EBA

03
2016

CAPTURING COMPLEXITY AND CONTEXT: EVALUATING AID TO EDUCATION

Joel Samoff, Jane Leer, Michelle Reddy

Capturing Complexity and Context: Evaluating Aid to Education

Joel Samoff, Jane Leer, Michelle Reddy

Stanford University

Rapport 2016:03
till
Expertgruppen för biståndsanalys (EBA)

Acknowledgement: Effective research is always a collective product. Ours reflects the experiences, expertise, and insights of colleagues in Tanzania, Bénin, Nepal, Sweden, France, Denmark, Norway, and the U.S. We are particularly grateful to the educators in aid-receiving countries who made our case studies possible and to our EBA reference group who provided timely comments and challenges. Margaret Irving contributed to our initial work.

ISBN 978-91-88143-14-3

Printed by Elanders Sverige AB
Stockholm 2016

Cover design by Julia Demchenko

From Kilimanjaro coffee farmers in Tanzania to militant bus drivers in Ann Arbor Michigan to the education activists of South Africa and Namibia, the orienting concern of **Joel Samoff's** work has been understanding how people organize themselves to transform their communities. With a background in history, political science, and education, he studies and teaches about development and underdevelopment. Consulting Professor in the Stanford University Center for African Studies, he has also been a faculty member at the Universities of California, Michigan, and Zambia, and he has taught in Mexico, South Africa, Sweden, Tanzania, and Zimbabwe. He received an honorary doctorate from the University of Pretoria. Concerned with the links between research and public policy, he works regularly with international agencies involved in African education.

Basic Education Learning Research Specialist at Save the Children U.S., **Jane Leer** is responsible for evaluation design, data analysis, and research capacity development. She completed the Stanford University MA program in International Comparative Education and International Education and Policy Analysis. Earlier she worked for several years on education projects in Latin America, initially with an NGO in rural Nicaragua and subsequently as a research assistant at the Inter-American Development Bank. Her recent research includes an exploration of the determinants and implications of participation in cross-national achievement tests. Her most recent publication reports on a difference-in-differences analysis of the effects of decentralization in Indonesia on education outcomes.

A PhD candidate at Stanford University in International Comparative Education and Organizations, **Michelle Reddy's** research interests center on innovation in peacebuilding, development and humanitarian aid, as well as organizations and civil society networks. She is currently a Fellow at the Stanford Center for International Conflict Resolution and Negotiation. Earlier she co-launched the Paris School of International Affairs at Sciences Po Paris, where she was Assistant Dean. She worked on research, partnerships, communications and program design and management for universities, NGOs, and the United Nations for seven years in Paris, Dakar, and New York. She is a graduate of Columbia University and Boston College.

Table of contents

Preface

One important challenge for development aid lies in the ability to (directly or indirectly) reinforce human capital in low- and middle-income countries, thereby positively affecting economic growth, and, ultimately, to achieve poverty reduction. It is hardly possible to envisage long-term poverty reduction in the world's low- and middle-income countries that is not preceded by strengthened education systems and a more educated population.
The links between education and economic growth, income distribution and poverty reduction are well established. On top of this, education is also a basic human right and a foundation for a more sustainable and inclusive society.

The central and prominent role of education in global development has recently been confirmed by Sustainable Development Goal 4: "Ensure inclusive and equitable quality education and promote lifelong learning". To increase the prospects of achieving the global goal of education for all, effective, good quality education policies, strategies and programmes must be in place.

The difficult part is finding out what type of intervention is likely to work best in a given community or school. There are also many context-specific problems in the education sector that need to be addressed, such as low school attendance, ineffective pedagogy and unsatisfactory school performance in terms of test scores. Studies and research conclude that many children in low- and middle-income countries leave the school system without being able to read simple texts or perform simple mathematical exercises.

In development research, education is repeatedly cited as crucial from a variety of perspectives. At the same time, this sector has not been prioritised in Swedish development aid, despite substantial and alarming needs in low- and middle-income countries and despite the lack of funding for education systems.
Donors and the research community on international education have built up a considerable knowledge base, with hundreds of evaluations and impact studies with (potentially) important conclusions to draw on for effective future investment in the sector. However, the question remains how accessible and useful this knowledge base is, and also whether it is actually used by policy-makers and officials deciding on aid to education. This was the starting point for the Expert group for Aid Studies when it decided to commission two synthesis evaluations on aid to education.

In this report, Professor Joel Samoff, Jane Leer and Michelle Reddy from Stanford University have taken a broad, holistic approach, addressing the question of what we can learn from evaluations undertaken in aid projects and programmes (focusing on aid to education). The team has reviewed and synthesised a diverse sample of evaluations from a large number of national and international donors and agencies. Key conclusions in the report stress the importance of context, effective inclusion of the surrounding community and the importance of taking complexity into account in the analysis of the aid relation. The authors conclude that the delivery of various 'inputs' (computers, school books, more teachers, schools, etc.) is rarely enough to achieve expected results, and that aid projects and programmes need to be more holistic, seeing education as an inclusive process and a system.

The issues of sustainability and local ownership are described as continuously important challenges, and participation is strongly emphasised along with the need for more appropriate time horizons in projects and programmes. The authors corroborate conclusions drawn in previous research when they conclude that "reaching the difficult to reach remains beyond reach".

The authors also argue that evaluations rarely promote learning and seldom contribute new knowledge. With some exceptions, the reviewed evaluations did not, for instance, summarise the findings of previous evaluations in which similar/comparable projects where analysed. This is highly likely to affect lesson-learning, making it probable that mistakes are repeated over and over again.

This report, together with the simultaneously published EBA report by Paul Glewwe, Amy Damon, Suzanne Wisniewski and Bixuan Sun (2016:02), contains important lessons for future Swedish aid to education, but also conclusions of importance for aid effectiveness in general and for the work on evaluation of aid projects and programmes.

The work on this report has been conducted in dialogue with a reference group chaired by Dr Kim Forss of the EBA. The analysis and conclusions expressed in this report are solely those of the authors.

Stockholm, May 2016

Lars Heikensten

Sammanfattning

Biståndsgivare ser regelmässigt över sin policy, sina prioriteringar och sina metoder för att försöka bedöma vilken roll de har och vilka resultat de leder till. Formella utvärderingar av utbildningsbistånd har blivit vanligare, mer systematiska och viktigare för de efterföljande besluten om politik och programplanering. Utvärderingarna har blivit en egen genre i utvecklingslitteraturen.

Vad kan vi lära oss av denna stadigt växande volym utvärderingar? Har utvärderingarna underlättat evidensbaserat beslutsfattande om politik och programplanering? Har biståndsmottagarna använt utvärderingarna för att förbättra sina metoder?

Såväl utbildningen i fattiga länder som det externa utbildningsbiståndet har många syften, många former och många kontexter. Utvärderingarna har olika mål, tillvägagångssätt och målgrupper. En informerad och informativ syntesutvärdering, som grundas på en bred läsning, måste därför både undersöka och belysa olika teman som är relevanta för dessa målgrupper och samtidigt ta upp det som kan vara problematiskt. Brett grundade insikter är mer användbara för både praxis och politik än försök att konstruera ett genomsnitt utifrån disparata och ojämförbara biståndsinsatser, där man riskerar att sudda ut viktiga skillnader, missa kontextuell komplexitet och få ett resultat som inte är särskilt användbart för någon av de tänkta målgrupperna. Något som ytterligare komplicerar arbete med syntesutvärderingar är utvärderingars begränsade spridning och att de i praktiken sällan diskuteras. Det är inte ovanligt att biståndsgivare beställer utvärderingar som därefter förblir relativt okända och knappt användbara för de som biståndet skulle hjälpa och som knappt verkar ha något inflytande på biståndets genomförande.

Globalt och internationellt pågår nu omfattande omvärderingar med koppling till bistånd och utveckling. Runtom i världen omprövas och sätts nya utbildningsmål och indikatorer, samtidigt som biståndsgivarna omprövar och ser över prioriteringar och metoder. Det är därför dags att ompröva även utvärderingsarbetet, från våra uppfattningar om utvärdering till de metoder som ska användas.

Genomgång och syntes

I sin strävan att förbättra både utbildning och bistånd har Expertgruppen för biståndsanalys (EBA) beställt den här syntesen av utvärderingar av biståndsfinansierade utbildningsinsatser. Komplexitet och kontext är viktiga faktorer som bildar ramen för vår genomgång. Utbildning, bistånd och utvärdering är alla mångfasetterade företeelser och de kräver därför ett mångfasetterat och flerdimensionellt angreppssätt.

Vi börjar med att gå igenom viktiga frågor som berör utvärdering, bland annat förväntningarna på den roll utvärderingarna kan spela och den ökande preferensen för kvasiexperimentella och experimentella tillvägagångssätt. Därefter går vi vidare till de huvudsakliga resultaten av vår genomgång, som berör skärningspunkterna mellan bistånd och utbildning och utvärderingsprocessen. Vi undersöker också biståndskontexten och avslutar rapporten med observationer om den roll som utvärderingar av biståndsfinansierade utbildningsinsatser kan spela. Här noterar vi vikten av en differentierad utvärderingsstrategi som matchar olika tillvägagångssätt med specifika behov, syften och målgrupper.

Vår genomgång och syntes riktar sig till flera överlappande men distinkta målgrupper, som alla har sina egna erfarenheter och expertkunskaper. Några av de frågor vi tar upp kommer att vara nya för vissa läsare och mycket välbekanta för andra. Vi har strävat efter en rimlig balans, och vi vill uppmana läsaren att fokusera på de delar av rapporten som han eller hon finner mest utmanande och mest användbara.

Vad fungerar?

Även om alla inblandade naturligtvis vill ha svar på frågan ”*Vad fungerar?*” så är inte det någon fruktbar frågeställning för en genomgång av utvärderingar av biståndsfinansierade utbildningsinsatser. Det är helt enkelt så att en lovande intervention kan leda till de avsedda målen i ett sammanhang, men inte i ett annat, och i ett tredje sammanhang kan den få oönskade konsekvenser. Den kan också framstå som effektiv för finansiärerna men inte för utförarna, eller för utförarna men inte för utvärderarna. En användbar syntes måste därför ta hänsyn till komplexitet och kontext.

Det är mer produktivt att fråga vad som fungerar för vem, under vilka omständigheter och på vilka villkor. Detta i sin tur kräver att man undersöker mer situationellt definierade specifikationer av

framgång. Inte nog med att ett utbildningsinitiativ kan förbättra resultaten i ett sammanhang, men inte i ett annat – samma initiativ kan också ses som framgångsrikt ur ett perspektiv (t.ex. provresultat) men misslyckat ur ett annat (t.ex. att kvinnor hoppar av).

Flera andra aspekter på komplexitet försvårar arbetet med att syntetisera utvärderingar och skapa klarhet kring utbildning och biståndseffektivitet. Utvärderare och forskare försöker ofta undvika dessa komplexiteter genom att förenkla sina antaganden – ”allt annat lika” – eller genom att förpassa dem till bedömningens marginal och sedan direkt eller indirekt hålla dem konstanta. Med denna typ av tillvägagångssätt försöker man få en klarare bild genom att ta det beteende eller förhållande man vill undersöka ur sitt sammanhang. Risken med detta är att bilden visserligen blir klarare men också mer begränsad, ofta till den grad att man inte kan dra rimliga slutsatser som kan vara till hjälp för biståndsgivare och -mottagare.

Vårt tillvägagångssätt är det motsatta, då vi insisterar på att fenomen måste förstås i sin kontext.

Felaktiga premisser

Utvärderingar krävs, med få undantag, av nästan alla biståndsprogram. Förutom att ge bekräftelse på att stödet kopplats till de angivna målen och att finansieringen använts på rätt sätt, förväntas utvärderingarna bidra till att förbättra biståndsprocessen. Den logiken vilar tydligt på tre premisser som är lockande, vid första anblicken övertygande, men som inte har särskilt mycket stöd i forskningen. För det första, och trots omfattande påståenden om vikten och värdet av att lära av erfarenheter, finns det inte mycket bevis för att man direkt lär sig något av de erfarenheter som rapporteras i utvärderingarna och man ser sällan spår av kunskaper som ackumulerats genom de utvärderingar som gjorts över åren. För det andra uppfattas utvärderingar regelmässigt som tillämpad forskning och som genererar relevant kunskap för en evidensbaserad policy. Evidensbaserad policy är ett lockande begrepp, men det finns inga tillgängliga bevis som stöder utgångspunkten att utvärderingar skulle spela en viktig roll för att generera kunskap som direkt används för att utforma politiken. För det tredje uppfattas ofta utvecklingen av den offentliga politiken som en i stort sett rationell och linjär process. I den mån utvärderingar bidrar till politikens utformning är det dock på kaotiska, motstridiga och ofta svagt sammanlänkande vägar.

Redan en snabb titt på dessa tre felaktiga premisser visar tydligt klyftan mellan den roll som utvärderingarna påstås spela – att generera kunskap som gör att man kan lära sig av erfarenheter, vilket i sin tur förbättrar såväl bistånd som utbildning i politik och praktik – och den roll som utvärderingarna faktiskt kan spela. Vårt syfte med att påpeka denna klyfta är inte att förringa svårigheterna med att forma och optimera en rationell politik. Vi vill snarare framhålla att dessa begränsningar måste erkännas, och mana till ödmjukhet i fråga om vad man kan få kunskap om, hur kunskap genereras och hur kunskap används.

En framväxande standard

Vi ser en konvergens, om än inte enhällig, mot ett visst tillvägagångssätt, nämligen effektutvärderingar, om möjligt randomiserade kontrollerade studier (RKS). Uppskattningsvis 150 miljoner US-dollar användes till RKS-utvärderingar av utbildningsprogram under 2013.

Användningen av RKS är inte på något sätt oomstridd. För det första är den här typen av studier kostsam. För det andra är randomisering omöjligt eller extremt svårt i många, för att inte säga de flesta, utbildningskontexter i fattiga länder – av praktiska, politiska och etiska skäl. Det praktiska problemet ligger i att utbildningsinitiativ och utbildningsreformer oftast genomförs på sätt som är svåra att anpassa till de krav som experimentliknande effektbedömningar ställer. Det politiska problemet ligger i att en ojämn fördelning av resurser, i det här fallet bättre utbildningsmöjligheter, kräver en politisk logik och politisk legitimitet. Det räcker inte med specifikationerna från projektledaren för den experimentella utvärderingen. Det etiska problemet har tre komponenter. Slumpvisa tilldelningar är oförenliga med de koncept om preferens och val som studerande, föräldrar och samhället värderar. I kontexter där det finns anledning att tro att vissa skolor eller elever kan gynnas mer än andra av ett visst program blir slumpvisa tilldelningar också etiskt problematiska. RKS jämför ofta en innovation eller en reform med kontrollgruppens ”ingen förändring”, ett tillvägagångssätt som inte uppfyller de etiska krav som ställs på jämförelser av alternativa erfarenheter.

För det tredje kan inte en metod som används inom hälso- och sjukvårdssektorn för att skydda personer vid experimentella behandlingar fungera oproblematiskt på utbildningsområdet, där skillnaderna mellan skolor och samhällen i fråga om institutionell

kapacitet och resurser, liksom samhällspolitiska och kulturella skillnader, innebär att programgenomförandet (behandlingen) sällan är stabil eller gemensam för de olika sammanhangen, även om de väljs ut slumpmässigt. För det fjärde kan man hävda att utvecklingen på utbildningsområdet *inte bör* uppfylla de krav som RKS ställer. Skillnaderna när det gäller programgenomförande är betydande, och bör till och med uppmuntras, snarare än dämpas i strävan efter en stabil behandling. För det femte är de resultat man får av en RKS, liksom av alla typer av utvärderingar, specifika för kontexten och för de villkor under vilka det utvärderade programmet genomförs.

För det sjätte har en relativt färsk granskning av sex metautvärderingar av utvärderingar av utbildningsprogram i låginkomstländer lett till ett ifrågasättande av antagandet att en stor mängd effektbedömningar med RKS kommer att kunna identifiera vilka interventionsformer eller lärandestrategier som är att föredra och lämpliga i ett vidare perspektiv. Ett liknande tillvägagångssätt användes i alla sex genomgångarna, och författarna fann nästan ingen överlappning i de slutsatser som dragits av utvärderingarna – man fann dramatiska motstridigheter där man förväntat sig konsensus. Detta resultat bevisar ytterligare att oavsett hur många utvärderingar som görs och hur mycket uppgifter som samlas in, så kan man inte få fram ett facit på vad som fungerar för utbildning. Det är faktiskt så att sökandet efter ett facit, en uppsättning standardmetoder eller praxis, inte är produktivt. Lärande är en deltagarstyrd, interaktiv och dynamisk process, djupt sammanflätad med de politiska, ekonomiska och historiska kontexter där formell och informell utbildning sker.

Såväl RKS begränsningar som de praktiska, finansiella och etiska problemen med att genomföra dem leder till slutsatserna att även om effektbedömningar och RKS kan vara användbara för att utvärdera biståndsfinansierade utbildningsinsatser så är användningsområdet i praktiken begränsat, och att varken RKS eller effektbedömningar mer generellt är den standard mot vilken andra utvärderingsmetoder bör bedömas.

När metoden avgör resultatet

Den senaste tidens forskning om fattigdom och tillväxt i Afrika visar tydligt att det finns risker med att förlita sig på ett enda tillvägagångssätt eller en forskningsmetod och att anta att om metoden är korrekt så måste också resultatet och rekommendationerna vara de rätta. För att minska dessa risker har vi i denna syntes använt oss av flera metoder och tillvägagångssätt snarare

än att föredra en enda metod, hur vetenskaplig den än verkar vara. För att minska risken för bias från utvärderarens sida krävs interaktion med utbildare, beslutsfattare och samhällen, inte distansering från dem. Systematisk och kritisk hänsyn till komplexitet och kontext är avgörande när man ska bedöma nyttan med ett föreslaget tillvägagångssätt eller en metod och dess resultat. Det är beaktande av historia parallellt med kvantitativa uppgifter, av utbildarnas, elevernas och de utomstående observatörernas synpunkter, och av erfarenheter i kombination med statistisk analys som gör ett visst tillvägagångssätt vetenskapligt.

Ett integrerat angreppssätt

Vad kan vi lära oss av utvärderingar? Vårt fokus ligger på utvärderingar mer samlat, inte på enskilda utvärderingar. Vi undersöker inte om en enskild utvärdering ger tydliga resultat som skulle kunna vägleda insatserna. I stället undersöker vi vad man kan lära av det breda utbud av utvärderingar som genomförs inom ramen för biståndsrelationer. Eftersom vi inser att väl underbyggda utvärderingsresultat inte kan förbättra utbildningen om inte slutsatser tillämpas, undersöker vi också hur utvärderingarna används.

Vi började med att göra en omfattande sökning efter utvärderingar av utbildningsinsatser som beställts av internationella och nationella biståndsfinansierande organ, OECD:s direktorat för utvecklingssamarbete, UNICEF, utbildningsinriktade organisationer i det civila samhället samt framstående utbildningsinriktade forskningsinstitut och konsultbyråer. En princip för arbetet var att ta fram en uppsättning utvärderingar som skilde sig åt med avseende på tillvägagångssätt, beställande myndighet, specifikt fokus och involvering av biståndsmottagare. Vi strävade alltså efter maximal diversitet, inte kvantitet. Sökandet resulterade i en första lista på 80 utvärderingar. Bland dessa valde vi ut 40 utvärderingar för en mer ingående genomgång. Efter att ha granskat dessa och andra listor är vi övertygade om att den uppsättning utvärderingar vi valt ut på ett rimligt sätt återspeglar mångfalden av utvärderingar av biståndsfinansierade utbildningsinsatser. Från den större uppsättningen utvärderingar valde vi dessutom ut tre för en mer djupgående bedömning och på flera nivåer, nämligen biståndsfinansierade insatser i Tanzania, Nepal och Benin.

Vår syntes är blygsam. Den syftar till att få en djupgående och detaljerad analys snarare än till att identifiera och klassificera varje utvärdering som någonsin gjorts. Så vitt vi vet är det den första

syntesen som omfattar ett såpass diversifierat urval av utvärderingar (med avseende på metod, typ av politik och program som utvärderats, finansierande organ, länder och kontexter) och som fokuserar på en uppsättning utvärderingar snarare än på ett fåtal välgrundade utvärderingar av enskilda insatser.

Utvärderingar av utbildningsbistånd i fattiga länder

I de utvärderingar som gåtts igenom är det flera observationer som framstår som särskilt tydliga.

Effektiva utbildningsinsatser når bortom skolorna. Utvärderingarna av biståndsfinansierade utbildningsinsatser bekräftar och bevisar tydligt att effektiva utbildningsprojekt når bortom själva insatserna (input) och bortom skolorna. Ett tydligt exempel på det är insatserna för att åstadkomma ”education for all”. De effektivaste strategierna för att öka skolinskrivningen verkar vara att minska kostnaderna för familjer i kombination med ihållande insatser för påverkan och medvetandegörande.

