methods.

2004-present: Evaluation Department in Norad (EVAL). Following a reorganisation of the Norwegian foreign service and aid administration, the evaluation function moved out of the Ministry and became an independent office in Norad, reporting directly to the Secretary Generals of the Ministry of Foreign Affairs and the Ministry of Climate and Environment. The office experienced a considerable expansion in staff and budgets. In 2006, the Office was granted a

distinct mandate asserting their organisational independence, formal role, and practical tasks. EVAL's resources expanded during the first years, before stabilising at approx. 10 staff members and 10 evaluation reports annually. According to its mandate, EVAL should commission evaluations that in combination cover the main areas, sectors, and priorities of Norwegian development aid over the course of 4-5 years. In addition to sector, thematic, strategic, and real-time evaluations, the department has increasingly commissioned evaluations of the Norwegian aid administrative system as such, including the use of evaluations, documentation of results, and strategic planning. In 2015, EVAL's mandate was adjusted to further emphasise its organisational independence and grant the Department more independence, including more responsibility for the communication and follow-up of evaluation reports.

References

Amba, T.A 1998. "Text in an Evaluative Context. Writing for Dialogue", *Evaluation* 4:4; 434–454.

Armitage, L. 2011. “Evaluating aid: An adolescent field of practice”, *Evaluation* 17(3): 274.

Asdal, K. 2008. “On Politics and the Little Tools of Democracy: A Down-to-Earth Approach,” *Distinktion: Scandinavian Journal of Social Theory*, no. 16, pp. 11-27.

Asdal, K. 2011. “The Office: The Weakness of Numbers and the Production of Non-Authority,” *Accounting, Organizations, and Society*, vol. 36, no. 1, pp. 1-9.

Asdal, K. 2015. "What is the issue? The transformative capacity of documents", *Distinktion: Scandinavian Journal of Social Theory* 16(1): 74-90.

Balthasar, A. & S. Rieder 2000. "Learning from Evaluations. Effects of the Evaluation of the Swiss Energy Programme", Evaluation 6(3): 245–260.

Bamberger, M. 1991. «The politics of evaluation in developing countries”, *Evaluation and Program Planning*, 14: 325-339.

Benjamin, L.M. 2008. "Evaluator’s Role in Accountability Relationships: Measurement Technician, Capacity Builder or Risk Manager?", *Evaluation* 14(3): 323-343.

Biggs S. & S Smith 2003. "A Paradox of Learning in Project Cycle Management and the Role of Organizational Culture", *World Development* 31 (10): 1743-1757.

Bjørkdahl, K. 2016. *Expanding the Ethnos. Rorty, Rediscription, and the Rhetorical Labor of Moral Progress.* Ph.D. dissertation, Faculty of Humanities, University of Oslo.

Bjørkdahl, K. (ed.) 2017a (forthcoming). *Rapporten: Sjanger og styringsverktøy.* Oslo: Pax.

Bjørkdahl, K. 2017b (forthcoming). "Den innflytelsesrike nøytraliteten: Rapportsjangeren som demokratiets retoriske reisverk", in Bjørkdahl, K. (ed.) 2017. *Rapporten: Sjanger og styringsverktøy.* Oslo: Pax.

Bjørkdahl, Kristian. 2017c (forthcoming). "Kunsten å ikke lese: Hvorfor så få faktisk leser rapporter", in Bjørkdahl, K. (ed.) 2017. *Rapporten: Sjanger og styringsverktøy.* Oslo: Pax.

Brodén, V.G. 2013. *Aiding Science. Swedish Research Aid Policy, 1973-2008.* PhD thesis. Linköping: The Department of Thematic Studies – Technology and Social Change, Linköping University.

Bøås, M. and D. McNeill, eds. 2004. *Global Institutions and Development: Framing the World?* London and New York: Routledge.

Callon, M., P. Lascoumes & Y. Barthe 2009. *Acting in an Uncertain World: An Essay on Technical Democracy*. Cambridge: The MIT Press.

Cassen, R. & Associates 1986. *Does Aid Work?* Oxford: Oxford University Press.

Cooper, F. and R. Packard (eds.) 1997. *International Development and the Social Sciences: Essays on the History and Politics of Knowledge.* Berkeley: University of California Press.

Cox R.W. ed. 1997. *The New Realism: Perspectives on Multilateralism and World Order*. New York: Macmillan and United Nations University Press.

Cracknell, B.E. 1996. "Evaluating Development Aid. Strengths and Weaknesses", *Evaluation* 2(1): 23-33.

Cracknell, B.E. 2001. "The Role of Aid-Evaluation Feedback as an Input into the Learning Organization", *Evaluation* 7(1): 132-145.

Curtis, D. 2004. "'How we think they think': Thought styles in the management of international aid", *Public Administration and Development* 24: 415–423.

Darnton, R. 1982. "What Is the History of Books?," *Daedalus*, vol. 111, no. 3, pp. 65-83.