Det räcker inte med input. De flesta biståndprogrammen fokuserar på input i form av visst stöd. Det är sällan som biståndsprogram bygger in tillhandahållandet av input i ett större ramverk som tar hänsyn till vilket stöd som behövs för att inputen ska användas väl, vem som ska ansvara för att ta emot och förvalta inputen, vilket fortlöpande stöd som kan behövas (t.ex. teknisk assistans och underhåll), hur stödet ska integreras i det nationella och lokala utbildningssystemet, eller reaktionerna från lärare, elever och samhällen. Biståndsprogram som helt eller primärt fokuserar på input är mindre effektiva än de som tar utgångspunkt i en helhetssyn på utbildning som en process och ett system, och där förståelsen för denna helhetssyn är inbyggd i biståndsprogrammet. Trots att utvärderare noterar detta problem, kan de i vissa fall bidra till problemet. Det är sällan utvärderare försöker minska, eller ens ta upp klyftan mellan de bredare utvecklingsmålen (fattigdomsbekämpning, social integrering, mänskliga rättigheter, hållbar utveckling) och de utbildningsinsatser som får stöd.

Effektivt externt stöd når bortom utbildningsministeriet. På samma sätt som fokuseringen på input kan vara begränsande, kan även koncentrationen på utbildningsministerierna vara det. De utländska biståndsmedel som är mest effektiva när det gäller att förbättra utbildningen når bortom centralmakten vid utbildningsministerierna.

Lokalt ägarskap i utbildningsinnovationer är grundläggande: men utvärderas sällan. Vikten av lokalt ägarskap är känd sedan länge och

framhålls ofta i biståndslitteraturen. I utvärderingarna konstateras ofta att insatser där känslan av lokalt ägarskap är stark har mycket större sannolikhet att vara effektiva, eller mer effektiva, mer inkluderande eller mer varaktiga än insatser som de inblandande betraktar med viss distans och kanske med en känsla av att de införts eller påtvingats utifrån. Trots detta är det sällsynt att biståndsfinansieringen uttryckligen lägger fokus på att utveckla, främja och finansiera en stark känsla av lokalt ägarskap i utbildningsinsatser som får stöd. Likaså är det få utvärderingar som granskar eller bedömer det lokala ägarskapet grundligt och systematiskt.

Det är helt avgörande att man erkänner den inneboende och kraftfulla spänningen mellan lokalt ägarskap och det finansierande organisationernas intressen och mål. Frågan gäller var den yttersta kontrollen och auktoriteten ska ligga. För att stödmottagare ska kunna utveckla ett starkt lokalt engagemang och ansvarstagande för biståndsfinansierade utbildningsinsatser krävs att de har en betydande kontroll över både verksamheten och finansieringen. De finansierande organen har dock egna mål och ansvars- och redovisningslinjer och kan vara ovilliga eller oförmögna att lämna över ansvaret till mottagarsidan.

Att nå ut till dem som är svåra att nå är fortsatt svårt. De utvärderingar vi har gått igenom bekräftar problemen med att nå ut med utbildningsmöjligheter till de befolkningsgrupper som är svåra att nå, och som fortfarande i hög grad är uteslutna från biståndsfinansierade utbildningsprojekt. Biståndsfinansiering som är avsedd att minska ojämlikhet kan i praktiken komma att omlokalisera densamma.

Centralisering trots decentralisering. Tidigare har Världsbanken och andra finansierande organ betraktat decentralisering – överföring av befogenheter och ansvar från central till lokal nivå – som en viktig del av utbildningsreformer. I många länder har dock den vanligaste praktiken på utbildningsområdet varit dekoncentrering, vilket innebär att vissa tjänstemän och poster omlokaliseras från centrala till regionala eller lokala utbildningsmyndigheter, utan att makt och befogenheter överförs till samma myndigheter i någon större utsträckning.

Utvärderingarna bekräftar att de flesta i de biståndsmottagande länderna (och i större delen av världen) anser att utbildningsområdet kräver en stark central myndighet och att decentralisering i praktiken är ganska ovanligt, men vad kan de mer

berätta om decentraliseringen? För det första att decentralisering är en viktig del av de officiella strategierna för utveckling på utbildningsområdet. För det andra bekräftar utvärderingarna att decentralisering kan ske i många former och i olika omfattning. För det tredje finns det betydande bevis för att decentralisering, trots de förväntade fördelarna, kan förvärra befintliga ojämlikheter mellan skolor och samhällen. För det fjärde har det visat sig att, samtidigt som decentraliseringsargumenten framhåller egenmakt och lokalt ansvarstagande, kan det i praktiken vara svårt att åstadkomma ett meningsfullt deltagande på lokal samhällsnivå, och att deltagandet ofta är begränsat till ekonomiska bidrag eller insatser för underhåll av skolor. För det femte stöter decentraliseringsstrategierna ibland på lokalt motstånd. För det sjätte – trots att många utvärderingar framhåller vikten av decentralisering, är det få som uttryckligen tar upp den som en del av utvärderingen eller undersöker hur biståndsmyndigheterna skulle kunna underlätta decentraliseringsprocessen.

Hållbarhet är viktigt, men utvärderas inte systematiskt. I september 2015 antog Förenta nationerna formellt ett antal mål för hållbar utveckling, de hållbara utvecklingsmålen. Men trots att de finansierande organen upprepar sina förväntningar på att biståndsfinansierade utbildningsinsatser ska vara hållbara, saknar biståndsprogrammen i praktiken, generellt sett, antingen uttryckliga punkter om vad som krävs för denna hållbarhet, eller finansiering som öronmärks för hållbarhetsarbete. Det är inte heller särskilt förvånande att många utvärderingar inte tar upp frågan om hållbarhet på något systematiskt sätt.

Information, evidens, data och indikatorer. Behovet av bättre informationshantering, data och indikatorer återkommer ofta i de utvärderingar vi har gått igenom. Det finns några viktiga undantag, men de flesta av utvärderingarna påpekar luckor och andra problem i de uppgifter om utbildning som finns tillgängliga. Men, det är förvånansvärt få av utvärderingarna som direkt tar itu med uppgiftsproblemen, antingen genom att samla in egna allmänna utbildningsuppgifter eller genom att utveckla strategier för att bearbeta bristfällig data. De flesta utvärderingarna integrerar inte heller de sannolikt omfattande felmarginaler som finns i tillgängliga utbildningsdata i sina resultat.

Något som generellt sett inte heller tas upp i de utvärderingar vi har gått igenom är avvägningarna mellan å ena sidan ökade

ansträngningar för att samla in mer och tillförlitligare utbildningsuppgifter, och å andra sidan insatser som inriktas på att använda ett mycket mindre antal befintliga indikatorer på ett bättre sätt. Utvärderingarna undersöker heller inte hur de finansierande organen skulle kunna komma vidare om de baserade stödprogram och utvärderingar på de begränsade, och inte sällan ofullständiga och inkonsekventa, uppgifter som de stödmottagande utbildningsministerierna regelmässigt använder sig av för att förvalta utbildningssystemen.

Vikten av institutionell kunskap och lärande hos finansierande organisationer. De utvärderingar vi har gått igenom ger ett starkt stöd för en välkänd rekommendation: Det behövs omfattande institutionell kunskap och lärande bland de finansierande organisationerna. Den största utmaningen när det gäller att förbättra biståndseffektiviteten ligger inte i att förvärva eller dokumentera kunskap, utan i att göra det möjligt för och att uppmuntra organisationerna att använda den kunskap som redan finns. Utvärderingarna i genomgången tog upp behoven av data och uppgiftsinsamling, men de innehöll inte någon analys av kunskapsdelning i nätverk eller i partnerskap mellan organisationer.

Utbildning, bistånd och utvärderingar

Vad har då dessa utvärderingar lärt oss om biståndsförhållandet och om utvärderingar och utvärderingsprocessen?

Biståndsrelationen

Från stöd till utbildningsinnovation till biståndsberoende

I många år var det externa utbildningsbiståndet till låginkomstländer fokuserat på specifika projekt avsedda att utöka och förbättra utbildningen. I den bemärkelsen var utlandsbiståndet en mycket liten del av de sammanlagda utgifterna för utbildning, kanske bara 1–3 %. Men trots den begränsade volymen hade stödet enorma hävstångseffekter. På senare tid har situationen förändrats, särskilt i världens fattigaste länder. Biståndsmyndigheterna gör nu direkt och indirekt via nationella budgetsstöd det som de tidigare sagt att de inte skulle göra – ger stöd till den ordinarie budgeten. Eftersom lönekostnaderna står för den största delen av de sammanlagda utbildningskostnaderna är det i vissa länder biståndsgivarna som i praktiken betalar lärarnas löner. Ett sådant arrangemang verkar ohållbart, men hittills har det inte förekommit särskilt mycket

diskussion kring någon strategi för att växla över till självbärande av utbildningskostnaderna. Kampanjerna inom ramen för Education For All har i stället förutsatt betydande och ökande utbildningsbistånd.

Trots de regelbundet återkommande löftena om ökat stöd till utbildning har de senaste trenderna gått åt motsatt håll. Globalt har biståndet till grundläggande utbildning stagnerat eller minskat. Det har dock inte minskat biståndets inflytande.

Felmatchande tidshorisonter

Utvecklingsbistånd har en tydlig cykel och tidshorisont. Eftersom de flesta anslag är årliga så är det svårt för biståndsgivande regeringar, och i vissa fall rättsligt sett omöjligt, att garantera ett långsiktigt stöd. Utbildningsinitiativ har dock i allmänhet tidshorisonter som sträcker sig längre än ett, tre eller till och med fem års finansiering. Ett annat problem är att tjänstemännen vid biståndsmyndigheter har en relativt kort anställningstid. Dessutom har trenden mot outsourcing och privatisering lett till nya roller för de finansierande organens fältpersonal, som numer oftare ingår avtal med ”Contract Managers” snarare än med experter och rådgivare på utbildningsområdet. Biståndets och utbildningens tidshorisonter stämmer därför inte alls överens.

Felmatchningen får betydande konsekvenser för utvärdering. Biståndets korta cykel kräver snabba utvärderingar på kort sikt, ofta långt innan de avsedda resultaten ens börjar bli märkbara. Föga förvånande är utvärderingarna därmed ofta ytliga, och ägnas mer åt det som kan mätas snabbt (”Hur många lärare deltog i workshopen?” ”Levererades böckerna?”) snarare än åt huruvida undervisningen och lärandet förbättrades.

Attribueringsproblem

Det är sällan som initiativ och reformer på utbildningsområdet skapar omedelbar nytta. I de fall positiva resultat kan mätas i ett senare skede, är det svårt att avgöra vilka faktorer framgången beror på, något som ofta kallas attribueringsproblemet. Inte sällan vill de finansierande organen få bekräftat att deras stöd har åstadkommit resultat, även när de deltar i budgetstöd som samlar bistånd från flera organisationer eller givare.

Detta är en svår nöt att knäcka för utvärderarna. Att fastställa attribueringen är på en och samma gång nödvändigt, problematiskt och kanske omöjligt. Biståndssystemet skapar starka incitament för att

fortsätta arbeta som om det var möjligt att klart fastställa en tydlig attribuering och sedan rapportera att attribueringen bekräftats, på grundval av de bevis som ändå finns tillgängliga.

Utvärderingar: Varför? För vem?

Nu övergår vi till utvärderingarna och utvärderingsprocessen. Utvärderingarna själva är sällan självreflekterande eller självkritiska.

Historien upprepar sig

För att utvärderingar ska vara användbara måste de bli lästa, granskade och ”smälta”, och utvärderingsresultaten måste införlivas i politik och program. Ändå försvinner utvärderingarna ibland ner i ett bottenlöst hål. Våra detaljerade fallstudier ger relevanta exempel på detta.

Med jämna mellanrum används utbildningsbiståndet för att ge stöd till insatser där teknik ska användas för att ersätta lärare i områden där många lärare har en begränsad utbildning eller lite eller inga yrkeserfarenhet. Tidigare har det varit radio och tv, och nu är det datorer och telefoner. Utvärderarna har därefter rapporterat ett lyckat genomförande, men de noterar också samtidigt att det finns kvarstående problem. De finansierande organisationerna, som tycker sig ha fått bekräftat att stödet är effektivt och därför är beredda att tillmötesgå förfrågningar om ny teknik, påbörjar då en ny cykel. I praktiken upprepar finansiärerna då ett bristfälligt tillvägagångssätt, med liknande resultat – framgångar på kort sikt och frustration på lång sikt, samt få urskiljbara positiva effekter på själva lärandet. Man lär sig inte heller särskilt mycket av erfarenheter, framför allt när personalen byts ut och utvärderarna inte granskar historien bakåt.

Den kunskapskumulation och det institutionella lärande som man förväntat sig uppstår ofta inte. Utvärderingar och underbyggd kunskap visar sig ha mindre betydelse för att forma de finansierande organisationerna beteende än andra former av inflytande och som gynnar särskilda projekt och medelstilldelning, oavsett tidigare bevisade problem. Såväl personalen vid de finansierande organen som utvärderarna fäster regelmässigt för lite avseende vid relevant historia, bland annat systematiska, detaljerade och kritiska utvärderingar, och har uppenbarligen för lite incitament att göra detta.

Att bortse från kontext och komplexitet

Genom sin fokusering på kontext och komplexitet belyser fallstudierna riskerna med att inte ta hänsyn till kontext och

komplexitet. Även de mest kompetent genomförda och insiktsfulla utvärderingar kan förbli okända och oanvända. Varför?

För det första har inte utvärderingar någon framträdande plats i vardagen för utbildarna i de stödmottagande länderna, ens när de kan ha en direkt relevans för deras arbete. Utbildarna använder sig inte av utvärderingarna för att få information och vägledning när de utvecklar nya initiativ.

För det andra verkar det som om lärandet, trots att de biståndsfinansierade initiativen skapar betydande lärande, ändå förblir begränsat till de personer som är inblandade i det finansierade projektet och det är sällan som det lärandet stimuleras av eller fångas upp i utvärderingen. Det är förmodligen därför som stödmottagarna inte betraktar utvärderingen som sina verktyg, som kan uppfylla deras behov, och som lätt kan anpassas till och införlivas i deras tänkande och beslutsfattande. Genom vad som på ytan framstår som en deltagarstyrd process betraktar utbildningspersonal ofta utvärderingarna i stort som en extern process, ett krav från biståndsprocessen. Ägarskap är dock viktigt, inte bara för biståndsfinansierade utbildningsinsatser, utan även för utvärderingen av insatserna.

För det tredje, och kanske viktigast, kan inga utvärderingsresultat påverka framtida beteenden om utvärderingarna begränsas till input och output, eller om de bara dokumenterar processen mekaniskt utan att utforska samband och interaktion – och alltså bortser från komplexitet och kontext. Utvärderare rapporterar regelmässigt om vad som gjorts och inte gjorts, men inte för vem eller vilka detta var viktigt. Att bortse från komplexitet och kontext begränsar, och till och med undergräver, såväl utvärderingens substantiella kvalitet som dess användbarhet.

Utvärdering är till sin natur interaktiv. Det är nästan alltid så att förståelsen för hur ett resultat (outcome) uppnås är minst lika viktigt, kanske ännu viktigare, än själva resultatet. Att bortse från komplexitet och kontext undergräver vår förmåga att förstå och förklara just detta.

Formativa och deltagande utvärderingar

Deltagande utvärderingar är mycket vanliga i samband med internationellt utvecklingssamarbete, och har fått ökad uppmärksamhet som en reaktion på begränsningarna i 1970- och 80-talens toppstyrda angreppssätt, där de finansierande organisationernas

prioriteringar ibland verkade oförenliga med de tänkta stödmottagarnas behov. Ett centralt mål här är att ge lokalsamhället makt och förmåga att göra egna analyser av behoven och prioriteringarna och samla dessa samhällsdrivna element i en handlingsplan.

Det är inte förvånande att de många varianterna av deltagande utvärderingar, och de ibland utpräglade metodskillnaderna dem emellan, ger bränsle åt fortlöpande diskussioner om dessa utvärderingars styrkor och begränsningar. Forskare på utvärderingsområdet debatterar huruvida syftet med dessa utvärderingar är lika expansivt som skiften i maktdynamiken och främjandet av sociala förändringar. De som kritiserar ett deltagande tillvägagångssätt ifrågasätter värdet av deltagarnas inblandning i utvärderingen, och menar att det hotar objektiviteten.

Deltagande utvärderingsmetoder är varken oproblematiska eller lämpliga i alla situationer. De kan dock bidra till att minska tre risker som framgått mycket tydligt av vår utvärderingsgenomgång. För det första kräver deltagande utvärderingar att man tar hänsyn till kontext och komplexitet, vilket är centralt för att förstå vilken roll biståndet spelar och vad det kan få för konsekvenser. För det andra kan deltagande utvärderingar, när de utformas för att ha en både formativ och summativ roll, vara en generativ input för stödmottagarna snarare än en påtvingad börda utan omedelbar relevans. För det tredje breddar stödmottagarnas deltagande deras ägarskap av utvärderingprocessen, vilket i betydande grad ökar sannolikheten för att utvärderingsresultaten och rekommendationerna kommer att användas av både finansiärer och mottagare.

För många utvärderingar används för lite

Vår genomgång gav få bevis för att utvärderingar används till ett av de avsedda syftena, nämligen för att förbättra kvaliteten på biståndsfinansierade utbildningsprojekt. Med några undantag kunde vi konstatera att de utvärderingar vi gått igenom inte sammanfattade eller noterade resultat från tidigare utvärderingar. Analyser inom ramen för fallstudierna bekräftade att trots att respondenterna konsekvent underströk vikten av utvärderingar i allmänhet, var det få som kunde ge konkreta exempel på att utvärderingar lett till förändringar i policy eller praktik.

Vi har lyft fram flera olika orsaker till detta. Dekontextualiserade utvärderingsmetoder, ytliga eller svagt

underbyggda analyser och rekommendationer, felmatchade tidshorisonter samt attribueringsproblem som leder till att utvärderingarna sällan genererar åtgärdbara slutsatser som kan användas direkt i utformningen och genomförandet av projekt. Professionella prioriteringar, institutionella belöningssystem, ett mycket begränsat institutionellt lärande och utökande anspråk på personalens tid gör utvärderingarna till ett krav som ställs på de finansierande organens utbildningspersonal utan att vara till särskilt stor direkt nytta för dem. Ett svagt ägarskap av utvärderingsprocessen gör utvärderingarna till ett periodiskt intrång istället för ett konstruktivt tillskott för både mottagarländernas och de finansierande organens utbildare.

När de utvärderingar som processen krävs går långt utöver vad utbildarna bedömer som användbart, och regelmässigt överbelastar kapaciteten, är det sannolikt att de blir till byråkratiska formaliteter som genomförs när man måste och ignoreras så snart man kan. Det visar sig att det inte är ovanligt att utvärderingarna är tekniskt väl genomförda, omfattande, kanske kostsamma, och i stort ignorerade. Mer utvärderingar, mindre användning.

Sammantaget ger dessa slutsatser stöd för slutsatsen att olika syften kräver olika typer av utvärderingar. De finansierande organen har intresse av att se till att deras medel används som avsetts och av att kunna avgöra vem och vad som ska finansieras. Regeringar vill försäkra sig om att deras utbildningspolitik är i linje med de nationella prioriteringarna och de politiska målen. Genomförandeorganisationerna vill förbättra sina insatser för att kunna attrahera fortsatt stöd. Lärare, familjer och samhällen vill veta hur de bäst kan stötta barnens lärande. Alla dessa mål kan inte stödjas med en enda typ av utvärdering.

Biståndsmyndigheternas behov av data och statistik

Då och då framhålls att de finansierande organen skulle kunna dra nytta av de system som utbildningspersonalen använder för att förvalta sina utbildningssystem. De finansierande organisationerna behov av mätningar och datainsamling är dock för närvarande betydligt mer omfattande än vad som behövs för den dagliga förvaltningen. Man kan säga att det ständiga kravet på låginkomstländer att samla in, hantera och analysera mer och mer data avleder erfarenhet och expertis från de utbildningsinsatser som

biståndet ska stödja. I biståndsrelationen blir styrningen av biståndet ett hinder för biståndseffektiviteten.

Omprövning av utvärdering och utvärderingarnas roll

Vad kan vår genomgång av utvärderingar av biståndsfinansierade utbildningsinsatser lära oss om utvärderingar? Med några få undantag framstår det som osannolikt att de allt mer komplexa utvärderingarna kommer att kunna förbättra utbildningen eller öka biståndseffektiviteten. I de fall där det lokala generativa deltagandet i utvärderingsprocessen är begränsat är det lokala ägarskapet av utvärderingarna sannolikt mindre, liksom det lokala engagemanget i utformningen och genomförandet av utvärderingar och det lokala intresset för utvärderingsresultaten. Så länge det saknas en bredare fokusering på utvärderingarnas roll kan inte en bättre utformning av utvärderingarna och en ökad vetenskaplig disciplin lösa problemen.

För de finansierande organisationerna har detta flera implikationer.

I de fall utvärderingar behövs för att bekräfta att biståndsmedlen används på avsett sätt, bör utvärderingarna begränsas till den rollen. För det ändamålet kan utvärderingarna göras mycket enklare, billigare och mindre tidsödande för både biståndsgivare och mottagare.

Om utvärderingarna ska tillgodose andra syften, exempelvis ökad lokal öppenhet och insyn, eller redovisning av biståndsflöden, så kan de utformas och hanteras för dessa syften.

Komplexa och kostsamma utvärderingar som genomförs av utomstående kan svara mot vissa smalt definierade mål, men deras allmänna användbarhet är begränsad. Att försäkra sig om lokalt ägarskap av utvärderingar utesluter inte heller möjligheten att genomföra experimentella eller kvasiexperimentella effektbedömningar. När de används tillsammans med processutvärderingar och kvalitativa bedömningar kan den här typen av effektuppskattningar användas för att bevara utvärderingsfrågor som ”Varför?”, ”Hur?” och ”Under vilka omständigheter?”.

Betydligt mer kostnadseffektiva, och lättare att använda, är utvärderingar som åstadkommer tillförlitlighet, validitet och legitimitet genom att systematiskt inkludera stödmottagarna i såväl

utformning som genomförande och tolkning, och som innefattar både formativa och summativa mål.

Utvärderingarna i sig kan bli en del av utvecklingssamarbetet. När de innefattar ett betydande mottagardeltagande, och särskilt när de är väl integrerade i de biståndsstödda insatserna och ger formativa resultat, kan utvärderingar ge empowerment. De kan också bidra till att strukturera formerna för ansvarigheten gentemot stödmottagarna.

I stället för standardiserade utvärderingsmetoder som används brett, kan de finansierande organen och de utbildningssystem som får stöd utveckla en portfölj med olika slags utvärderingar som passar till sammanhang. Både biståndsgivarna och mottagarna kan ha nytta av att öka andelen formativa utvärderingar i förhållande till andelen summativa. Om man fokuserar på utbildarnas behov och användning av utvärderingar är det mer sannolikt att man förbättrar utbildningsresultatet, än om man som vanligt fokuserar på biståndsgivarnas krav på uppföljning.

De finansierande organisationerna tar regelmässigt risker då de ger stöd till innovationer på utbildningsområdet. Om man samtidigt är beredd att ta risker i samband med utvärderingar, kommer detta att främja utvecklingen av innovativa metoder för att försöka förstå konsekvenserna (avsiktliga och oavsiktliga) och effekterna (önskade och oönskade) av både utbildningsreformer och externt stöd.

I stället för att utforma utvärderingarna med det i stort sett ouppnåeliga målet att avgöra vad som fungerar eller vad som fungerar bäst, kan utvärderingarna utformas för att undersöka hur vissa saker fungerar under specifika omständigheter och sedan användas för att förbättra både utbildningen och biståndsprocessen.

Tillvägagångssättet att anlita utomstående – eller team som leds och styrs av utomstående ”objektiva” bedömare – för att genomföra utvärderingar kan i en del fall stärka utvärderingsarbetet, medan det i andra fall gör utvärderingen mindre användbar. Både utbildningen och biståndet kommer att gynnas av utvärderingar och utvärderare som har sina rötter i de insatser som ska bedömas, och av att administratörer, lärare och elever uppmuntras att införliva reflektion och utvärdering i det dagliga arbetet.

Summary

Aid providers have periodically reviewed their policies, priorities, and practices and sought to assess the roles and consequences of their support. Formal evaluations of aid to education have become more frequent, more systematic, and more important in subsequent policy and programmatic decisions. Indeed, those evaluations have become a new branch in the development literature.

What do we learn from that increasing volume of evaluations? In what ways have they facilitated evidence-based policy and programmatic decisions? How have aid recipients used those evaluations to improve their practice?

Both education in poor countries and external aid to support it have many purposes, many forms, and many contexts. Evaluations have differing objectives, approaches, and audiences. Based on a broad reading, an informed and informative synthesis must therefore both explore and highlight themes relevant to those audiences and at the same time address what is problematic. Broadly grounded insights are more useful to both practice and policy than an effort to construct an average across disparate and not readily comparable experiences, which risks blurring important distinctions, missing contextual complexity, and remaining little helpful to any of the intended audiences. Compounding the synthesis challenge is limited dissemination and discussion. Not infrequently, aid providers commission evaluations that remain little known and little useful to those whom the aid was intended to assist and that seem to have little influence on aid practices.

We are in a time of reappraisal. As the world re-thinks and resets education goals and indicators, aid providers reassess and revise their priorities and approaches. So, too, is it timely to re-think evaluations, from conception through method to use.

Review and Synthesis

Seeking to improve both education and foreign aid, the Swedish Expert Group for Aid Studies commissioned this synthesis of evaluations of aid-funded education activities. Framing our review is the recognition of the importance of complexity and context. Education, aid, and evaluation are multi-layered and therefore require attention that is multi-layered and multi-dimensional.

We begin by reviewing important evaluation issues, including expectations of the roles evaluations can play and the increasing preference for quasi-experimental and experimental approaches. We turn then to the major findings of our review, concerning the intersection of aid and education and the evaluation process. We explore the aid context and conclude with observations on the roles of evaluations of aid-funded education activities, noting the importance of a differentiated evaluation strategy that matches approaches to specific needs, purposes, and target constituencies.

Our review and synthesis are addressed to several overlapping but distinct audiences, each with its own experience and expertise. Some of the issues raised here will be new to some readers and thoroughly familiar to others. We have sought a reasonable balance, and we encourage readers to concentrate on the sections of this report they find most challenging and most useful.

What works?

Though everyone involved wants to know *what works*? that is not a fruitful organizing query for a review of evaluations of aid-funded education activities. Quite simply, a promising initiative may achieve intended objectives in one setting but not another and may have undesirable consequences in a third. Or it may seem effective to funders but not to practitioners, or to practitioners but not evaluators. A useful synthesis must incorporate attention to complexity and context.

Productive, therefore, is to ask what works for whom? in what circumstances? under what conditions? That, in turn, requires exploring situationally specific specifications of success. Not only may an education initiative improve results in one setting but not another, but that same initiative may be deemed successful from one perspective (exam results) and a failure from another (female attrition; cost).

Several other complexities confound efforts to synthesize evaluations and to develop clarity on education and aid effectiveness. Often, evaluators and researchers seek to avoid those complexities through simplifying assumptions—"other things being equal"—or by relegating them to the margin of the assessment and then directly or indirectly holding them constant. Those approaches seek a clearer view by dissecting the behaviour or relationship of interest out of its setting. The risk in those approaches is that the view will be clearer but

more limited, often so limited that it precludes drawing reasonable inferences useful to aid providers and recipients.

Our approach is just the opposite, insisting that phenomena must be understood in their context.

Flawed Premises

With rare exceptions, aid programs require evaluations. Beyond confirming that support was linked to stated objectives and that funds were used appropriately, evaluations are expected to improve the aid process. That rationale rests, it seems clear, on three premises that are engaging and initially persuasive but that have little research support. First, notwithstanding expansive claims about the importance and value of learning from experience, quite simply, there is little evidence of direct learning from experiences reported in evaluations and rarely a trace of cumulation of learning from the succession of evaluations over many years. Second, evaluations are regularly understood as applied research that generates relevant knowledge for evidence-based policy. Evidence-based policy is an appealing notion. But the premise that evaluations play an important role in generating knowledge that directly shapes policy is not supported by available evidence. Third, developing public policy is often understood as a largely rational and linear process. If evaluations contribute to policy formulation, it is through chaotic, discordant, and often poorly linked pathways.

Even brief attention to three flawed premises demonstrates clearly the gap between the claimed role of evaluations—to generate knowledge that permits learning from experience, which in turn improves aid and education policy and practice—and the roles evaluations can play. We note this gap not to decry the constraints on rational policy making and optimization but rather to encourage recognition of those limits and humility in claims about what is knowable, how knowledge is generated, and how knowledge is applied.

The Emerging Standard

There has been a convergence, though not unanimity, on a particular approach: impact evaluations, where possible with randomized controlled trials. An estimated USD 150 million was spent on RCT evaluations of education programs in 2013.

The push for RCTs has by no means gone uncontested. First, RCTs are expensive. Second, randomization is unfeasible or extraordinarily difficult in many, if not most, poor country education contexts—for practical, political, and ethical reasons. The practical challenge is that most often education initiatives and reforms are implemented in ways that are not readily amenable to the requirements of experiment-like impact assessments. The political challenge is that uneven distribution of resources, in this case improved education opportunities, requires a political rationale and political legitimacy, not simply an experimentalist's specification. The ethical challenge has three components. Random assignment is incompatible with notions of preference and choice that students, parents, and communities value. In contexts where there is reason to believe that certain schools or students may benefit more than others from a particular program, the ethics of random assignment are problematic. RCTs often compare an innovation or reform with no change, an approach that does not meet the ethical standards for comparing alternative experiences.

Third, applying an approach used in the health sector to protect recipients of experimental treatments does not work seamlessly in education, where differences in institutional capacity and resources between schools and communities, along with socio-political and cultural differences, mean that program implementation (the treatment) is rarely stable or common across settings, even when they are randomly selected. Fourth, education development arguably *should not* meet the requirements of RCTs. Differences in program implementation are important, and should even be encouraged, rather than stifled in the push for a stable treatment. Fifth, like all types of evaluations, the findings of an RCT are specific to the context and to the conditions under which the evaluated program operates.

Sixth, a recent review of six systematic reviews of evaluations of education programs in low-income countries calls into question the presumption that a large volume of impact assessments with RCTs will identify preferred and widely appropriate education content and teaching strategies. While all six systematic reviews used a similar approach, the authors found almost no overlap in the conclusions drawn from these evaluations—dramatic discord where we expect consensus. That finding provides further evidence that no volume of evaluations and data collection can uncover a blueprint of what works for education. Indeed, that search for a blueprint, or set of standard approaches or practices, is not productive. Learning is a participatory,

interactive, and dynamic process, deeply intertwined with the political, economic and historical contexts within which formal and non-formal education take place.

Both the limitations of RCTs and the practical, financial, and ethical challenges in their implementation lead to the conclusions that while impact assessments and RCTs can be useful in evaluating aid-funded education activities, their effective domain is constrained, and that certainly neither RCTs, nor impact assessments more generally, are the standard against which other approaches to evaluation must be assessed.

When Method Determines Outcomes

Recent research on poverty and growth in Africa shows clearly the risks of relying on a single research approach or method and of assuming that if the method is correct, its results and recommendations must also be correct. To reduce those risks we have employed multiple methods and approaches rather than privileging a single method, however scientific its aura. Minimizing the risk of evaluator bias requires engagement with educators, decision makers, and communities, not distance from them. Systematic and critical attention to complexity and context are essential for assessing the utility of a proposed approach or method and its findings. It is that attention to history along with quantitative data, to educators and learners and their voices along with detached observers, and to experience along with statistical analysis that make an approach scientific.

An Integrated Approach

What can we learn from evaluations? Our focus is evaluations as a set, not individual evaluations. We are not asking whether or not a particular evaluation provides clear findings that might guide action. Rather, we are exploring what can be learned from the broad range of evaluations undertaken within the aid relationship. Recognizing that a well supported finding cannot improve education if it is not applied, we explore as well the uses of evaluations.

To begin, we conducted a comprehensive search of evaluations of education activities commissioned by international and national funding and technical assistance agencies, the OECD Development Cooperation Directorate, UNICEF, education-focused NGOs, as well as prominent education-focused research institutes and consulting firms. A guiding concern was to develop a set of

evaluations diverse in approach, commissioning agency, specific focus, and involvement of aid recipients. That is, we sought to maximize diversity, not quantity. This search resulted in an initial list of 80 evaluations. From this set we selected 40 evaluations for more detailed review. Through our subsequent examination of those and other lists we are confident that the selected set reasonably reflects the broader universe of evaluations of aid-funded education activities. In addition, from the larger set of evaluations we selected three for detailed assessment across multiple layers: aid-funded activities in Tanzania, Nepal, and Bénin.

Ours is a modest synthesis, aimed at in depth and detailed analysis, rather than identifying and classifying every evaluation that has been undertaken. To the best of our knowledge, our synthesis is the first to include such a diverse sample of evaluations (in terms of methods used, types of policies and programs evaluated, funding agencies, countries and contexts) and to address evaluations as a set rather than focusing on a few well-grounded evaluations of particular activities.

Evaluations of aid to education in poor countries

Several observations stand out from the evaluations reviewed.

Effective education efforts reach beyond schools. Evaluations of aid-funded education activities provide confirmation and rich evidence: effective education efforts reach beyond inputs and beyond schools. A clear example is efforts to achieve education for all. The most effective strategies for increasing enrolment appear to be the reduction in the costs for families combined with sustained advocacy and awareness activities.

Inputs are not enough. Most aid programs focus on inputs of some support. Only rarely do aid programs embed the provision of inputs in a larger frame that is attentive to the supports needed for the inputs to be used well, to who is responsible for receiving and managing the inputs, to needed on-going support (including technical assistance and maintenance), to integration into the national and local education system, to responses by teachers, learners, and communities. Aid programs that focus entirely or primarily on inputs are less effective than those that start with a holistic notion of education as a process and education as a system and that embed that understanding in the aid program. Although they note this problem, evaluators may contribute to it. Only rarely do they seek to close or even address the

gap between broader development goals (poverty, social inclusion, human rights, democracy, sustainable development) and supported education activities.

Effective external support reaches beyond the education ministry. Just as the focus on inputs is limiting, so too can be concentrating attention on the education ministry. Foreign aid funds that are most effective in improving education reach beyond the centralized authority of the education ministry or department.

Local ownership of education innovation: essential but rarely evaluated. The importance of local ownership has long been clear and is often highlighted in the aid literature. Evaluations have regularly noted that activities for which there is a strong sense of local ownership are much more likely to be effective, or more effective, or more inclusive, or better sustained than activities which those involved regard with some distance and perhaps with a sense that they have been delivered or imposed by outsiders. Yet, only rarely does aid funding focus explicit attention on developing, nurturing, and funding a strong sense of local ownership of the education activities that are supported. Similarly, few evaluations study or assess local ownership systematically and thoroughly.

It is essential to recognize the inherent and powerful tension between local ownership and funding agency interests and objectives. The issue is locus of authority. Achieving strong local engagement in and responsibility for aid-funded education activities requires that recipients have significant control over the activities and the funding. Funding agencies, however, have their own objectives and lines of responsibility and accountability and may be unwilling or unable to cede authority to the aid recipients.

Reaching the difficult to reach remains beyond reach. The evaluations we have reviewed confirm the challenges of extending education opportunities to the most difficult to reach populations, which remain largely excluded from aid-funded education projects. Aid funding intended to reduce inequality may in practice relocate it.

Centralization despite decentralization. Earlier, the World Bank and other funding agencies regarded decentralization—transfer of authority and responsibility from central to local levels—as an essential component of education reform. In many countries, however, the most common practice in the education sector has been deconcentration—relocation of some officials and roles from central

to provincial or local education ministry offices, without a significant transfer of power and authority to local communities.

Beyond confirming that in aid-receiving countries (and in most of the world) most people think education requires a strong central authority and there has not been much decentralization, what else do the evaluations tell us about decentralization? First, decentralization is an important component of official education development strategy. Second, the evaluations confirm that decentralization comes in many shapes and sizes. Third, there is also substantial evidence that notwithstanding its expected benefits, decentralization can exacerbate existing inequalities between schools and communities. Fourth, while the rhetoric of decentralization highlights community empowerment and local accountability, in practice, meaningful participation at the community level may be difficult to achieve and is often limited to financial contributions or school maintenance activities. Fifth, decentralization strategies sometimes encounter local resistance. Sixth, even as many evaluations stress the importance of decentralization, few address it explicitly as part of the evaluation or explore how aid agencies might facilitate the decentralisation process.

Sustainability: important but not systematically evaluated. In September 2015 the United Nations formally adopted Sustainable Development Goals. Yet, while funding agencies regularly reiterate their expectation that aid-funded education activities be sustainable, in practice aid programs generally do not include either explicit attention to what is required for that sustainability or funding specifically dedicated to achieving sustainability. Not surprisingly, many evaluations do not address sustainability systematically.

Information, evidence, data, and indicators. The need for better information management, data and indicators is a pervasive finding across the evaluations we have reviewed. With important exceptions, most of those evaluations point to gaps and other problems in the available education data. Yet surprisingly few of these evaluations address data problems directly, either by collecting their own general education data or by developing strategies for working with seriously flawed data. Nor do most evaluations integrate into their findings the very large probable margins of error in most of the available education data.

Also generally unaddressed in the evaluations we reviewed are the trade-offs between increased efforts to collect more and more

reliable education data on the one hand and on the other, efforts focused on making better use of a much smaller number of indicators. Nor do the evaluations explore how the funding agencies might proceed if they based both their support programs and their evaluations on the limited, and not infrequently partial and inconsistent, data that aid-receiving education ministries use regularly to manage education systems.

The importance of institutional knowledge and learning among funding agencies. The evaluations we have reviewed provide strong support for a familiar recommendation: the need for substantial institutional knowledge and learning among funding and technical assistance agencies. The major challenge in improving aid effectiveness is not in acquiring or documenting knowledge, but in enabling and encouraging organizations to act on existing knowledge. While they addressed data needs and data collection, the evaluations reviewed did not analyse knowledge sharing among networks or inter-organizational partnerships.

Education, aid, and evaluations

What, then, do we learn from the set of evaluations about the aid relationship and about evaluations and the evaluation process?

The Aid Relationship

From Support for Education Innovation to Aid Dependence.

For many years, external support to education in low income countries was focused on specific projects intended to expand and improve education. In that role, foreign aid was a very small part of total spending on education, perhaps 1-3%. Though its volume was limited, that aid had tremendous leverage. Most recently, especially in the world's poorest countries, that situation has changed. Both directly and indirectly through national budget support, foreign aid agencies are doing what previously they said they would not do: supporting the recurrent budget. Since the wage bill is the major portion of total education spending, in some countries, effectively the aid providers are paying the teachers. While that arrangement seems unsustainable, to date there has been little discussion of a strategy for shifting to self-reliant education spending. Indeed, the education for all campaign has presumed substantial and increased provision of education aid.

Notwithstanding periodic promises of increased education assistance, the most recent trend has been in the opposite direction.

Globally, aid to basic education has stagnated or declined. That has not, however, reduced its influence.

Mismatched Time Horizons

Foreign aid has a clear cycle and time horizon. Since most appropriations are annual, aid-providing governments find it difficult, or are legally unable, to assure long-term support. Education initiatives, however, generally have time horizons that extend beyond one year, or even three-to-five year funding. Also problematic is the relatively short job cycle of funding agency officials. As well, a major consequence of the push toward out-sourcing and privatization is the transformation of the role of the funding agency's field staff, who are more likely to be contract managers than education experts and advisers. The aid and education horizons are thus sharply mismatched.

That mismatch has powerful consequences for evaluation. The short aid cycle requires near-term evaluations, often well before the intended outcomes can become clearly visible. Not surprisingly, evaluations are often correspondingly superficial, attentive to what can be measured quickly (how many teachers participated in the workshop? were the books delivered?) rather than whether or not teaching and learning improved.

Attribution Challenges

Only rarely do education initiatives and reforms yield instant benefits. When positive outcomes can later be measured, it is difficult to determine what were the major causes, commonly termed the attribution problem. Often the funding agencies seek confirmation of the benefits of their assistance, even when they participate in budget support that combines the aid of several agencies.

Thus a conundrum for evaluators. Establishing attribution is simultaneously necessary, problematic, and perhaps impossible. The aid system creates strong incentives for proceeding as if it were possible to establish clear attribution and then to report that on the basis of available evidence, attribution has been confirmed.

Evaluations: For What? For Whom?

We turn now to the evaluations and the evaluation process. Evaluations themselves are rarely self-reflective or self-critical.

Déjà Vu All Over Again

For evaluations to be useful, they must be read, reviewed, digested, and their findings incorporated in policy and programs. Yet, sometimes evaluations disappear into a bottomless pit. Detailed case studies provide relevant examples.

Periodically, aid supports efforts to use technology, earlier radio and television, currently computers and telephones, to substitute for teachers where many teachers have limited education and little or no professional preparation. Having verified successful implementation, evaluators also note persisting problems. Assured of the effectiveness of their support and responsive to requests for new technology, funding agencies subsequently start a new cycle. In practice, the funders repeat a flawed approach, with similar results: short term success and longer term frustration, with little discernible positive effect on learning. Especially as professional staff changes and evaluators do not review the earlier history, there is little learning from experience.

The expected cumulation of knowledge and institutional learning often do not occur. Evaluations and well grounded knowledge prove less important in shaping funding agency behaviour than other influences that favour particular projects and allocations, notwithstanding the evidence of problems. Regularly both funding agency staff and evaluators pay little attention to relevant history, including systematic, detailed, and critical evaluations, and apparently have little incentive to do so.

Ignoring Context and Complexity

Through their attention to context and complexity, case studies highlight the perils of ignoring context and complexity. Even the most competent and insightful evaluations may be little known and little used. Why?

First, evaluations, even where they are directly relevant to their work, do not feature prominently in the daily lives of educators in aid-receiving countries. When they develop new initiatives, educators do not turn to evaluations for information and guidance.

Second, while there is important learning in aid-funded initiatives, that learning may remain limited to those involved in the funded project and only rarely stimulated by or captured in the evaluation. That is most likely where the aid recipients do not regard

that evaluation as their tool, responsive to their needs, readily appropriate and incorporated into their thinking and decisions. Often, throughout what appears to be a participatory process, education officials regard evaluations largely as an external event, a requirement of the aid process. Ownership matters, not only for aid-funded education activities, but also for their evaluations.

Third, perhaps most important, evaluations that limit their view to inputs and outputs, or that document process mechanically without exploring interconnections and interactions—that ignore complexity and context—are unable to produce findings that influence subsequent behaviour. Regularly, evaluators report on what was and was not done, but not to whom that mattered. Inattention to complexity and context sorely limits, indeed undermines, both the substantive quality of the evaluation and its utility.

Education is by design interactive. Nearly always, how an outcome is achieved is at least as important and perhaps more important than the outcome itself. Inattention to complexity and context undermines our ability to understand and explain that.

Formative and participatory evaluations

Participatory approaches are widespread in international development, attracting increased interest as a response to the limits of top-down approaches in the 1970s and 1980s, especially where funding agency priorities sometimes seemed incompatible with the needs of intended beneficiaries. A key objective is to empower the community to conduct its own analysis of its needs and priorities, and organize these community-driven elements into a plan of action.