Dayton, D. 2002. "Evaluating Environmental Impact Statements as Communicative Action", *Journal of Business and Technical Communication* 16(4): 355-405.

Devitt, A. 2008. *Writing Genres*. Carbondale: Southern Illinois University Press

Ebrahim, A. 2005. “Accountability Myopia: Losing Sight of Organizational Learning”, *Nonprofit and Voluntary Sector Quarterly* 34 (1): 56-87

Eyben, R. 2005. *Donors’ Learning Difficulties: Results, Relationships and Responsibilities.* IDS Bulletin 36:3.

Eyben R. 2013. *Uncovering the Politics of ‘Evidence’ and ‘Results’: A Framing Paper for Practitioners.* Online: www.bigpushforward.net

Eyben, R., I. Guijt, C. Roche & C. Shutt 2015. *The Politics of Evidence and Results in International Development: Playing the Game to Change the Rules?* IDS: Practical Action Publishing.

Feinstein, O.N. 2002. "Use of Evaluations and the Evaluation of their Use", Evaluation 8(4): 433-439.

Fløttum, K. and T. Dahl 2012. ”Different contexts, different ‘stories’? A linguistic comparison of two development reports on climate change”, *Language and Communication* 32: 14.23.

Forss, K 1985. *Planning and evaluation in aid organizations.* Dissertation for the Doctor’s Degree, Stockholm School of Economics. Stockholm: Gotab.

Gasper D. 2000. "Evaluating the `logical framework approach': towards learning-oriented development evaluation", *Public Administration and Development* 20: 17-28.

Gasper, D. A.D. Portocarrero & A.L. St.Clair 2013. "The framing of climate change and development: A comparative analysis of the Human Development Report 2007/8 and the World Development Report 2010", *Global Environmental Change* 23: 28–39.

Haas, Peter M. 1992. "Introduction: Epistemic Communities and International Policy Coordination", *International Organization* 46 (1) 1-35.

Hilgartner, S. 2000. *Science on Stage: Expert Advice as Public Drama.* Stanford: Stanford University Press.

Hirschman, A. 1967. *Development Projects Observed.* Washington, D.C.: Brookings Institution Press.

Hull, M. 2012. *Government of Paper. The Materiality of Bureaucracy in Urban Pakistan.* University of California Press.

Jasanoff, S. 2004. *States of Knowledge: The Co-Production of Science and the Social Order.* London: Routledge.

Jensen, C.B. and B.R. Winthereik 2013. *Monitoring Movements in Development Aid: Recursive Partnerships and Infrastructures.* Cambridge: The MIT Press.

Johnson, Paul L. 1991. "Ray Rist Talks about the IIAS Working Group on Policy and Program Evaluation", in *Evaluation Practice* 12 (1): 45-53.

Kafka, B. 2009. «Paperwork: The State of the Discipline», *Book History* 12: 340-353.

Karlsson, B. 2013. "Writing Development," *Anthropology Today* 29(2): 4-7.

Lehtonen, M 2005. "OECD Environmental Performance Review Programme: Accountability (f)or Learning?", *Evaluation* 11(2): 169-188.

Latour, B. 1999. *Pandora's Hope. Essays on the Reality of Science Studies.* Harvard University Press.

Latour, B. 2010. *The Making of Law: An Ethnography of the Conseil d'Etat.* London: Polity Press.

Lockheed, M.E. 2009. "Evaluating Development Learning: The World Bank Experience", *Evaluation* 15(1): 113-126.

McNeill, D. 1981. *The Contradictions of Foreign Aid.* London: Croon Helm.

McNeill, D. 2017 (forthcoming). "Å skrive rapporter: En dialog", in Bjørkdahl, K. (ed.) 2017. *Rapporten: Sjanger og styringsverktøy.* Oslo: Pax.

Miller, C. 1984. "Genre as Social Action," *Quarterly Journal of Speech*, vol. 70, pp. 151-167.

Milligan, S., S. Bertram and A. Chilver 2015. "The rhetoric and reality of results and impact assessment in donor agencies: a practitioners' perspective" in Jakupec, Viktor and Max Kelly (eds.): *Assessing the Impact of Foreign Aid: Value for Money and Aid for Trade.* Elsevier/Academic Press

Moretti, F. & D. Pestre 2015. "Bankspeak. The Language of World Bank Reports." *New Left Review* 92: 75-99.

Mosse, D. 2005. *Cultivating Development: An Ethnography of Aid Policy and Practice.* London: Pluto Press.

Mosse, D. (ed) 2011. *Adventures in Aidland: The Anthropology of Professionals in International Development*. Berghahn.

Nielsen, S.B. & D.M. Winther 2014. "A Nordic evaluation tradition? A look at the peer-reviewed evaluation literature", *Evaluation* 20(3): 31-331.

O'Connor, A. 2001. *Poverty Knowledge. Social Science, Social Policy, and the Poor in Twentieth-Century U.S. History.* Princeton University Press.

Odén, B. 2006. *Biståndents idéhistoria. Från Marshallhjälp til Milleniemål.* Studentlitteratur.