Not surprisingly, the many variations of participatory evaluation and their sometimes sharp methodological differences fuel continuing contention about its strengths and limitations. Scholars of evaluation debate whether or not the purpose of evaluation is as expansive as shifting power dynamics and promoting social change. Critics of participatory approaches contest the inclusion of participants in evaluation, citing a threat to objectivity.

Participatory evaluation approaches are neither unproblematic nor universally appropriate. They can, however, reduce three risks that have emerged sharply in our review of evaluations. First, participatory evaluation approaches require the attention to context and complexity that is essential for understanding the roles and consequences of development assistance. Second, where they are designed to play a

formative as well as summative role, participatory evaluations can be a generative input for aid recipients rather than an imposed burden that has no immediate relevance. Third, by broadening the ownership of the evaluation process, recipient participation substantially increases the likelihood that evaluation findings and recommendations will be used, by funders as well as recipients.

Too many evaluations have too little use

Our review found limited evidence that evaluations are used for one of their intended purposes: to improve the quality of aid-funded education projects. With some exceptions, the majority of the evaluations we reviewed did not summarize or note findings from previous evaluations. Case study analyses confirmed that while our respondents consistently emphasized the importance of evaluations in general, few could provide concrete examples of evaluation-induced changes in policies or practices.

We have highlighted multiple reasons for this. De-contextualized evaluation approaches, superficial or weakly supported analyses and recommendations, mismatched time-horizons, and attribution challenges mean that evaluations rarely provide actionable results that feed directly into project design and implementation. Professional priorities, institutional reward systems, sharply constrained institutional learning, and over-stretched demands on their time make evaluations both required and at the same time of limited direct utility to funding agency education staff. Narrow ownership of the evaluation process makes evaluations a periodic intrusion rather than a constructive contribution for funding agency and recipient country educators.

Where required evaluations go far beyond what educators deem useful and regularly overwhelm capacity, they are likely to become formalistic exercises, completed when necessary and ignored as soon as possible. Not infrequently, it turns out, evaluations are technically sound, extensive, perhaps expensive, and largely ignored. More evaluations, less use.

Together, these findings support the conclusion that different purposes require different types of evaluations. Funding agencies are interested in ensuring that their funds are used as intended, and in determining who and what to fund. Governments want to ensure their education policies align with national priorities and political objectives. Implementing organizations want to improve their

operations in order to attract continued support. Teachers, families, and communities want to know how to support children's learning. No single type of evaluation will meet all of these objectives.

Aid Agencies' Data Demands

Periodic voices note that funding and technical assistance agencies could draw on the measures that education officials use to manage their education system. Currently, however, funding agencies require measurement and data collection that far exceed the needs of day-to-day education management. Put sharply, the incessant demand that low income countries collect, manage, and analyse ever more data diverts experience and expertise from the education activities that the aid is intended to support. In the aid relationship, *aid management becomes an obstacle to aid effectiveness.*

Re-thinking evaluations and their role

What do we learn about evaluations from our review of evaluations of aid-funded education activities? With occasional exceptions, more and more complex evaluations are unlikely to improve education or increase aid effectiveness. Especially where there is little local generative participation in the evaluation process, there is likely to be little local ownership of evaluations, little local engagement in their elaboration and implementation, and little local attention to their findings. In the absence of broader attention to their roles, better evaluation design and increased scientific rigor cannot solve these problems.

For funding agencies, the implications are several.

Where evaluations are needed to confirm that aid funds were used as intended, limit the evaluations to that role. For that purpose, evaluations can be much simpler, less costly, and less time consuming for both providers and recipients.

Where evaluations are intended to serve other purposes, say increasing local transparency and accountability for aid flows, they can be designed and managed for those purposes.

Complex and expensive evaluations by detached outsiders can serve occasional narrowly defined objectives but have limited general utility. Ensuring local ownership of evaluations does not exclude the possibility of conducting experimental or quasi-experimental impact evaluations. When accompanied with process evaluations and

qualitative assessments, these types of impact estimates can be used to answer why, how, and in what circumstances evaluation questions.

Far more cost-effective and more likely to be used are evaluations that achieve reliability, validity, and legitimacy through the systematic inclusion of aid recipients from conception through implementation to interpretation and that incorporate both formative and summative objectives.

Evaluations can themselves become part of development assistance. Where they incorporate significant recipient participation, and especially where they are well integrated into aid-supported activities and provide formative results, evaluations can be empowering. They can as well structure accountability to aid recipients.

Rather than a standard evaluation approach to be used broadly, funding agencies and supported education systems can develop a portfolio of evaluation sorts and types, appropriate to different circumstances. Both aid providers and aid recipients will find it useful to increase the proportion of evaluations that are formative, rather than summative. Focusing on educators' evaluation needs and uses is more likely to improve education outcomes than the common focus on aid providers' monitoring requirements.

Regularly, funding agencies take risks in supporting innovation in education. A parallel willingness to take risks in evaluation will encourage the development of innovative approaches to understanding the consequences (intended and unintended) and impacts (desired and problematic) of both education reform and external support.

Rather than the generally unachievable objective of determining what works or what works best, evaluations can be designed to examine how things work in specified circumstances and then used to improve both the education and the aid process.

While evaluation by detached outsiders, or teams led and managed by detached outsiders, will strengthen some evaluations, that approach renders other evaluations less useful. Both education and aid will benefit from evaluations and evaluators rooted within the activities to be assessed and from encouraging administrators, teachers, and learners to incorporate reflection and evaluation in their daily work.

1. Capturing complexity and context: evaluating aid to education

Education for all

For analyses of development, whether excitedly optimistic or persistently pessimistic, 2015 was a drama year of global targets, global assessments, global reappraisals, and global recommitment. The international flow of documents that reported on what has happened and what is to be done was dizzying. Equally energetic were the major international events that specified revised and new development and education objectives. Noting both progress and unachieved objectives, the world promised to do more. The 2015 World Education Summit and then the United Nations Sustainable Development Summit took stock, adopted goals, objectives, and indicators, and reset the targets to 2030.

For education in the world's poorest countries, the moment is sobering. Education has been an explicitly affirmed and reaffirmed high priority development domain for more than a half century. Meeting in Thailand in 1990, the world—governments, the United Nations system, other international organizations—formally adopted its commitment to education for all. The world convened in Senegal a decade later to assess progress toward that commitment. Frustrated that the initial objectives had not been met, the world reaffirmed its commitment to education for all, resetting most of the target to 2015.

Yet, the current global picture is troubling. In the world's poorest countries, far too many children remain out of school. Younger and older adults who have missed their schooling moment, especially women, have few opportunities to develop proficiency in reading and writing and to use that learning to transform their own and their societies' future. For many of those in school in those countries, there are very large classes led by teachers with very limited professional education, there are too few books, even pencils and chairs, and schooling functions as an inverted funnel, with few reaching the top and nearly all pushed aside. Millions of the world's citizens do not have access to the learning opportunities that their and their societies' development require.

Well before the formal declarations of global responsibility for achieving education for all, foreign aid, increasingly formalized in the

international system in mid-20th Century, regularly assigned high priority to education. International and national aid policies and strategies continue to do so.

Although the initial education for all declarations did not address directly how achieving the EFA objectives was to be financed, there was a clear global understanding that external support must play a significant role in expanding education opportunities. Historically that role had been limited to development, but not recurrent, expenditures. Providing and managing education remain a national responsibility. Over several decades, however, analyses indicated that national resources are insufficient to meet the projected costs of achieving education for all. Foreign aid was to close the gap. The 2000 Dakar Framework for Action made that commitment explicit, "No countries seriously committed to Education for All will be thwarted in their achievement of this goal by lack of resources." Even though in recent years foreign aid has stagnated and aid to basic education has declined, the new agenda also presumes that continued, indeed increased, foreign aid will be essential.

For education in the poorest countries, foreign aid has come to play a prominent role. In some, both the development and the recurrent budgets are heavily dependent on foreign assistance. What has that aid accomplished? Regularly its critics have responded: not enough, or even, not much. In part due to that frustration, funding and technical assistance agencies have insisted on the importance of explicit national policies, clearly stated and staged objectives, and improved monitoring and assessment. For many funding agencies that frustration has also fuelled attention to policies and allocations shaped by results.

Though aid to education has fallen short of projections, it has been substantial. Especially those countries that have regularly met the international aid targets have been frustrated that major objectives of that support remain unachieved.

2. Reviewing and synthesizing evaluations of aid-supported education activities

Aid providers have periodically reviewed their policies, priorities, and practices and sought to assess the roles and consequences of their support. Formal evaluations of aid to education have become more frequent, more systematic, and more important in subsequent policy and programmatic decisions. Indeed, those evaluations have become a new branch in the development literature.

What do we learn from that increasing volume of evaluations? In what ways have they facilitated evidence-based policy and programmatic decisions? How have aid recipients used those evaluations to improve their practice? Focused reviews using a narrowly defined subset of those evaluations have sought to assess the effectiveness of particular education initiatives. But what of the evaluations more generally, our focus here?

Both education in poor countries and external aid to support it have many purposes, many forms, and many contexts. Evaluations have differing objectives, approaches, and audiences. Based on a broad reading, an informed and informative synthesis must therefore both explore and highlight themes relevant to those audiences and at the same time address what is problematic in the evaluation process and thereby in the aid relationship. Broadly grounded insights are more useful to both practice and policy than an effort to construct an average across disparate and not readily comparable experiences, which risks blurring important distinctions, missing contextual complexity, and remaining little helpful to any of the intended audiences. Compounding the synthesis challenge is limited dissemination and discussion. Not infrequently, aid providers commission evaluations that remain little known and little useful to those whom the aid was intended to assist and that seem to have little influence on aid practices.

Our core concern, therefore, is to step back from the common query—do evaluations confirm the effectiveness of a particular education initiative?—in order to explore the large volume of evaluations as a set. What can we learn, especially about the

intersection of aid and education and about the evaluation process, from evaluations of aid-funded education activities?

Seeking to improve both education and foreign aid, the Swedish Expert Group for Aid Studies commissioned this synthesis of evaluations of aid-funded education activities. The Terms of Reference are in Annex G. This final report has been revised following review by the project reference group.

The major challenges of our work are to understand better, and through that understanding, to develop strategies for making more effective use of evaluations. For that, we must review evaluations of many different sorts. And for that we must address the needs and expectations of several different constituencies, from the creators and managers of foreign aid to those who commission evaluations to those who are expected to benefit from the external support.

We are in a time of reappraisal. As the world re-thinks and resets education goals and indicators, aid providers reassess and revise their priorities and approaches. So, too, is it timely to re-think evaluations, from conception through method to use. The ultimate goal, of course - important to keep in focus though beyond the reach of this limited project—is to improve education access and quality, to make the right to education the practice of education.

Review and Synthesis—the roadmap

What can we learn from evaluations of aid-supported education activities?

Framing our review is the recognition of the importance of complexity and context. Education, aid, and evaluation are multi-layered and therefore require attention that is multi-layered and multi-dimensional.

We begin by reviewing important evaluation issues, including expectations of the roles evaluations can play and the increasing preference for quasi-experimental and experimental approaches and randomized control trials. We turn then to the major findings of our review, concerning both education and evaluations. Next we consider the aid context, that is, evaluations initiated largely by and for funding and technical assistance agencies. We conclude with observations on the roles of evaluations of aid-funded education activities, noting the

importance of a differentiated evaluation strategy that matches approaches to specific needs, purposes, and target constituencies.

References to academic literature follow in the main report. Annexes include the list of evaluations reviewed, a discussion of our approach, our selection strategy, summary reviews of the larger set of evaluations considered, more detailed attention to a selected subset of those evaluations, reports of our case studies, and the terms of reference for our work.

Our review and synthesis are addressed to several overlapping but distinct audiences, each with its own experience and expertise. Some of the issues raised here will be new to some readers and thoroughly familiar to others. We have sought a reasonable balance, and we encourage readers to concentrate on the sections of this report they find most challenging and most useful.

What works?

Though everyone involved wants to know *what works?* that is not a fruitful organizing query for a review of evaluations of aid-funded education activities. Quite simply, a promising initiative may achieve intended objectives in one setting but not another and may have undesirable consequences in a third. Or it may seem effective to funders but not to practitioners, or to practitioners but not evaluators.

Productive, therefore, is to ask what works for whom? in what circumstances? under what conditions? That, in turn, requires exploring situationally specific specifications of success. Not only may an education initiative improve results in one setting but not another, but that same initiative may be deemed successful from one perspective (exam results; simplified implementation) and a failure from another (female attrition; cost).

A useful synthesis must incorporate attention to complexity and context. Our task is broadly analytic and synthetic, not more narrowly advisory on the problems and prospects of particular education reforms.

Education is Multi-Layered

Education is multi-layered. Addressing *what works* requires unpacking both *what* and *works*.

After many years of focus on education inputs, major attention has shifted to outcomes and results. Does a particular approach to teaching reading, for example, lead to improved reading abilities? Commonly, that is measured by scores on a national examination or perhaps an international test. Then, for aid-supported activities, both the providers and the recipients have empirical grounding for selecting a better approach. So far, so good, but that is not sufficient.

Education always has multiple objectives. A strategy for teaching mathematics that emphasizes rote learning may be associated with improved examination scores, at least in the short run, but may also undermine pedagogies focused on encouraging curiosity, promoting concept formation and problem solving, and developing self-confidence and self-reliance. From that perspective, improved mathematics scores may be a very poor measure of achieving desired outcomes. A narrow specification of objectives to facilitate assessment risks devaluing other objectives to which educators, learners, or parents may assign higher priority.

Education is context specific. In practice, learning objectives vary widely, are regularly revised and re-specified, and are generally negotiated. The notion of a global standard for, say, mathematics or reading, may be more obscuring than clarifying.

Perhaps most important, generally what matters most in education is process rather than outcomes. In this respect, education differs from many other activities for which evaluations that are indifferent to process in their focus on outcomes are appropriate. If learning, rather than examination scores, is the critical concern, and if the ways in which learning occurs are at least as important to communities as what is learned, process must be the central focus of education evaluations. The common education black box approach (focus on inputs and outputs, with little or no attention to what happens in between) ignores the core of education. Evaluations that are inattentive to the learning process cannot generate useful findings on *what works*, either in education or in foreign aid.

Technical issues in specifying education and effective education are often more important that is commonly assumed. There

is significant evidence, for example, that national examinations measure language competence, and perhaps test ability, much more than subject competence. Though they are commonly used proxies for achieved learning objectives, with very rare exceptions they are flawed and partial measures whose inherent biases are generally not noted or examined critically.

Aid is Multi-Layered

Foreign aid is multi-layered. Indeed, there are least three issue clusters. Here, we focus on development aid and do not address short-term humanitarian aid (that is, emergency assistance provided after a flood or drought or tsunami).

For the purposes of this synthesis, what, exactly, is aid? The OECD Development Assistance Committee's specification is a reasonable starting point. Overseas development assistance is a concessional transfer of resources provided by official agencies intended to promote the economic development and welfare of recipients (most often countries whose national income is below a specified threshold), with or without conditions. Global discussions of aid, however, regularly reach more widely. Some transfers are loans, with limited or no concessional features. Assistance may take the form of seconded personnel, or products (books; computers), or services (quality assurance for purchases), commonly on terms specified by the provider. Investment, especially by parastatals, may be categorized as aid. Support that is by design not overseas development assistance (military; humanitarian) may have significant education components. Overseas education for educators and students from poor countries is sometimes included in aid allocations and sometimes funded separately. An effective synthesis must recognize that what is aid is regularly negotiated, that the major sources of aid data may not capture all of the transfers that those involved consider to be aid, and that, more generally, reported aid flows over time may have a significant margin of error.

Aid has multiple providers: countries, multi-country groups (European Union), international agencies (UNDP; UNICEF), foundations (Ford; Gates), development banks (African Development Bank; World Bank), non-governmental organizations of several sorts (churches; unions), development or technical assistance funds (GPE; GFATM), companies (through their attached social responsibility

units or foundations), special purpose events (Live Aid), and more. While there are many reports on and studies of foreign aid, few have sought to address comprehensively the range of providers. Why is that important? Just as aid providers and recipients may differ on objectives and appropriate assessment strategies, so there may be sharp differences among the providers.

Aid has multiple pathways: direct government transfers to ultimate recipients; transfers via national (Church of Sweden), international (World Bank; World Vision), and local intermediaries (education ministry; early childhood association); transfers via contracted implementing agencies; project, program, sector, and budget support.

Compounding the challenge of evaluating aid and its roles is that all of these clusters intersect and interact, making a large number of combinations of objectives, forms, pathways, and modalities, sometimes within a single aid program. Where there are so many differentiating factors and so many possible mixes, it is difficult to generalize assigning weight or significance to any of them. The analysis must therefore be constantly attentive to context.

The technical problems in studying foreign aid are well documented. Available data sources are often widely discrepant (for example, allocations and timing reported by the funding agency differ sharply from receipts and timing reported by aid recipients) and have large error margins. While precision on those volumes and flows is not essential to this synthesis, we need to be attentive to the data issues.

Evaluation is Multi-Layered

Evaluations have multiple purposes and constituencies.

Some evaluations are used to justify policies and allocations decided by parliaments, government ministries, or governing bodies of international organizations. Some evaluations are intended primarily to assist decision makers (in funding agencies; in governments) by assessing alternative approaches or tools or funding. Some evaluations, both formative and summative, are designed primarily to assist educators and learners on the ground (teachers; students; local education offices).

Given that diversity of purpose, evaluations that serve one constituency well (for example, Members of Parliament reviewing the foreign aid allocation) may not be very useful for another constituency (teachers). At a minimum, it is reasonable to expect that

funding agencies, technical assistance institutions, implementing organizations, and aid recipients will have very different evaluation needs and therefore very different assessments of the utility (and quality) of evaluations of education and of aid to education. If the aid net is cast broadly, then banks, churches, and unions are also likely to vary in their evaluation needs and assessments.

An effective synthesis of evaluations of aid-supported education activities must recognize that objectives, expectations, and assessment needs vary across the aid relationship.

Confounding Complexities

Several other complexities confound efforts to synthesize evaluations and to develop clarity on education and aid effectiveness. Often, evaluators and researchers seek to avoid those complexities through simplifying assumptions—"other things being equal"—or by relegating them to the margin of the assessment and then directly or indirectly holding them constant. Those approaches seek a clearer view by dissecting the behaviour or relationship of interest out of its setting. The risk in those approaches is that the view will be clearer but more limited, often so limited that it precludes drawing reasonable inferences useful to aid providers and recipients.

Our approach is just the opposite, insisting that phenomena must be understood in their context. Unlike laboratory flasks in which chemicals are mixed that do not affect the mixture, contexts for aid and education are active containers that are themselves part of the mixture. Nor are contexts for aid and education like fixed classroom walls. They are people and groups and institutors, with values, preferences, interests, rigidities, fragilities, and will. An effective synthesis requires not their exclusion but rather, their active participation.

Flawed Premises

Evaluations of aid-funded activities have increased in number, complexity, and cost. All involved presume that every aid program and every aid-supported activity requires a formal evaluation, sometimes several. Why? Why are experienced observers' reports and direct feedback from aid recipients deemed insufficient to assess progress? What warrants allocating resources, sometimes very substantial

resources, to an evaluation rather than to the aid-funded activity directly?

Three rationales for the insistence on evaluations are common. The first is the requirement that aid allocations be confirmed to ministries and parliaments and that proposed allocations be explained and justified. The second has to do with monitoring activities and spending. The third presumes institutional learning.

Evaluations of aid-funded education activities can be used to determine that specified activities were undertaken, that the formal requirements of the aid agreement have been met, and that funds were spent and documented appropriately. The primary concerns are to enable the funding agency to confirm that the aid was used as intended and to be able to report on that to its parent ministry or agency, or government, or governing body. That information may be most important when the funding agency seeks renewed or additional funding.

Since the EBA commission for our synthesis did not highlight these two roles for evaluations, we did not address them directly. In practice, however, they may be the most important use of evaluations, notwithstanding the rhetoric that focuses on what works. If so, then evaluations could be simpler, less costly, and less distracting to funding recipients.

The third rationale is the expectation that improved knowledge will improve policy. That rationale rests, it seems clear, on three premises that are engaging and initially persuasive but that have little research support.

Improved knowledge improves policy

The logic of the effort to determine what works—to assess aid effectiveness—is clear. The assumption is that systematic and critical observations of particular activities will generate empirically grounded information about more and less effective courses of action. That knowledge will in turn enable both the aid providers and the aid recipients to select activities that are more likely to achieve desired objectives or to implement particular approaches more effectively or more efficiently. Individuals, and more important, since individuals move on, institutions learn from experience.

Yet, evaluations themselves regularly decry the lack of institutional learning. The same flawed programs, evaluators report, are repeated, with little or no evidence of learning from past

experience. Computers are delivered, for example, without the maintenance and technical support necessary to render them useful. Aid officers are unfamiliar with earlier evaluations and their findings, sometimes even those barely off the press. Focused, critical, and detailed evaluations are regularly ignored or their findings dismissed (the World Bank's annual evaluations are good examples; note the dismissal in the major education strategy paper, World Bank, 2011). Even when operations staff attribute a new project to lessons learned, their explanations reflect much more what might be termed received wisdom, that is, widely articulated general observations, than explicit findings from systematic evaluations. We shall return to this theme.

Quite simply, there is little evidence of direct learning from experiences reported in evaluations and rarely a trace of cumulation of learning from the succession of evaluations over many years. Even more scarce is evidence that whatever is learned from evaluations is appropriated, owned, and used by aid recipients.

The premise that evaluations are a primary vehicle for learning from experience is not supported by available evidence.

Evidence-based policy

The core logic, that evaluations generate knowledge on what works and thus improve policy, rests on a second premise that warrants critical attention. Widespread is the insistence on evidence-based policy. Here too the thinking is clear. Policy that rests on solid evidence about more and less successful courses of action will make foreign aid and the activities that aid supports more effective. Within that thinking, evaluations are understood as the applied research that generates relevant knowledge.

As many years of research on the relationship between research and policy have shown, research influences policy, if at all, through complicated and largely indirect pathways. Hardly ever can one find evidence of a direct march from research to revised policy.

Public policy is a mechanism for addressing and resolving conflicting and sometimes incompatible interests in society. Policy makers must be much more attentive to expressed demands, to constituencies and constituents, to political alliances and coalitions, to financial and other practical constraints, and to broader and narrower political objectives than to research findings. That on many important policy issues the research findings are inconsistent enables policy makers to claim research support for whatever policies they propose.

While that claim of research support is integral to policy debates, in practice, research more often justifies and legitimizes than shapes policy.

Evidence-based policy is an appealing notion. It may well be that through complicated pathways new knowledge has some influence on policy. Improving the base for evidence-informed policy warrants significant effort. But the premise that evaluations play an important role in generating knowledge that directly shapes policy is not supported by available evidence.

Muddling through and satisficing

The core logic that the knowledge generated by evaluations informs and guides education and aid policy and practice rests on a third problematic premise. In that logic, the policy process is understood as largely rational and linear. In this view, whatever the weights assigned to the different inputs, policy makers organize those inputs into policies that are then promulgated and that then guide action. Both the construction of the policy and its implementation are largely orderly, rational, and systematic endeavours.

Here too research suggests otherwise. As Lindblom and others have shown, very rarely is policy making characterized by optimizing objectives and refining approaches (Lindblom, 1959, 1979). Rather, policy making and implementation are best understood as a good deal of stumbling about, trying to find and develop courses of action that are politically tenable and feasible. Little optimization and a good deal of "muddling through." In this perspective, muddling through is not an indication of incompetence or failure but rather an effective strategy for integrating conflicting interests and demands in a contested environment. Most often, policy makers are inclined toward what they regard as feasible solutions, even when they use the terminology of ideal, optimize, and maximize.

With a grounding in psychology and economics rather than public administration, Simon and others have reached similar conclusions (Simon, 1956, 1982, 1997). Rationality is sharply bounded. Policy making can best be understood as satisficing. Policy makers seek and develop policies that will do rather than policies that are best, and policies that are incrementally better than their predecessor rather than radical departures and grand solutions.

A review of the research on public policy is beyond the scope of this synthesis of evaluations. What is important here, however, is

that the premise that making and implementing public policy is an orderly and rational process in which research and scientific knowledge play the central role is not supported by available evidence. If evaluations contribute to policy formulation, it is through chaotic, discordant, and often poorly linked pathways.

Even this brief attention to three flawed premises demonstrates clearly the gap between the claimed role of evaluations—to generate knowledge that permits learning from experience, which in turn improves aid and education policy and practice—and the roles evaluations can play. We note this gap not to decry the constraints on rational policy making and optimization but rather to encourage recognition of those limits and humility in claims about what is knowable, how knowledge is generated, and how knowledge is applied. Where even extensive and expensive evaluations cannot generate knowledge that is useful and used, there is a strong case for evaluations directed more toward and by aid recipients than aid providers.

The Emerging Standard

To explore what can be learned from evaluations of aid-funded education activities we must address the evolution of thinking and practice in evaluation. While expectations and standards in evaluation are not our primary concern in this synthesis, the increasing attention to, and for some aid providers, insistence on, impact assessments and quasi-experimental methods require review here.

Evaluation is central to foreign aid. Researchers and practitioners have long sought to assess the role of development assistance to low-income countries. That is not new. Over time, common practice has changed. Earlier, experienced educators provided informed and detailed observations on aid-funded activities. Increasingly, attention has shifted away from reports on activities to a focus on effectiveness—has the support achieved specified objectives?—assessed through readily quantifiable indicators of intended outcomes. Generally characterized as more scientific, that orientation is expected to yield more reliable and more broadly applicable information. Most recently, there has been a convergence, though not unanimity, on a particular approach: impact evaluations. That is especially visible in the education sector, where it is

increasingly recognized that counting inputs (how many computers were delivered?) and outputs (how many teachers were trained?) says little about whether or not development assistance improves education systems in a meaningful and sustainable way (Chapman and Quijada, 2009). In response, scholars, practitioners, aid agencies, and advocacy groups call for evaluations of aid programs to focus on quantifiable impacts such as student enrolment, attrition, repetition, and test scores (Gertler et al., 2010; Lloyd and Villanger, 2014; Sturdy, Aquino, and Molyneaux, 2014; White. 2007). The ideal model for this approach is the laboratory experiment, to be adapted to field settings.

The clamour for impact evaluation and randomized controlled trials, which we discuss below, is strident, widespread, and influential. Many of the largest funding agencies expect impact assessment to be at the core of the evaluations they commission. There are of course critical voices though even the most prominent struggle to be heard. In his review of a recent synthesis, a major contributor to the development of evaluation strategies noted "my arguments (and even those of other more senior and respected development economists like Angus Deaton) had the effectiveness of a pea-shooter against a tank" (Pritchett, 2015; Deaton, 2009, 2010). Since we find this approach important, potentially useful, but crippled by its assumptions and especially by its narrow gauge, it is important to review it here. We thus join the larger discussion. Since our report is addressed to several audiences, that review has additional importance. While evaluators may be thoroughly familiar with these issues, funding and technical assistance agency staff are likely to be less conversant with debates that have appeared largely in the academic arena, and aid recipients may well not have encountered them at all.

This focus on impact evaluations is made possible in large part by the dramatic increase over the past several decades in the volume of data and the extent of computer power available to conduct statistical analyses linking aid projects to measurable outcomes (Olofsgard 2014; Reddy 2012). Heavily promoted, impact assessments are expected to determine causality, isolating a quantifiable impact of a particular program on a measurable outcome of interest.

To be clear, there are multiple types of impact. Impacts can be direct or indirect, short-term or long-term, intended or unintended, or some mix of all of those. A program can achieve its intended impact, such as improving school completion rates, while having the unintended impact of perpetuating low levels of classroom learning,

for example, as teachers are encouraged to pass as many students as possible in order to meet target completion rates. Conversely, a program can have no effect on a quantifiable impact, such as test scores, while indirectly contributing to an important impact that is more challenging to measure, such as improved teacher morale or increased parental engagement in learning. Evidence-informed policymaking requires understanding these multiple, and at times conflicting, impacts.

However, international efforts promoting the use of impact evaluations most frequently focus on only one type of impact: a quantifiable difference in the outcome of interest (Y) with what is termed the intervention (Y1) and without the intervention (Y0); impact = Y1 – Y0. This type of impact estimate provides tangible, concrete findings that facilitate cost-benefit and cost-effectiveness analyses, thereby enabling policymakers to make informed decisions about how to distribute limited resources (or so the thinking goes).

Central to this notion of impact is the issue of attribution: how can we be sure that the outcomes we observe are actually due to the initiative or reform or program under study? Thus, the primary objective of an impact evaluation is to isolate the effects of a particular program from all other environmental, socioeconomic, cultural, historical, institutional, and political factors that shape the outcomes of interest. In order to do so, evaluators regularly compare observed changes in outcomes (enrolment, attendance, school completion, test scores), to the counterfactual: how outcomes *would have* changed in the absence of the program.

Barring the invention of a time machine, however, it will always be impossible to know the true counterfactual. We cannot go back and forth in time and compare the same students' test scores with and without the program we want to evaluate. Thus, evaluators typically rely on a comparison group of students (or schools, or teachers, or communities) who do not participate in the program, but who closely resemble those who do. While seemingly a straightforward approach, this method proves challenging in practice.

Constructing explanations

In a simple example, suppose we want to know which type of grade 5 mathematics curriculum works best: Curriculum A or Curriculum B. We look at examination results and find that students

who have experienced Curriculum B have scored higher than those who have experienced Curriculum A.

Our initial thought is to associate improved learning with the curriculum. Curriculum B is better, should be funded, and must be implemented across the country. With more detailed investigation, though, we might discover that males have scored higher than females and that when we remove sex from consideration (control for sex), the effect of curriculum disappears.

Similarly, we might explore socioeconomic status and find that higher SES students score higher on the examination than lower SES students. When we control for SES, we find that the effect of curriculum disappears. Or perhaps the underlying critical factor is age: older students score better than younger students, eliminating the effect of curriculum. We might find other explanations that are not immediately obvious. Perhaps the teachers who implemented Curriculum B were better prepared, or more experienced, or had stronger pedagogical skills. In that case, what matters most may be the teachers and their competence, not the curriculum. Or perhaps Curriculum B was implemented in schools that excel in other subjects, or have better facilities, or serve breakfast and lunch, or provide free, high quality supplementary tuition.

In sum, the initial observation that associates curriculum and improved learning may prove to be misleading. To have confidence in that observation we need to be able to control for selection bias. In other words, we need to account for all factors that could confound our results due to their association with both curriculum and test scores. Every factor we add to the list of factors to control increases the complexity and cost of the analysis. Very quickly it becomes clear that in the most common school-to-school or school cluster-to-school cluster comparisons, at best the numbers permit controlling for only a few potentially important factors. As well, the choice of which factors to control and which to ignore rests on prior assumptions and perhaps on research, but certainly reflects sharp disagreements among educators.

Quasi-experimental and experimental research designs

Evaluators interested in measuring a quantifiable impact typically address this problem through a quasi-experimental or an experimental research design. Quasi-experimental evaluations attempt to identify a comparison group that is as similar as possible to those

participating in the program. One example is propensity score matching, in which evaluators use statistical techniques to identify pairs of participants and non-participants who are identical on all observable characteristics deemed important (say, age, SES, religion, gender). Another example is regression discontinuity design (RDD). RDD is the preferred method when there is some selection criterion (such as age, income, or test scores) that determines whether or not individuals are eligible to participate in the program under study. This enables evaluators to compare outcomes between individuals at the threshold: that is, between those who are just above and just below the eligibility criterion (and therefore, it is assumed, very similar to one another).

When done well, quasi-experimental methods can overcome much of the selection bias inherent in the simple comparison between Curriculum A and B described above, but never all. Practical constraints require that evaluators choose a few characteristics that are to be matched (controlled) out of a much larger set of potential influences on outcomes. That selection is not itself the subject of the evaluation. Not infrequently, evaluators do not explain why some characteristics were selected and others ignored, or do not address characteristics that others may deem important. Equally important and problematic is that evaluators commonly monitor individual attributes and ignore community characteristics and collective behaviour. The embedded assumption that communities are simply the sum of atomized individuals is neither presented and defended explicitly nor analytically useful.

Experimental evaluations address the issue of selection bias by randomizing program participation. This is widely considered the most valid and reliable approach—the claimed gold standard—in impact evaluation. The notion is that if large pools of teachers and students are assigned to Curriculum A and B randomly, then there will be no reason to expect the proportion of boys, or skilled teachers, or better equipped classrooms to differ across Curriculum A and B. This method is called a randomized controlled trial (RCT). Previously limited to clinical trials used to test the efficacy of drugs and medical treatments, RCTs are now widely used in social settings, where they have become especially popular among the development aid community. For many, RCTs have become the preferred tool for ensuring that aid money is directed toward activities more likely to achieve intended outcomes (Clements, Chianca, and Sasaki, 2008; Gertler et al., 2010; Lloyd, Poate, and Villanger, 2014; Olofsgard,

2014; White, 2007). Indeed, an estimated USD 150 million was spent on RCT evaluations of education programs in 2013 alone (Pritchett, 2013).

Many evaluators and others are convinced that a great deal of learning has come out of this process. The findings have been used to influence and modify aid policy and programs. Multiple studies from a parenting program in Jamaica, for example, report the efficacy of non-formal, community-based education efforts designed to encourage parents to play an active role in their children's early learning (Gertler et al., 2014; Grantham-McGregor et al., 2007). These findings have been used to direct foreign aid to early childhood education, an area that was mostly overlooked by the international community until recently. RCTs of conditional cash transfers (CCTs) provide another example. Perhaps the best known CCT is the Mexican program Progresa (later renamed Opportunidades), which provides low-income families cash transfers that are conditional on children's school attendance. Multiple RCT evaluations of Progresa have found a positive impact of the program on school enrolment and attainment for rural and socioeconomically disadvantaged students (see Schultz, 2001, for one example). However, the same RCTs have also demonstrated that although CCTs can improve access, they do not or at best rarely improve learning—a finding that has provided an important counterargument to the initial support for CCTs as an easy way to improve educational outcomes.

Some funding agencies are strong advocates of randomized controlled trials. Others prefer them over alternative approaches. Still others rarely commission RCTs (for example, Sida) or consider them to have limited utility (Agence Française de Développement: AFD). Since the insistence that only RCTs provide credible evidence on which to shape policy, since even agencies that do not assign high priority to RCTs report pressure to do so, and since aid recipients have at best a limited role in specifying the evaluation approach, it is important here to review what is problematic.

Randomized controlled trials: theoretical, practical, and ethical problems

The push for RCTs has by no means gone uncontested.

Pritchett's recent commentary highlights major critiques of the logic and theory of change embedded in the claim that increased

RCTs will improve development practice and thereby improve human welfare (Pritchett, 2015):

"Claims that RCTs of impact evaluation could (even in principle) produce useful codifiable knowledge with external validity about development policies and practices were wrong.

Claims about the political economy of policy adoption and scaling were wrong.

Claims about how organizations learn and change practices on the basis of evidence were wrong.

The claim that RCTs would or could address issues of first order importance in development was wrong."

Let us review several of the major problems.

First, RCTs are expensive. In order to achieve the statistical power necessary to identify a causal impact, RCTs require large sample sizes—ideally with an equal number of participant and comparison group individuals—all of whom must be surveyed at least twice (before and after program implementation, baseline and end line). Consequently, the cost of the evaluation may become a major portion of the resources allocated to the project. Aid providers, aid recipients, and the evaluators may all wonder whether the information generated warrants the large expenditure and whether or not those funds might be used more productively elsewhere.

Second, randomization is unfeasible or extraordinarily difficult in many, if not most, poor country education contexts—for practical, political, and ethical reasons. The practical challenge is that most often education initiatives and reforms are implemented in ways that are not readily amenable to the requirements of experiment-like impact assessments. Informed by broad objectives, education ministries and departments organize implementation around available human and financial resources, principals and teachers who can play key roles, availability of requirement equipment and materials, state of facilities, the national and local politics of resource allocations, and a good deal more. As well, since schools are generally community based, student assignments are not readily randomized. Teachers and students move around for non-random reasons. Situational and contextual influences, say flood, famine, epidemic illness, or war, may not be evenly experienced. Sometimes evaluators seek to disrupt decision making and implementation in education, occasionally even to specify the location of new programs and their participants, to

establish the conditions required for an experimental or quasi-experimental impact assessment.

The political challenge is that uneven distribution of resources, in this case improved education opportunities, requires a political rationale and political legitimacy, not simply an experimentalist's specification.

The ethical challenge has three components. Random assignment is incompatible with notions of preference and choice that students, parents, and communities value. Few politicians (let alone citizens!) like the idea of randomly assigning and denying program participation. Societies generally frown on experimenting on humans, even where a positive result can be expected. Kenyan parents, for example, reacted strongly to the inclusion of de-worming medication in school meals, both because they were not consulted and because in the parents' view, the authorities were treating their children as farmers might treat pigs.[1] As well, standards for RCTs that have evolved initially in the health sector require that those who do not participate in the new activity have access to what is regarded as the best current practice. The searing lessons of the Tuskegee experiment have made no-treatment an unacceptable comparison to the treatment that is to be assessed.

Of course, financial and institutional constraints mean that new programs rarely reach simultaneously all of the individuals who could potentially benefit from them. Those excluded or not yet included thus constitute a comparison group. For this reason, RCTs often make use of programs that are gradually phased in, randomizing the order in which participants (individuals, families, schools, or communities) become eligible to participate in the program (Gertler et al. 2010). Moreover, RCTs do not always require a comparison group that receives no new or revised program (treatment). The same method can be used to compare two implementations of a particular program. In this case, the question becomes "which variant of Program X is more effective?" rather than "is Program X better than no program?"

[1] Deworming was popularized as a cost-effective strategy to improve educational outcomes after an RCT in Kenya found positive effects of deworming campaigns on school attendance (Miguel & Kremer, 2004). These findings have since been challenged by a group of epidemiologists who conducted a replication analysis using the same data and failed to find similar results. The resulting controversy, deemed, "The Worm Wars," is a reminder that even RCTs do not provide unequivocal answers to policy-relevant questions.

However, in contexts where there is reason to believe that certain schools or students may benefit more than others from a particular program (or program variation), the ethics of random assignment are problematic (Fives et al., 2015). As Reddy (2012) points out, the fact that the current boom in RCTs has almost exclusively involved experimentation on poor people in low-income countries is not a coincidence. It is much less politically palatable to assign a particular program or benefit randomly to middle or upper class recipients.

RCT advocates counter this criticism with the argument that even if there are reasons to believe that a program is beneficial, it is unethical not to do an RCT. This argument is based on the "honest null-hypothesis." According to the honest null hypothesis, despite the fact that policy makers or educators may think a particular program will improve learning, the null hypothesis, that the program has no effect, cannot be rejected until the program has been empirically evaluated (Fives et al.,2015).

In the context of aid to education the honest null hypothesis holds particular sway. Billions of dollars have been spent on aid to education, but to what end? At least 60 million children remain out of school, and of those who do attend, many complete basic education without mastering basic literacy and numeracy skills (UNESCO, 2015). In light of this slow progress towards sustainable educational development, the argument is that it is ethically imperative to evaluate the impact of aid to education. The ethical claims clash. It is unethical to implement a program without a systematic evaluation vs. the evaluation itself is unethical or includes unethical components. The history of experimentation on human beings suggests that the latter must have priority.

Third, the health and education sectors differ sharply. RCTs are arguably the best method available to do what they were initially designed to do: identify a causal impact between a medical treatment and changes in health outcomes. The health metaphor does not apply to education as seamlessly as is often assumed, however. RCTs require a stable treatment, a short and straightforward causal chain, and a large group of individuals who are directly affected by the treatment, but who have a very limited capacity to change how the treatment operates (Bernard, Delarue, and Naudet, 2012; Reddy, 2012). This makes sense in clinical settings, where researchers ensure that everyone in the treatment group takes the same exact dosage of the pharmaceutical

product being tested, while everyone in the control group takes a placebo, or a lower dosage, or no treatment at all, but no less than the best current treatment. No one in either group interacts with the treatment in an intentional manner to change the way that that it operates—they simply swallow the medication (or the placebo).

Education programs rarely meet these criteria. Differences in institutional capacity and resources between schools and communities, along with socio-political and cultural differences, mean that program implementation (the treatment) is rarely stable. Program participation rates and implementation practices vary significantly between schools, communities, and districts, even when they are randomly selected (Bernard et al., 2012; Culbertson et al., 2014). The effort to apply health sector experiences to education is superficially attractive but ultimately unhelpful. To regard education as a vaccine, or to look for education's vaccine—focused action that can directly, sharply, and quickly change outcomes—is not productive. At best there might be an analogy to managing chronic conditions rather than avoiding or curing illness, but evaluation of education activities must incorporate its distinctive purposes, forms, and characteristics.

Fourth, education development arguably *should not* meet the requirements of RCTs. The quality and sustainability of any educational program depend on the program's capacity to adapt to the local context, respond to operational challenges quickly and organically, and encourage participation and buy-in at all stages (Chapman and Quijada, 2009; Grindle, 2010; Riddell, 2012). Effective learning settings are interactive, regularly modifying how they do things. Especially where few learners are able to progress to higher school levels, education process is far more important than the commonly measured outcomes. Thus, differences in program implementation are important, and should even be encouraged, rather than stifled in the push for a stable treatment.

By prioritizing certain types of evidence over others, we ignore ideas that do not fit the mechanistic input-output notion of social change. The only valid solutions become those in which "interveners within a system are viewed as standing outside it, and their possible actions are well defined and without reference to how the *system acts upon the intervention*" (Reddy, 2009; p. 64, emphasis added). If all efforts to improve the quality of education systems around the world are based on this model, then achieving inclusive and

sustainable educational development becomes more difficult, likely unattainable.

The importance of evaluating aid-funded education programs is clear. However, what is the risk of prioritizing particular impacts or types of impacts, and then a particular approach to evaluation, above all others?

Fifth, like all types of evaluations, qualitative, quantitative, or mixed methods, the findings of an RCT are specific to the context and to the conditions under which the evaluated program operates. This is particularly true in the case of education, a highly complex, and often politicized process. Education systems, both formal and non-formal, function in specific historical, socio-political, cultural, and economic contexts and thus cannot be understood fully without attention to their interaction with those contexts. Likewise, as Pritchett and Sandefur argue, the importance of context is especially pronounced in RCTs (2013). The high political, financial and ethical costs of implementing RCTs mean that the contexts (people/places/organizations) that chose to use RCTs are atypical. This poses a challenge to the external validity of RCTs. Context matters.