Olejniczak, K., E. Raimondo & T. Kupiec 2016. "Evaluation units as knowledge brokers: Testing and calibrating an innovative framework", *Evaluation* 22(2): 168-189.

Patton, M.Q. 1984. "An Alternative Evaluation Approach for the Problemsolving Training Program: A Utilization-Focused Evaluation Process", *Evaluation and Program Planning*, 7:189-192.

Patton, M.Q 2008. *Utilization-Focused Evaluation.* 4. edition. Sage Publications.

Patton, M.Q. 2015. "The Sociological Roots of Utilization-Focused Evaluation", *The American Sociologist* 46: 457-462.

Picciotto, R. 2008. *Evaluation independence at DFID: An independent assessment prepared for IACDI.* Independent Advisory Committee for Development Impact.

Power, M. 1994. *The Audit Explosion.* London: Demos.

Power, M. 1997. *The Audit Society. Rituals of Verification.* Oxford University Press.

Power, M. 2000. "The Audit Implosion: Managing Risk From the Inside", ICAEW.

Reeger, B. et.al. 2016. "Exploring ways to reconcile accountability and learning in the evaluation of niche experiments", *Evaluation* 22(1): 6-28.

Reinertsen, H. 2016. *Optics of evaluation. Making Norwegian foreign aid an evaluable object, 1980-1992.* Ph.D. dissertation. Faculty of Social Sciences, University of Oslo.

Reinertsen, H. 2017 (forthcoming). "40 år og like langt? Om etableringen av evalueringsoptikker i norsk bistand", in Bjørkdahl, K. (ed.) 2017. *Rapporten: Sjanger og styringsverktøy.* Oslo: Pax.

Riddell, R. 2007. *Does Foreign Aid Really Work?* Oxford University Press.

Riles, A. 2006. *Documents: Artifacts of Modern Knowledge*. University of Michigan Press.

Rist, R. & K. Joyce 1995. "Qualitative research and implementation evaluation: a path to organizational learning", *International Journal of Educational Research* 23 (2): 107-190.

Rottenburg, R. 2000. "Accountability for Development Aid", in Kalthoff, H., Rottenburg, R. and Wagener, H.J. (eds), *Facts and Figures. Economic Representations and Practices*, Metropolis, Marburg, pp. 143-173.

Rottenburg, R. 2009. *Far-Fetched Facts. A Parable of Development Aid*. Cambridge: The MIT Press.

Schaumburg-Müller, H. 2005. "Use of Aid Evaluation from an Organizational Perspective", Evaluation 11(2): 207-222.

Schwartz, C. 2006. *Evaluation als modernes Ritual. Zur Ambivalenz gesellschaftlicher Rationalisering am Beispiel virtueller Universitätsprojekte.* Hamburg: LIT Verlag.

Schwarz, C. & G. Struhkamp 2007. "Does Evaluation Build or Destroy Trust? Insights from Case Studies on Evaluation in Higher Education Reform", *Evaluation* 13(3): 323-339.

Shutt, C. 2016. *Towards an Alternative Development Management Paradigm?* EBA Report 07/2016.

Simensen, J., K.A. Kjerland, F. Liland and A.E.Ruud 2003. *Norsk utviklingshjelps historie 1952-2002.* Bergen: Fagbokforlaget.

Stirrat, R.L. 2000. “Cultures of Consultancy”, *Critique of Anthropology* 20(1): 31-46.

Strathern, M. (ed) 2000. *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*. London: Routledge.

Uusikulä, P & P. Virtanen 2000. "Meta-Evaluation as a Tool for Learning. A Case Study of the European Structural Fund Evaluations in Finland", *Evaluation* 6(1): 50–65.

Vähämäki, J. 2015. "The results agenda in Swedish development cooperation: cycles of failure or reform success?", in Eyben, R. et.al (eds.) *The Politics of Evidence and Results in International Development: Playing the Game to Change the Rules?* IDS: Practical Action Publishing.

Vähämäki, J. 2017. *Matrixing Aid. The Rise and Fall of ‘Results Initiatives’ in Swedish Development Aid.* Doctoral Thesis in Business and Administration at Stockholm University. Stockholm: Stockholm Business School.

Vedung, E. 1995. ”Utvärdering och de sex användningarna”. In Rombach, B. and Sahlin-Andersson, K. (eds.). *Från sanningssökande till styrmedel: Moderna utvärderingar i offentlig förvaltning.* Stockholm: Nerenius & Santérus förlag.

Vedung E. 2010. "Four Waves of Evaluation Diffusion", *Evaluation* 16(3) 263–277.

Weiss, C. 1993. “Where politics and evaluation research meet”, *Evaluation Practice* 14 (1): 93-106.

Winther, T. 2015. "Impact evaluation of rural electrification programmes: what parts of the story may be missed?", *Journal of Development Effectiveness* 7(2): 160-174.

White, J. 1974. *The Politics of Foreign Aid*. Bodley Hea