Still, implicit in the prioritization of RCTs is the idea that if we conduct enough randomized evaluations we will eventually understand what works to improve education outcomes. This is not to say that the advocates of RCTs believe in a magic bullet solution. However, central to the claim that impact assessments with RCTs are the most scientific and most reliable form of evaluation and thus the standard against which other approaches should be assessed is the notion that if we conduct enough experiments we will eventually uncover blue-print like solutions that are transferable and scalable (Abhijit Banerjee & Duflo, 2011).

Sixth, a recent review of six systematic reviews of evaluations of education programs in low-income countries calls into question this presumption (Evans and Popova, 2015). All six systematic reviews included in the study consist primarily of RCT evaluations, but the authors find almost no overlap in the conclusions drawn from these evaluations—dramatic discord where we expect consensus. That is not the fault of RCTs as an evaluation methodology. Rather, that finding provides further evidence that *no* volume of evaluations and data collection will ever uncover a blueprint of *what works* for education. The conflicting results of numerous experimental and quasi-

experimental estimates of the effects of class size provide another example. In Tennessee (Krueger, 1997) and Israel (Angrist and Lavy, 1999), studies found large and statistically significant effects of class size reductions on students' test scores, while in Kenya (Duflo, Dupas, and Kremer, 2009) and India (Banerjee, Cole, Duflo, and Linden, 2005) there were no effects. Even something as seemingly straight forward as reducing the number of students per teacher was found to have a strikingly different effect on student learning in different places.

Indeed, that search for a blueprint, or set of standard approaches or practices, is not productive. There is no blueprint that can be identified and scaled up. Learning is a participatory, interactive, and dynamic process, deeply intertwined with the political, economic and historical contexts within which formal and non-formal education take place.

Neither more RCTS nor more refined RCTs—no reliance on any single approach to evaluation—will uncover *what works*. The socioeconomic and political transformations in high-income countries that we term developed was neither dependent on nor significantly shaped by RCTs. No one claims that the world's most equitable and high performing education systems were built through RCT-driven policymaking (Pritchett 2013).

Both the limitations of RCTs and the practical, financial, and ethical challenges in their implementation lead to the conclusions that while RCTs can play a useful role in evaluating aid-funded education activities, their effective domain is constrained, and that certainly neither RCTs, nor impact assessments more generally, are the standard against which other approaches to evaluation must be assessed. Improving aid to education requires a comprehensive, inclusive, and contextually grounded understanding of role and consequences—one in which RCTs are one of many evaluative tools, but not the only valid tool.

It is important to note here that the Expert Group on Aid Studies has commissioned a parallel synthesis that we understand will focus largely or entirely on impact assessments and randomized controlled trials. That synthesis, we trust, will provide rich and documented attention to the strengths and limitations of impact assessments and RCTs and to the situations where they are most useful and cost-effective.

When Method Determines Outcomes

Our concern in this section of our report has been to lay the foundation for our review of evaluations of aid-funded education activities. The frequent insistence that only impact assessments and quasi-experimental methods warrant inclusion in that review required critical attention to those approaches. The availability of large datasets and expanded access to the computing power necessary to manipulate and analyse them require similar attention to the risk that confidence in seemingly sound methodology may disable critical assessment of reported findings and related interpretations.

In his penetrating critique of the validity, reliability, and use of the most common measures of economic growth in Africa, Jerven shows in careful detail the analytic and policy consequences of assuming that if the method is correct, the results must be correct (Jerven, 2013, 2015). The dominant strategy for explaining poverty in Africa has been to use regression analysis, a statistical tool that permits estimating the strength of relationships among variables of interest. Earlier focused largely on economic policies, analysts are now more attentive to political institutions and governance. How do they know their understanding is correct? Confidence in the method produces confidence in the explanation. Where there seems to be contradictory empirical evidence, the faith in the method leads to efforts to find flaws in the apparently contradictory evidence rather than to re-think the approach and method and their consequences.

Jerven's critique is both technical and analytic. The starting premise of a good deal of the writing on development in Africa, that poverty and stunted growth are Africa's standard condition, is empirically wrong, largely a function of selecting a limited time period and failing to recognize earlier and subsequent growth. The basic data on gross domestic product have a wide margin of error and are at least as much political as technical:

"the most basic metric of development, GDP, should not be treated as an objective number but rather as a number that is a product of a process in which a range of arbitrary and controversial assumptions are made." (Jerven, 2013: 121)

Although economists regularly reiterate the cautions that correlation is not causation and garbage in, garbage out (using flawed data will yield flawed results), they often ignore both. Relying on regression, they establish correlation and then work backward to fit a plausible

explanation to the observed correlation, and then use that to derive policy recommendations. To strengthen their claim, the recommendations are then presented in the form "Research shows that . . . ." Alternative explanations can be rejected as unscientific. Confident in their approach, they regard employing the correct method as confirmation of the results it produces. The method becomes self-validating, largely impervious to error margins and conflicting data.

Note here the consequences of inattention to context and complexity. A broader historical analysis shows both the situational specificity of the period of slower growth in Africa and the large margins of error in GDP numbers. That shows as well that the economists had posed the problem poorly. What needed explanation was not Africa's inherent or characteristic slow economic growth but rather Africa's strong economic growth over several centuries and the periodic declines in particular countries. Reposing the question also shows that causal factors cannot be located entirely within Africa but clearly include Africa's relationships with the global political economy.

That faith in method also plagues analyses of education in Africa. For example, a 1995 World Bank policy review declared that "Human capital theory has no genuine rival of equal breadth and rigor" (World Bank, 1995: 21). The associated method, rates of return analysis, has "withstood the tests of more than three decades of careful scrutiny." Accordingly, although both the approach and the calculation were criticized from the outset, the World Bank's calculated rates of return on education in Africa provided the rationale for the focus on basic education—the research underpinning for the campaign for education for all—and for the insistence that education resources be redirected from higher to basic education. In hindsight, most observers agree that starving Africa's universities reduced the quality of education in general and likely impeded Africa's development well into the future.

Our primary concern here is not with explanations for Africa's economic growth or with the consequences of impoverishing higher education, but rather with the risks of relying on a single approach or method and of assuming that if the method is correct, its results and recommendations must also be correct. In our synthesis we have employed multiple methods and approaches rather than privileging a single method, however scientific its aura. Minimizing the

risk of evaluator bias requires engagement with educators, decision makers, and communities, not distance from them. Systematic and critical attention to complexity and context are essential for assessing the utility of a proposed approach or method and its findings. *It is that attention to history along with (flawed) quantitative data, to educators and learners and their voices along with detached observers, and to experience along with statistical analysis that make an approach scientific.*

An Integrated Approach

What can we learn from evaluations?

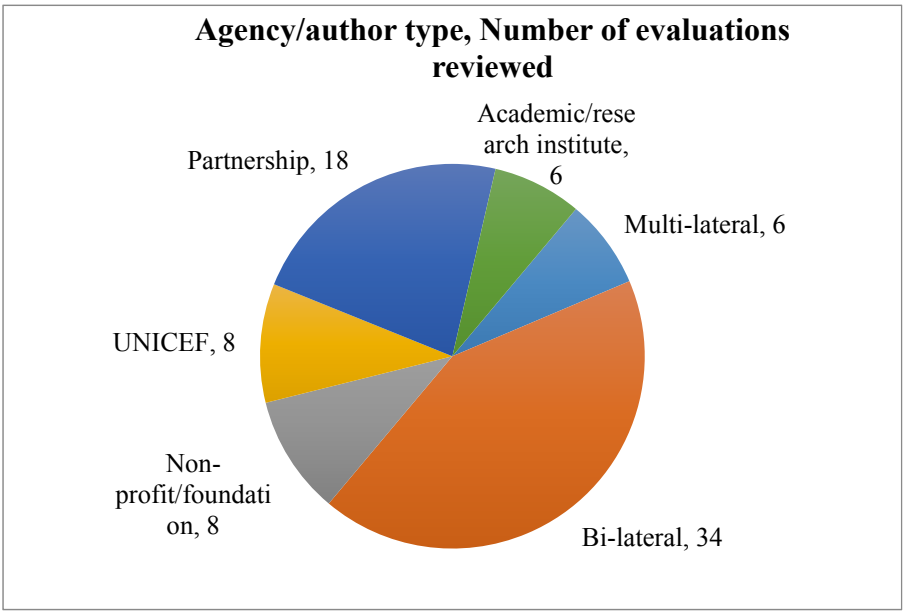
Our task is not a new one. Efforts to identify *what works* to improve aid to education are numerous (see McEwan, 2015, Masino and Niño-Zarazúa, 2015, and Krishnaratne, White and Carpenter, 2013, for three recent examples). We build on these efforts by going beyond what works (say, to raise enrolment rates, to raise test scores, to keep girls in school), to explore what we can learn from evaluations about the intersection of education, aid and evaluation. To do so, rather than relying on academic research, we focus on evaluations conducted by and for those who are directly involved in the aid relationship, since these are the evaluations that are expected to be directly linked to changes in practices and policymaking. From this, it follows that the ultimate value of an evaluation depends on the extent to which it enables funding agencies, governments, education officials and educators to improve their practices. Thus, where possible, we explore how different constituencies, from funding agencies, to implementing organizations, and aid recipients, *use* evaluations.

In his recent comment, Pritchett distinguishes between knowledge-focused evaluations and decision-focused evaluations (2015). The former are intended to contribute to development theory, while the latter focus on improved decisions by funding agencies at the agency or program level. He adds a third category, accountability-focused evaluations, for which he emphasizes the independence of the evaluator. Others have proposed additional categories, for example, learning evaluations, or combinations of those categories. Strikingly absent from nearly all of those discussions is a notion of evaluations primarily by and for the recipients of the aid, perhaps recipient-focused-evaluations, concerned not only with improved decisions but directly with the programs and health of schools and other recipient

organizations. With some exceptions, similarly absent are process-focused evaluations whose primary focus is education and learning as a process rather than an outcome and whose primary audience includes aid recipients. Since our orientation is inclusive, we do not use those or other categories to determine which evaluations to review, but rather to be confident that the evaluations we have reviewed include both those that conform to the common models and others that do not.

Note that our focus is evaluations as a set, not individual evaluations. We are not asking whether or not a particular evaluation provides clear findings that might guide action. Rather, we are exploring what can be learned from the broad range of evaluations undertaken within the aid relationship. Recognizing that a well supported finding cannot improve education if it is not applied, we explore as well the uses of evaluations.

To begin, we conducted a comprehensive search of evaluations of education activities commissioned by international and national funding and technical assistance agencies, the OECD Development Cooperation Directorate, UNICEF, education-focused NGOs, as well as prominent education-focused research institutes and consulting firms (Annex D). A guiding concern was to develop a set of evaluations diverse in approach, commissioning agency, specific focus, and involvement of aid recipients. That is, we sought to maximize diversity, not quantity. Cognizant of the need to be as inclusive as possible across a wide range of aid providers, no exclusion criteria were applied at this phase beyond the requirements that the evaluation examine education activities that were at least in part aid-funded, that the evaluation be published after 2005, and that the evaluation be written in English, French, or Spanish. This search resulted in an initial list of 80 evaluations. From this set we selected a subset of 40 evaluations for more detailed review. Through our subsequent examination of those and other lists we are confident that the selected set reasonably reflects the broader universe of evaluations of aid-funded education activities.

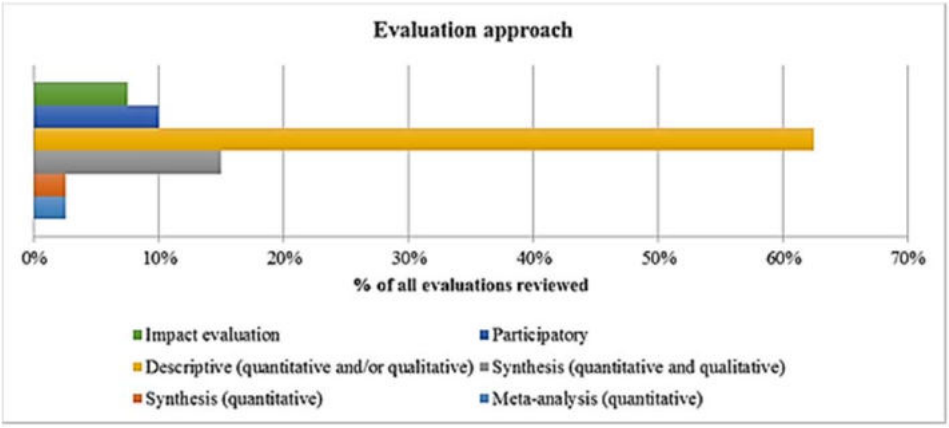


Our selection and review process is informed by realist synthesis, a methodology designed to explore complex and varied programs applied across multiple contexts (Greenhalgh, Wong, Westhorp, and Pawson, 2011; Pawson, 2002; Pawson and Tilley, 1997; Westhorp, Walker, and Rogers, 2012). The objective of a realist synthesis is to achieve depth of understanding by exploring context, mechanisms, and processes that lead to outcomes and impact, rather than to produce a verdict on a program's effectiveness. Especially productive for our multi-mode, multi-method review strategy, the realist approach challenges the common approach to explaining causality. Rather than focusing on does A cause B, the realist approach explores the circumstances, interactions, and institutions—generative mechanisms—that make it possible for A to influence B. That orientation assures attention to complexity and context. A realist synthesis draws from a diverse group of purposively selected studies, using two main criteria: (1) relevance (to the theories or concepts under exploration), and (2) rigor. Importantly, rigor refers to the adequacy and appropriateness of the methods used in relation to the context, interactions, and processes under study, rather than to the evaluation's internal or external validity.

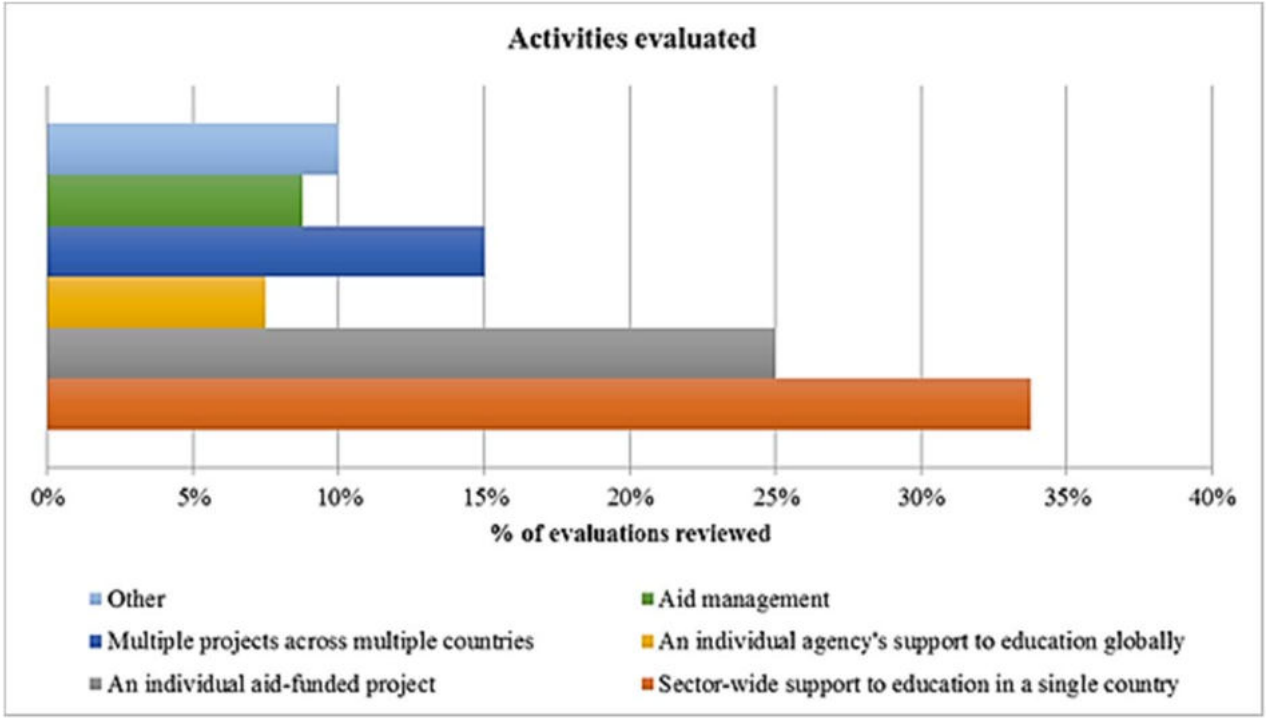
Our synthesis draws on these criteria and adds a third: *diversity*. We therefore modified and added to the subset of evaluations selected for more detailed review in order to ensure that

the studies to which we gave most attention reasonably reflect the diversity of funders, implementers, programs evaluated, contexts, and methodological approaches present across the 80 evaluations we initially identified. That attention to diversity among the evaluations reviewed enables us to report on what we learn from the evaluations as a set, rather than particular evaluations.

Our synthesis is also informed by the Real World Evaluation (RWE) approach developed by Bamberger et al., which addresses constraints of budget and time, missing baseline data, and political pressures, which may put at risk sound research design. That is, given rushed, incomplete, and otherwise flawed evaluations, how to assure methodological rigor in reviewing and synthesizing them? As we have noted, we consider the common strategy—ignore flawed evaluations—a problem, not a solution. Even flawed evaluations may yield useful information, about both aid-supported education activities and the evaluation process. RWE responds to specific constraints unique to particular contexts, where for example, an evaluator might need to reduce sample size. In this instance RWE presents techniques to ensure statistically acceptable standards in that constrained environment. A mixed methods approach can manage those constraints, especially where the supported education activities are complex and the environment challenging. Anchored with a Real World approach, inclusive of local voices, experimental evaluation methods may strengthen attention to complexity, context, and process in the effort to understand outcomes and assess impact.



To promote rigor and consistency across readers, we developed a common list of dimensions and assessment criteria that we used to select and classify evaluations. These are: relevance; program description; evaluation objective; approach; rigor; target audience; participatory evaluation; explicit assessment of process; explicit assessment of outcomes; external quality measures; activities evaluated; lessons learned; utility (see Annex D.2). A thorough reading of each evaluation permitted classifying them as strong/moderate/weak across these dimensions. Did they provide clear information for each dimension? Were the observations, interpretations, and conclusions explicitly supported by relevant evidence? We used these classifications to guide our selection process, not as strict inclusion/exclusion criteria.


Activities evaluated
0%
5%
10%
15%
20%
25%
30%
35%
40%
% of evaluations reviewed
Other
Aid management
Multiple projects across multiple countries
An individual agency's support to education globally
An individual aid-funded project
Sector-wide support to education in a single country

As we use common selection criteria to determine which evaluations warranted fuller examination, we were at the same time attentive to the ways in which the selection process itself may constrain or specify the eventual findings. To that end, the evaluations selected for detailed review scored high on most of the quality dimensions described above, but not necessarily all. That enabled us to explore directly the complexity of the relationship among education, aid, and evaluation.

To review the 40 evaluations selected for focused attention we developed an interactive and iterative analytic process. Rather than locating the synthesis after the completion of our review of evaluations, we synthesized while we reviewed, identifying, connecting, and exploring concepts about education, aid and evaluation as they emerged. Throughout the review process we asked, *what observations or findings of more general interest have emerged? What do these emerging ideas tell us about education, aid and evaluation?* Likewise, in our effort to build on the syntheses that have been conducted to date, we asked, *given what we know from the existing syntheses of aid to education, what might we expect to see in evaluations? Do we find what we expected?*

We used these questions to develop and test propositions against the broad set of evaluations without losing their content, nuances, and details. Importantly, rather ignore the evaluations not selected for focused attention; we drew on these evaluations throughout the review process in order to test our hypotheses and identify common strengths and weaknesses across the larger set of evaluations.

In addition, from the larger set of evaluations selected for focused review, we selected three for detailed assessment across multiple layers:

1. *Evaluation of ICT in Teachers' Colleges Project in Tanzania*, conducted by InDevelop for Sida (2014)

2. *Joint Evaluation of Nepal's Education for All 2004-2009 Sector Programme*, conducted by Cambridge Education Ltd. for Norad (2009)

3. *Évaluation à mi-parcours du Plan décennal de développement du secteur de l'éducation du Bénin*, commissioned and conducted by France (AFD), Denmark (DANIDA), and Bénin (MCPD) (2012)

We selected evaluations for case study analysis based on the criteria described above (relevance, rigor, diversity) and *feasibility*. That is, we chose evaluations for which we were confident we could establish direct contact with the funding agencies, implementing partners, and aid recipients involved.

Guided by an appreciation for the diversity of evaluation purposes, approaches and uses, each case study asks:

1. Through what processes do organizations determine what to assess, how to assess, and how to *use* evaluation findings?

2. Which sorts of evaluations are most useful for different constituencies involved in aid to education, and why?

3. What evidence is there of evaluation-induced learning or change?

To answer these questions we conducted open-ended interviews (and an e-mail questionnaire, for the Nepal/Norad case study) with actors at different constituencies involved with the evaluations under study (see Annex D.3 for a full description of each case study). To the extent possible, we sought to trace each case study evaluation from the policy and decision makers who commissioned it, to consultants who conducted it, to aid recipients involved in the production of the evaluation and/or the implementation of the funded program.

In practice, this proved much more challenging than we anticipated. Thorough investigation and frank discussions at all stages required a dense network of contacts and face-to-face interaction. The time and budget available for this synthesis limited what could be accomplished. Many of the case study interviews were conducted remotely (via Skype and electronic mail). Each case study includes interviews with officials from the funding agency and aid-recipient governments involved in the evaluation with the exception of the Nepal/Norad case study, for which it was not possible to interview those in the aid-recipient country who were directly involved in producing the evaluation under study. Regardless, the inclusion of voices from a diverse group of actors responsible for different aspects of aid to education is an integral component of our synthesis. The findings from the case study analyses inform and shape our findings from the larger set of evaluations, recognizing and building on the context specificity of both evaluations and their use.

Ours is a modest synthesis, aimed at in depth and detailed analysis, rather than identifying and classifying every evaluation that has been undertaken. A broader search strategy, especially a country-by-country and agency-by-agency investigation, could have yielded a much larger pool. That effort was outside the scope of our commission and would have required effort better allocated to analysis. More important, we are confident that the evaluations included in our synthesis adequately reflect the set of evaluations of interest. Moreover, to the best of our knowledge, our synthesis is the first to include such a diverse sample of evaluations (in terms of methods used, types of policies and programs evaluated, funding agencies, countries and contexts) and to address evaluations as a set rather than focusing on a few well-grounded evaluations of particular activities.

Cognizant of their limitations, our synthesis draws on a diverse group of evaluations, including some that employ RCTs and many others that neither seek to identify a quantifiable impact nor rely on RCTs to do so. Our approach requires attention to evaluations that others might ignore or discard. This is because we insist on a holistic approach to education, aid, and evaluation. In this, our approach is consistent with numerous scholars across a range of disciplines who argue that rather than *what works*, it is important to know *how, why*, and *in what circumstances* policies affect outcomes (Deaton, 2010; Greenberg & Shroder, 2004; Tikly, 2015; White, 2009).

This approach requires reviewing evaluations of individual programs with clearly defined participants and non-participants, as well as evaluations of sector wide programs that cannot be assessed through experimental or even quasi-experimental techniques. This approach requires examining evaluations that explore education and aid as an interactive, dynamic process, and where possible, evaluations that regard program participants as co-evaluators, rather than subjects. In sum, just as there is no one size fits all education program, there is no one size fits all evaluation. Our synthesis, therefore, explores a diverse group of evaluations, selected not for their claimed methodological rigor, but for their combined capacity to speak to the multi-layered complexities of education, aid and evaluation.

3. Evaluations of aid to education in poor countries

We turn now to our review's major observations, followed by our commentary and analysis. Our major concern was to identify points of commonality and difference across diverse contexts, thereby facilitating triangulation of findings across different evaluation approaches and perspectives and permitting generalization while simultaneously accounting for differences in context. Details are annexed.

While a few evaluations assess the impact or utility of a particular education innovation or strategy, most address the implementation of specified aid funding (say, support for textbook revision or increasing girls' enrolment) or aid to the education sector more generally (for example, multi-agency support to a ten-year basic education development plan). Accordingly, their primary findings concern the intersection of education initiatives and their implementation, the forms and efficacy of aid, and the strengths and limitations of alternative evaluation strategies. Many address the education environment, for example capacity building, knowledge transfer, local ownership of education initiatives, decentralization, and institutional learning. Not infrequently, recent evaluations confirm observations and interpretations that are regularly discussed in the education and aid communities.

That few well-grounded and broadly generalizable findings emerge from a broad set of evaluations of aid-funded education activities over the past decade should not surprise us. Notwithstanding the periodic search for global best practices, effective learning approaches are necessarily tuned to setting, place, and time. Even in centralized national systems, the practice of education is largely local and thus situationally specific. Several evaluations confirm the very local character of education (among them, Eval: Sida, 2005).[2] As that evaluation insists, primary findings are specific to particular settings and cannot readily be generalized, notwithstanding the regular inclination to do so.

For example, an evaluation might examine achievement outcomes associated with the introduction of new instructional materials or the implementation of a new instructional approach. As it

[2] Citations in the form "Eval:" refer to evaluations reviewed and listed in Annex A.

does so, the evaluation can confirm to those involved that the aid funds were used as intended, that the planned activities were undertaken, and that the innovation was or was not associated with, say, higher examination scores. That might well provide sufficient evidence for a funding agency and recipient government to proceed with the innovation, or alternatively to drop it. But in the absence of confirming evaluations over an extended period and in other settings, that evaluation does not provide solid grounds for generalizing beyond the setting studied.

Periodically evaluators, funding agencies, and less frequently education ministries, do just that. The apparent benefits of, say, a new method for teaching mathematics or for encouraging girls' enrolment is deemed so clearly preferable that it is applied elsewhere. Subsequent assessments then report that the new method has worked well in some settings, poorly in others, and in some not at all. Accordingly in our review we sought evaluation evidence of innovations and reforms tested over time and in diverse settings, with less weight assigned to those reported as successful in a single evaluation or setting.

Evaluations might have a holistic or system mandate. Incorporating elements of the Real World Evaluation approach Eval: UNICEF 2012 notes that evaluating the activities, process, and results of education programs requires examining the planning context, the specific activities, the results and the impact of the program, with particular attention to successes, weaknesses, and constraints during implementation. In practice, few evaluations have that broad mandate.

That evaluations are commonly expected to focus narrowly on a specific activity and are rarely accorded sufficient time to cast a broad net limits the breadth of the observations they report and the explanations they develop. While many of the evaluations we have reviewed set out to investigate the effectiveness of education activities, or the education sector as a whole, most end up describing, whether directly as a finding or indirectly, a limited set of specified activities and their immediate outcomes. That is, the structure of the evaluation itself regularly constrains its scope and reach, making it difficult for evaluators to explore critical political, economic, and social influences on education outcomes or to analyse in depth the process of knowledge transfer, capacity building, decentralization, and institutional learning. Where evaluations seek to control for, that is analytically ignore, the effects of complexity and context, that constraint is even sharper.

Unfortunately, far too many evaluations are frustratingly ahistorical or inattentive to relevant previous experiences and earlier research. Perhaps pressed to narrow their focus or to complete their work quickly, evaluators regularly present as a new finding an observation that could be substantially enriched by linking it to earlier experiences or to findings on similar education activities elsewhere. Especially frustrating to those who seek to use evaluations to improve both education and aid are evaluators' limited efforts to *explain* what they have found and to seek explanations for persisting problems they identify.

Major Findings: Education

In the discussion that follows we highlight several well grounded findings that emerge from the evaluations reviewed. Since our task was a broad canvas of evaluations rather than the assessment of particular education initiatives, the well-supported findings concern systemic change efforts and the role of foreign aid. Experienced educators will find many of those observations familiar, confirming understandings regularly discussed among educators and funding agencies. Those confirmations are not unimportant. What is taken for granted may turn out to be incomplete or incorrect. An important added value of our synthesis is that we are especially attentive to the context in which particular initiatives have been found to be effective. Providing an input, for example, may improve outcomes more when it is accompanied by a modification of teacher education or administrative reform. An initiative may affect outcomes indirectly, for example, creating school clusters that in turn increase community engagement, that in turn generates support for quality improvement through an expanded reading programme. Since the focus of our synthesis is on what we can learn from evaluations, we have not sought to supplement (or confront) these observations with a parallel review of relevant scholarly research: "The synthesis evaluation should be focused on evaluations and not synthesize research more generally" (TOR).

For our overview we distinguish between a finding (directly linked to the evaluation data and objectives) and the comments that follow (the evaluators' conclusion or recommendations, but that are not observations derived directly from the method employed to evaluate the education activity).

Effective education efforts reach beyond schools

Evaluations of aid-funded education activities provide confirmation and rich evidence: effective education efforts reach beyond inputs and beyond schools. A clear example is efforts to achieve education for all. The most effective strategies for increasing enrolment appear to be reducing the costs for families combined with sustained advocacy and awareness activities (Eval: GIZ 2012). Many evaluations note that aid-funded education programs that focus on input provision are successful only where they are accompanied by substantial efforts to work with educators, officials, students, and families, to develop skills in using the input (for example, textbooks, computers, and tablets) effectively (Eval: IADB 2013). For instance, in the Democratic Republic of Congo awareness activities had a positive impact on political support for education and facilitated cooperation between aid funders and recipients (Eval: UNICEF 2012).

Initiatives that reach beyond schools are especially important in the effort to reach universal school access. An evaluation of Sida support confirms that the determinants of access to schooling are context-specific as shown by the three cases: Bolivia, Honduras and Nicaragua (Eval: Sida 2005).

Another example is initiatives to improve education quality. Organizing schools in clusters—creating small groups of schools that work together— facilitates community participation, which in turn increases support for instructional quality reforms through partnerships with parents, community, NGOs, and other government institutions that provide social services (Eval: MiET Africa, SDC, EKN 2009).

Inputs are not enough

As a set, evaluations develop this theme further. Most aid programs focus on inputs of some support. The inputs vary. In one setting, facilities. In another materials (textbooks; computers). In still another, professional support (examination design; accounting systems) or services (access to the internet). Only rarely do aid programs embed the provision of inputs in a larger frame that is attentive to the supports needed for the inputs to be used well, to who is responsible for receiving and managing the inputs, to needed on-going support (including technical assistance and maintenance), to integration into the national and local education system, to responses

by teachers, learners, and communities. Research on education and experience with foreign aid have over many years indicated that providing inputs without accompanying attention to the environment in which those inputs must function yields limited results. Earlier evaluations have confirmed that understanding.

The evaluations we have reviewed provide additional support for that assessment. Aid programs that focus entirely or primarily on inputs are less effective than those that start with a holistic notion of education as a process and education as a system and that embed that understanding in the aid program. Across different education domains—curriculum, instructional materials, pedagogical approaches, teacher education—evaluators find that providing inputs without simultaneously addressing how those inputs are to be provided, to whom, in what circumstances, and with what accompanying authority and resources, sorely limits the utility and effectiveness of those inputs. Indeed, even well-designed and potentially useful inputs may lie unused. Inputs that are not accompanied by parallel work on their operating environment are less effective. An evaluation of French support notes that making more narrowly focused aid effective requires improved teacher training, attention to improving the school environment, support to South-South cooperation, and country-specific practices (Eval: French Ministry of Foreign Affairs 2007).

An evaluation of Swedish aid suggests what is needed to make an input, in this case support to school committees in Zanzibar, effective: "the roles of school committees will need to be broadened and capacity strengthened to ensure their participation in managing schools is done in a more meaningful way" (Eval: Sida 2007b). Drawing on document analysis, interviews with education officials, and group discussions with educators and community members, the evaluation notes that "outputs are evidence," that is, schools have been constructed and classrooms have been refurbished. Missing, the evaluators note, are "well thought out framework and methodology" and clarity on overall objectives and priorities. The emphasis on consultancies and training, another type of input, is insufficient to link the input with intended education outcomes.

Education innovations often focus on producing a specific output (for example, a new training methodology, a new financing or budget system, new tablets or computers) without complementary attention to the needed infrastructure, labour skills and policy (Dahlman, 2013). In practice, there may not be effective mechanisms

and governance systems to coordinate stakeholders at the local, regional, national, and international levels (OECD, 2005).

The evaluation of Swedish Support in the Education Sector in Zanzibar, 2002 2007, stands out for its focus on "progress as processes," which requires focusing on systematic issues set in the context of constraints at the national, district and school levels (Eval: Sida 2007b).

Recipient governments regularly encourage the focus on inputs. For example, they may request computers, despite limited evidence that computers improve learning (Eval: IADB 2013). Sometimes evaluations contribute to the focus on inputs. An evaluation of aid-supported provision of technology reports on the number of teachers who use computers in their classrooms or who are satisfied with the new approach, but only in its supplementary comments addresses what is needed for the new technology to improve learning (Eval: Sida 2014).

Riddell provides important cautions here (2012). Funding agencies that focus largely or entirely on demonstrable short-term impact contribute, perhaps unwittingly, to undermining long-term impacts on education systems. Similarly, where the focus on inputs is associated with an insistence on demonstrable short-term impacts, the longer term consequence, is weakening rather than strengthening the education system.

In her review of aid effectiveness Riddell demonstrates the distortions caused by focusing on enrolments and insufficiently on quality, on products such as plans and educational management information systems, and inputs, rather than processes and outcomes, what goes on in the classroom, what the students learn, whether the teachers' pay and status are sufficient to keep them in the classroom and continuing to teach (2012).

The limitations of the focus on inputs are clear even where the input is an advice or a service. An evaluation of Swedish support argues against a technocratic approach to results-oriented budgeting, since resource allocation is a negotiated process that must consider not only the expected impact of policies and spending, but also political and economic context and institutional arrangements (Eval: Sida 2005). Several evaluations insist on the importance of attention to documenting the use of inputs. One recommendation that emerges from an evaluation of USAID support to education in Guinea is that for complex and multifaceted educational programs, the collection and

analysis of program implementation documentation is critical for developing a deeper understanding of effectiveness (Eval: USAID 2006).

Although they note this problem, evaluators may contribute to it. With rare exceptions, even as they report that funding agencies provide inputs without sufficient attention to what is required to make those inputs effective, evaluators do not explore why aid agencies continue to go down that path. Across the evaluations we have reviewed, we do find a call for political economy analysis and a more holistic approach. Yet, in general evaluators do not seek to close or even address the gap between broader development goals (poverty, social inclusion, human rights, democracy, sustainable development) and supported education activities (Eval: Sida 2013).

Several evaluations highlight the importance of effective communication, awareness activities, and the widespread diffusion of information along with the provision of inputs.

Overall, evaluations have confirmed that the focus on inputs is not sufficient to achieve intended objectives and is regularly in tension with efforts to assure program sustainability, to reduce inequality, and to reinforce capacity building. Yet funding agencies continue to focus entirely or largely on inputs, and evaluators mostly do not ask why or explore the broader consequences of that focus.

Effective external support reaches beyond the education ministry

Just as the focus on inputs is limiting, so too can be concentrating attention on the education ministry. Foreign aid funds that are most effective in improving education reach beyond the centralized authority of the education ministry or department. Supported by evidence from several evaluations, for example, CfBT (2011) and AFD/DANIDA/MCPD (2012), assuming that recipient governments function like a strong, coordinated and unified team is problematic, since that is more the exception than the common experience. While poverty reduction strategy papers and other documents may suggest an orderly and coherent policy process, in practice, specifying policy, setting targets, and developing a strategy are often chaotic, spasmodic, disconnected, and not infrequently, discordant. As well, governments regularly set targets they know cannot readily be met.

Many of the evaluations we reviewed stress the importance of a political economy analysis of context, expected to improve the effectiveness of aid-funded activities (Eval: SIDA 2013; Eval: ADB-Uzbekistan 2010; Eval: World Bank 2006; Eval: Inter-American Development Bank 2011). Yet, as we have noted, most often that perspective appears as commentary or recommendation, rather than as a strong component of the evaluation of the aid and its effectiveness.

Where the supported activity is the responsibility of the education ministry, that ministry reasonably represents the government in aid discussions. Where strengthening the education ministry is deemed important, funding may be broadened from project to programme or sector support. That orientation, however, may not fit well with the emphasis on inputs and impacts. The evaluations we have reviewed note the tension between providing core support, perhaps through sectoral support or institutional strengthening programs, and a results-based agenda (Eval: SIDA 2013: 38-40).

Local ownership of education innovation: essential but rarely evaluated

The importance of local ownership has long been clear and is often highlighted in the aid literature. Evaluations have regularly noted that activities for which there is a strong sense of local ownership are much more likely to be effective, or more effective, or more inclusive, or better sustained than activities which those involved regard with some distance and perhaps with a sense that they have been delivered or imposed by outsiders. That is well known and widely agreed. That understanding appears in many aid analyses and in several important conventions intended to shape the aid process.

Yet, only rarely does aid funding focus explicit attention on developing, nurturing, and funding a strong sense of local ownership of the education activities that are supported. Similarly, many, perhaps most evaluations note the importance of local ownership. Yet, beyond organizing public events at the end of a project, few evaluations study or assess local ownership systematically and thoroughly. Some stress the importance of local ownership at the formulation stage, yet provide limited, if any guidance on accomplishing that. Others point to the role of parent and other community organizations in reinforcing local engagement and ownership but generally do not address how that local involvement does or can promote a national sense of responsibility for aid-supported education programs (Eval: Norad 2009).

It is essential to recognize the inherent and powerful tension between local ownership and funding agency interests and objectives. The issue is locus of authority. Achieving strong local engagement in and responsibility for aid-funded education activities requires that recipients have significant control over the activities and the funding. That in turn requires that the aid recipients play a prominent or the central role in setting the education development agenda. Funding agencies, however, have their own objectives and lines of responsibility and accountability and may be unwilling or unable to cede authority to the aid recipients. Evaluations regularly conclude that aid-funded programs and policies are most effective when there is clear political support from the beginning, and when local actors play a leading role in all stages of design and implementation. However, often in the same breath, evaluators note the need to define and achieve agency-specific, measurable objectives (Eval: CfBT 2011; Eval: IDB 2010; Eval: Sida 2007a; Eval: World Bank 2006). This tension is especially evident in sector-wide approaches to aid-funded education support (SWAPs). Evaluations highlight the need to involve teachers and other local change agents in the SWAP preparation and design process. However, efforts to do so are often hindered by funding agencies' conflicting agendas and differing levels of risk aversion (for example, different interpretations of local government's ability to manage SWAP design and implementation) (Eval: CfBT 2011). Since this tension is structural, it cannot be wished away, but rather must be recognized and managed.

Those evaluations that do address local ownership as a major concern go down one of several paths: (1) They report that there is little sense or a weak sense of local ownership. (2) They equate local ownership to consistency with formal documents (that is, they report that aid documents are consistent with published national education policy) or official government assent and then confirm that there is significant local ownership without much attention to what that is or how it works—very much like ticking in the local ownership box on a standard form. (3) They interpret positive assessments by people surveyed or interviewed as confirmation of local ownership. (4) Having noted its importance, they do not incorporate local ownership in the evaluation. (5) A few evaluations study local ownership systematically and relate it to the objectives of the aid-funded activities.

Note that funding agencies and evaluations are not consistent in their reference to "local ownership." Several evaluations mention

the need for local ownership in the design stage, referring to the involvement of central government officials. Local ownership is thus contrasted with foreign ownership: national government participation, not local officials, community members, families, teachers, students (Eval: Asian Development Bank 2010; Eval: Belgian Development Cooperation 2007; Eval: Norad 2009)

What approaches to increasing local ownership have proved productive?

Bottom-up approaches introduced by funding agencies include (1) beneficiary consultation and participatory planning, (2) community development support, (3) engagement of nongovernment organizations, (4) local government involvement, and (5) private sector participation (Eval: Asian Development Bank 2004).

Some evaluations focus on communication and dissemination. For example, "Community events and close interaction with civil society help disseminate information about educational interventions and evaluations" (Eval: USAID 2006). Yet, that evaluation tells us little about what "close interaction with civil society" means, or how community events are conceptualized.

Some evaluations point to the importance of political support or political will, which can be understood as another form of local (national government) ownership (Eval: IADB 2013; Eval: World Bank 2006). Not surprisingly, evaluations generally do not assess the political environment or political will. Nor have we seen terms of reference that require them to do so.

Some evaluations link attention to local ownership to the assessment of sustainability. In the absence of attention to local ownership and aid as partnership, aid-funded activities are less likely to be locally integrated or sustained once the external support has been used.

The largest number of evaluations that address local ownership do so through the lens of participation. The expectation is that increased national and local participation in the funded activities will correspondingly increase the sense of engagement and ownership. From that perspective, evaluations note differing understandings of participation and strategies to increase participation. With few exceptions, they find that aid programs characterized as participatory in practice often provide for limited, or very constrained participation

by aid recipients. Only a few seek to involve educators and communities in project design and throughout the project cycle.

Funding and technical assistance agencies themselves have different interpretations of what is participatory and what constitutes effective work at the local level. Some understand participation as working with the education ministry, while others note the importance of working with local aid recipients.

The AFD/DANIDA/MCPD 2012 evaluation links ownership, participation, and effective implementation, insisting that inclusion (decision makers, parents, students, and teachers) is a basic requisite for well-performing education systems. Yet, notwithstanding the recognized importance of involving aid recipients at the formulation stage to ensure ownership and better implementation, evaluations have much less to say about how to include them, how to assess the integration of local actors, or why that early participation has not occurred. One exception is an imaginative joint evaluation of support to girls' education in Bénin (Eval: USAID & World Learning 2006). Undertaken by a national funding agency and an NGO, using a mixed methods approach, the evaluation explored whether or not support to community organizations could increase girls' access and success. Support to local NGOs, the evaluation found, facilitated local participation at all phases of the project and thereby increased local buy-in and ownership of the girls' education strategies. At the same time, the project's period was too short and its indicators too limited to confirm confidently its long-term and sustainable benefits. As well, as we have noted, evaluators frequently report that aid programs characterized as participatory in practice often provide for limited, or very constrained participation by aid recipients.

Limited or constrained participation by aid recipients is clear in the evaluation of Swedish support to community-based school management in Zanzibar (Eval: Sida 2007b). Here, the evaluation concluded that "achieving local ownership and improved local management requires involving local partners from the outset of an activity, including the formulation stage." In this project parents and community leaders were actively involved in school construction. However, once the schools were completed (roofs were built), the sense was that the community had done its part and that responsibility should be returned to the education ministry. The evaluators suggest that communities should have been more involved in school

management and decision-making from the onset, but the evaluation does not explicitly assess the aid-funded project's attempts to do so.

Several evaluations emphasize the participation of parent organizations in developing local ownership (Eval: GIZ 2005; Eval: USAID Guinea 2006; Eval: USAID Benin 2005).They also report that a significant consequence of community participation has been the promotion of greater transparency and improved governance (Eval: USAID Guinea 2006; Eval: USAID Benin 2005). These evaluations suggest that democratic principles are taking root in the practices of parent associations and are generating a ripple effect in the political life of the communities.

An evaluation of aid funding in Chad highlights the benefits of creating parent associations, particularly for girls' education (Eval: GIZ 2005). Training sessions for parents increased participation and improved outcomes. An evaluation of an education initiative in Sudan found that parent and teachers associations can potentially address financial resource gaps by conducting their own fundraising activities targeting the broader community, thereby reducing the financial burden on learners and their families (Eval: GIZ, Sudan 2014). Yet, analytic comments in the GIZ evaluations indicate that PTA members generally lack needed skills: organisation, management, project planning and implementation as well as fundraising.

An evaluation of USAID support in Benin found that strengthening the capacity of grassroots organizations can help increase enrolment (Eval: USAID Benin 2005). The evaluation suggests that communities should be encouraged to participate in the co-management of schools.

Evaluations have also found that approaches to increasing the roles of local organizations differ, depending on whether the starting point is formative dialogue with the community or the study of the evolution of civil society in the particular national or local context. These two approaches are complementary. It may be futile to enter into a dialogue without an adequate understanding of the context. On the other hand, only through community consultations and interpretations does the context become clear.

Yet, evaluations have typically not examined the roles of local organizations and are not consistent in assessing those roles. While some are excitedly enthusiastic, others are far more reserved. That inattention is puzzling, both because many aid-supported activities are implemented by local organizations and because the expected benefits

of decentralization can be achieved only where there is a strong and active network of local organizations. Even so, most evaluations say little about the organizations with which funding agencies work (Eval: DFID 2010). Most often, national aid agencies rely on organizations within their own countries to maintain relationships with recipient country organizations. A review of several evaluations of Norwegian voluntary organisations concluded that short term objectives are often achieved, but that little is known about whether they achieve their intended long term objectives (Bye 2000). Even when there has been training and an emphasis at capacity development at the local level, "the assumption that these new skills will be applied and that organizations will welcome new ways of working is unwarranted" (Eval: Sida 2013).

Where funding agencies pay too little attention to the education system's institutional arrangements they may over-estimate the role of local organizations. The evaluation of USAID support to Bénin indicates that "the vision of a centralized school system clashes with one of the school as a responsibility of local government" (Eval: USAID Benin 2005: 2). An evaluation of AFD/DANIDA support to education in Benin found the central role of the state in education provision more consequential than the roles of community organizations. A World Bank evaluation (2006) notes that projects designed to provide technical assistance to central governments rest on a weak institutional-political analysis base. Even more problematic, community management has been linked to improved facilities and staffing but not to improved instructional quality or learning (Eval: World Bank, 2006).

One possible explanation for the disconnect between the regularly reiterated importance of local ownership and the absence of focused strategies for achieving and evaluating it is the lack of consensus on what, exactly, is needed. Our interviews revealed disagreement among aid agencies in their understanding of local ownership. Are the important owners local or national government institutions, organizations (as suggested by AFD), communities, teachers, learners and their families? The evaluations we have reviewed provide limited information on how funding agencies work with civil society networks, rather than particular organizations. Where local ownership means little more than limited advance consultation with selected officials and discussion of findings at a project's conclusion, the reports of limited local ownership will continue. Frustration will

be more common than genuinely shared responsibility and development partnership.

Just as funding agencies differ on how they understand local participation, so too do evaluators, an issue to which we will return. Should participation in evaluation be understood to mean working with a few local people to distribute surveys and perhaps gather and review basic data, or does participation require including local people in the conceptualization of the evaluation or surveys? Even as they confirm the importance of local ownership and review strategies for promoting local involvement, evaluators have not explored systematically and thoroughly why aid funding apparently does not strengthen either local ownership or aid as partnership. With rare exceptions, evaluations do not use the recognition of the importance of ownership and participation to explore or assess the consequences of prescribed and asserted local roles in aid-funded activities.

Outcome measures that permit characterizing a project as successful generally do not capture either local ownership or aid as partnership and thus are at best limited indicators of the achievement of intended objectives and at worst may well obscure what should be obvious—that without local advocates, defenders, and teachers, students, and communities committed to the activity, when the funds are exhausted, the activity will cease.

Reaching the difficult to reach remains beyond reach

The evaluations we have reviewed confirm the challenges of extending education opportunities to the most difficult to reach populations.

Funding agencies periodically affirm their commitment to inclusive education and to bringing into schools those on the margins of the education system, for example, children in remote rural areas and children of transhumant groups, as well as children who do not see or hear well. Evaluations reflect that commitment, noting the importance of improving efforts to target the most vulnerable (Eval: UNICEF 2013; Eval: Sida 2007a; Eval: IDB 2010; Eval: 3ie 2013).

Many evaluations, however, report that the most difficult to reach populations remain largely excluded from aid-funded education projects. An evaluation of USAID support to Bénin found little progress in integrating children with special needs into the education system (Eval: USAID Benin 2006). As well, despite commitments to education equity, the distribution of aid to education can and

periodically does lead to greater socioeconomic disparities (Eval: French Ministry of Foreign Affairs, 2007). Most public resources benefit a small minority: 39% of resources go to the 10% most educated (Eval: French Ministry of Foreign Affairs, 2007).

Evaluators often follow their reports on exclusion and uneven benefits by suggesting that new strategies are needed to reach the communities least well served. Those are generally supplementary observations and are certainly well known in the international education community. Those comments, however, are just that, informed observations rather than systematic analysis. Few evaluations explore carefully who are the most difficult to reach learners, what obstacles they encounter, and what can be done to improve their learning opportunities.

There are some exceptions, especially evaluations of specific programs managed by smaller organizations that are directed exclusively at communities that have been identified, either through direct consultation or survey analysis, as the most disadvantaged in a particular context (such as education programs for refugees). Still, where they are concerned with unequal access, evaluations of sector-wide programs and large scale ministerial-led projects for the most part focus on comparing education outcomes across factors that are known to be related to education inequality, such as gender and socioeconomic status. While useful for addressing equity, that analysis across binary categories (male/female, rich/poor, urban/rural) does not contribute directly to exploring how to reach the hardest to reach.

Aid funding intended to reduce inequality may in practice relocate it. Some evaluations focus on aid support intended to address inequalities in access and school progress. Girls' education is the outstanding example. On that, evaluations report great progress, usually associated with the particular approach or strategy that was evaluated. On that, the assumption seems to be that doing more of whatever seems to be working will eventually achieve substantive equality. With rare exceptions, evaluators do not address the shifting locus of inequality and the persistence of inequality. Once equal numbers of boys and girls enter primary school, the differentiation is not access but attrition, or selection for secondary school, or subject specialization. Nor do evaluations charged with examining a particular initiative explore more generally the gendered nature of power and authority and its consequences for inequalities in education.

As well, it is striking that other systematic patterns of advantage and disadvantage, for example, different education experiences for Muslim and Christian populations, receive far less attention, both from funding agencies and from evaluators.

Centralization despite decentralization

Earlier, the World Bank and other funding agencies regarded decentralization—transfer of authority and responsibility from central to local levels—as an essential component of education reform. In many countries, however, the most common practice in the education sector has been deconcentration—relocation of some officials and roles from central to provincial or local education ministry offices, without a significant transfer of power and authority to local communities. Apprehensive about losing control, central authorities have been willing to delegate authority for some administrative and discipline issues and have found it useful to direct complaints and challenges to local education offices. Only rarely have local education authorities been granted authority to develop curriculum, assess achievement, employ and transfer personnel, and raise revenue. Even where the decentralization of education authority has a strong legal foundation, for example South Africa, a powerful political alliance supports the reassertion of central authority.

Support for decentralization among funding agencies is now more muted, at least in part a recognition that decentralizing education authority nurtures education reform in some settings but not in others. Decentralization can entrench resistance to change and exacerbate inequalities among schools. Research has found that the appropriate balance between central direction and local autonomy is likely to vary over time and circumstances, perhaps even within the same setting (Maclure 1993; Samoff 2013: 421-422).

In their assessments of where foreign aid is most effective, many evaluations address decentralization (Eval: Sida 2005; Eval: USAID 2005; Eval: USAID/World Learning 2005; Eval: Sida 2007a; Eval: ADB 2008; Eval: AFD/DANIDA/MCPD 2012; Eval: IADB 2013; Eval: Norad 2009). The AFD-DANIDA case study in Benin (2012), the USAID-World Learning evaluation in Benin (2005), the USAID assessment of the USAID Assistance Program to the Reform of the Benin Primary Education System (2005), the Asian Development Bank Evaluation of the Education Sector (2008), the Program Evaluation for USAID-Guinea Basic Education (2006), among others, all discuss decentralization at great length though it was

not an initial object of study for the evaluations themselves, which for the most part were either sector-wide or programmatic.

Many evaluations conclude that aid to education is more effective where some resources are directed to local levels and managed directly by local authorities. At the same time, evaluators regularly note the gap between the rhetoric of decentralization and the practice of strong central authority. Several evaluations confirm that there has been little decentralization within the Ministry of Education, while at times limited deconcentration has transferred some responsibility to lower levels of administration but with limited decision-making authority (Eval: AFD/DANIDA/MCPD 2012; Eval: The Asia Foundation 2013).

What have been the major obstacles to more extensive decentralization of education authority? The most common perspective emphasizes insufficient skills and capacity at the local level and ambiguities in how decentralization is expected to be accomplished. An alternative understanding highlights instead political resistance to the transfer of authority.

Frequently, the implementation of decentralization is murky and spasmodic. Possible explanations suggested by the evaluations we have reviewed, include: (1) the design of decentralization as specified in legislation and decrees may create uncertainty as to which level of government or which decision-maker is responsible for what (Eval: Norad 2009); (2) the capacities of school boards to govern schools, or for school directors to manage schools, or for teachers to implement school reforms may be limited (Eval: World Bank 2005; Eval: DFID and IOE 2014); (3) there is no support system to the newly decentralized authorities (Eval: USAID 2005). Conversely, comparative education research suggests that aid-funded initiatives may in fact *undermine* autonomous local efforts to improve school quality. Yet, few evaluations of aid-funded activities entertain and explore this possibility (Eval: DFID and IOE 2014).

Decentralization is more difficult in some contexts than others, particularly in countries that have a history and policy of high centralization (Eval: USAID 2006).

Yet, there are clear examples of local authorities that have effectively exercised authority and mobilized community resources to support schools (Eval: Norad 2009; Eval: World Bank 2006). That suggests a critical interpretation: the major obstacle to

decentralization is not lack of local capacity but lack of political commitment.

An evaluation of support to education in Bénin found that decentralization has proceeded much further in the health, water, and sanitation sectors than in education. The ministries in charge of education are not inclined to transfer significant competencies to the commune level. When services are decentralized, there are limited resources to accompany their management; this is particularly notable in educational quality, equity, and delivery. Critical information remains centralized at the national level. (Eval: AFD/DANIDA/MCPD, 2012)

The Bénin case study illustrates that even where there is funding agency support and broad community consultation on education sector plans, decentralization has not been achieved because of lack of political will (Eval: AFD/DANIDA/MCPD 2012). A key finding from this case study is that when services are decentralized, there are limited resources to accompany their management, and this is particularly the case in educational quality, equity, and delivery. Information systems and education statistics remain highly centralized at the national level.

Some of the evaluations we have reviewed point to potentially undesirable consequences of decentralization. Decentralization can exacerbate existing inequalities between schools. In Indonesia, for example, evaluations found that the impact of aid-supported decentralization efforts has differed sharply across regions (Eval: AusAID 2010; Eval: RTI 2010). Through interviews with aid officials, teachers, and local education officials, evaluators found that the most significant element in the project's success, above and beyond local capacity, "was the level of commitment of the district or province and the capacity of the implementation team to leverage and build that commitment" (Eval: RTI 2010: 11).

Moreover, while the rhetoric about decentralization refers to community empowerment and accountability, in practice, decentralization is often a strategy for transferring financial responsibility to parents or local governments. Even where there appears to have been significant decentralization, teachers and school committees have little or no decision-making power (Eval: Norad 2009).

Let us summarize. Beyond confirming that in aid-receiving countries (and in most of the world) most people think education

requires a strong central authority and there has not been much decentralization, what else do the evaluations tell us about decentralization? First, decentralization is an important core component of official education development strategy. For instance, the IDB evaluation notes that decentralization is one of the Bank's priorities, despite the presence of mixed results on decentralization's effectiveness (Eval: IADB 2013). Second, the evaluations confirm that decentralization comes in many shapes and sizes. Third, as the Inter-American Development Bank (2013) evaluation notes, there is also substantial evidence from the development economics literature that notwithstanding its expected benefits, decentralization can exacerbate existing inequalities between schools and communities (Eval: IADB 2013).

Fourth, while the rhetoric of decentralization highlights community empowerment and local accountability, in practice, meaningful participation at the community level may be difficult to achieve and is often limited to financial contributions or school maintenance activities (Eval: Norad 2009; Eval: Sida 2007b; Eval: IDB 2013). Fifth, decentralization strategies sometimes encounter local resistance. For example, teachers may resist increased local authority, apprehensive that head teachers or communities will use their authority unfairly in evaluating teachers.

Sixth, even as many evaluations stress the importance of decentralization, few address it explicitly as part of the evaluation or explore how aid agencies might facilitate the decentralisation process. One exception is the evaluation of AFD/DANIDA support to Bénin, which reports that despite two years of significant technical institutional assistance provided by DANIDA to the Benin Ministry of Education, education remains highly centralized.

It is important to note here that while decentralization was for some countries and some funding agencies a very high priority education objective, the process of decentralizing is not readily amenable to quasi-experimental or experimental approaches to evaluation. Assessing progress on decentralization, including determining what in fact was intended, requires moving beyond rhetoric through systematic and detailed attention to complex and barely visible interactions. Without that attention to context, evaluators cannot determine intent, why there has been limited progress, which obstacles are most significant, or how those obstacles can be addressed.

Sustainability: important but not systematically evaluated

In September 2015, as we were preparing this synthesis, the United Nations formally adopted Sustainable Development Goals (SDGs). Yet, with few exceptions, in the evaluations we reviewed sustainability is discussed but not an explicit design goal of aid-funded education activities or, it seems, a major focus of aid funding or evaluations of aid-supported activities. Most often, sustainability is noted as a quick afterthought or an item on the list of desirables.

Sustainability is of course not a new concern. It became an important theme in aid programs by the 1960s. Rhetorically, at least, its importance has increased since then.

Why is sustainability so often an afterthought in evaluation? Where sustainability is not an explicit objective of an aid program, evaluators will not be expected to assess it, though they may do so on their own initiative. Chapman and his colleagues have argued that lack of attention to sustainability is a reflection of the diversity of opinions regarding what should be sustained (internet connections to schools? new management practices? teacher training programs?) (Chapman and Moore, 2010; Chapman and Quijada, 2009; Nkansa and Chapman, 2006). Likewise, discussions of sustainability generally ignore the highly political character of development in general and education in particular (Chapman and Moore, 2010; Pritchett, 2009). The likelihood that a program continues to receive funding depends on much more than the evaluation results. Sustainability requires that "the right people know that the project was successful" (Chapman and Moore, 2010). Technically flawed programs often yield political payoffs that make their continued funding attractive to governments (and/or to aid agencies). Likewise, many technically successful programs remain unfunded (or under-funded) because they are insufficiently visible at the national level or unpalatable politically.

As well, as we have noted, the focus on demonstrable short-term impacts may in practice undermine the long-term impact of the activities they fund on the education systems they support and may weaken the very institutions the aid seeks to strengthen.

How might sustainability feature more prominently in the aid relationship, both in aid-supported activities and in evaluations?

Long-term institutional co-operation might increase the priority assigned to the sustainability of individual projects or other aid-funded activities. In the current environment, however, the

preference for out-sourcing—often education support as well as evaluations—reduces the direct connection between the funders and recipients and may reduce the education professional expertise of the funding agencies.

The aid model organized around funding pilot experiments and then reproducing those deemed effective seems to have an inherent commitment to sustainability. Yet, as Riddell points out, sustainable education outcomes will not be achieved merely by reproducing more, and more successful, but individual projects (2012). As well, attempts to reproduce successful pilots have regularly stumbled, often failing to achieve the intended expansion of scale and sometimes undermining the original pilot (Samoff, Dembélé, and Sebatane, 2011, 2012).

Several evaluations report that increased attention to knowledge transfer will increase sustainability. For example, an evaluation of Swedish support to the education sector in Mozambique asserts that the transfer of knowledge, including attention to daily routines and record keeping, is very important in development co-operation (Eval: Sida 2004). Noting obstacles to education development, this evaluation found neither an explicit strategy nor coherent planning for knowledge transfer through institutional training, capacity development, or organization support in the education sector. Note that this perspective presumes that it is the funding agencies rather than the recipient education systems that have the critical knowledge. That presumption remains to be assessed empirically.

Might more participatory approaches increase sustainability? An evaluation of UNICEF support to basic education in the Democratic Republic of Congo found that participatory approaches and the involvement of local actors facilitate sustainability (Eval: UNICEF 2012). Sustainability can be increased by government management of sector financing, incorporating collaboration with regional, provincial, and local officials and involvement of local communities (Eval: UNICEF, 2012). That local participation may require orientation and training, for example, in the constitution and operations of school committees. Where the rhetoric of community based support is not accompanied by direct community involvement the outcome is likely to be frustration and disengagement. This evaluation recommends increased involvement of grassroots

organizations and increased expertise and training for technical directors.

An evaluation of the Inter-American Development Bank's support to secondary education argues that development of national assessment systems and participation in international assessments will increase sustainability of funded projects (Eval: IDB 2013).

What, then, do the reviewed evaluations tell us about the sustainability of aid-funded education activities in poor countries? The general observation is that while funding agencies regularly reiterate their expectation that aid-funded education activities be sustainable, in practice aid programs generally do not include either explicit attention to what is required for that sustainability or funding specifically dedicated to achieving sustainability. Not surprisingly, many evaluations do not address sustainability systematically, or do so only in their supplementary comments.

Major Findings: Challenges to Evaluators and Funding Agencies

Thus far our synthesis has focused on education and education outcomes. Though they are generally not very self-reflective, the evaluations also provide insight into challenges to evaluators and to funding and technical assistance agencies. We note two themes that stand out. Recall that we explore the evaluations as a set.

Information, evidence, data, and indicators

Everyone agrees that effective education planning and management require reliable and regularly updated information. For many countries, especially where distances are great, infrastructure is not well developed, and human resources are sorely strained, collecting important information on the education system is a persisting challenge. Sometimes the international community's data demands compound the challenge, overwhelming data collection and analysis capacity. Periodically, external support focuses on improving that capacity. As well, as Jerven's and others' work illuminates, available datasets commonly have gaps and large error margins and are not readily compatible or comparable.

The need for better information management, data and indicators is a pervasive finding across the evaluations we have reviewed. Impact assessments, an increasingly common approach to

evaluation, depend on reliable and valid large scale data and focused indicators.

With important exceptions, most of the evaluations we have reviewed include a note pointing to gaps and other problems in the available education data. Yet surprisingly few of these studies address data problems directly, either by collecting their own general education data or by developing strategies for working with seriously flawed data. Nor do most evaluations integrate into their findings the very large probable margins of error in most of the available education data.

Evaluators note that where information does exist, it may not be reliable (Eval: AFD/DANIDA/MCPD, 2012). Some evaluations suggest that improved data collection and use require building local capacity in data collection (Eval: WB 2006; Eval: Sida 2007b).

UNICEF emphasizes the importance of reinforcing monitoring and evaluation capacities at the provincial and local level (Eval: UNICEF 2011). Especially important are training in monitoring and evaluation, increased coordination between UNICEF field staff and institutional actors, and improved circulation of information so all involved can develop a broad and informed view of program implementation. For that, funding and technical assistance agencies can provide technical support, including assistance in developing data repositories and electronic communications.

Collaboration with civil society organizations can improve data triangulation (Eval: European Commission 2010; Eval: UNICEF 2012). Working with local reference groups may increase access to data sources and improve information flows and use (Eval: AFD/DANIDA/MCPD 2012).

Yet, the collection and analysis of data remain highly centralized in many countries, particularly those with a history of a strong central government, (USAID 2005; Eval: AFD/DANIDA/MCPD, 2012).

Most evaluations note the lack of quantitative indicators, especially of education quality, and the overall tendency to focus on inputs rather than outputs. Few discuss why this is the case (besides blaming "low levels of educational planning capacity among national ministries of education" [Eval: BTC 2007]). The Evaluation of Belgian Aid to Education (2007) summarizes the challenges:

1. educational quality is culturally defined,

2. there is no international consensus on how to measure/define educational quality,

3. education systems are slow to respond to inputs,

4. educationalists differ on the use of testing, and

5. education results are politically sensitive.

It is important to note here an important distinction that has not caught the attention of many evaluators. The clamour is for improved evidence, deemed essential to improving education in poor countries. The exhortations to collect more and better data are frequent and persistent. So too is the critique that data collection is inadequate. The common assumption is that countries lack the capacity for effective data collection. Yet, notwithstanding the rhetoric, perhaps education policy makers and managers do not see a need for more indicators and more data collection. Overall, evaluations have yet to grapple with this distinction—lack of capacity vs. no perceived need. As a result their observations and recommendations may be profoundly misdirected.

Also generally unaddressed in the evaluations we reviewed are the trade-offs between increased efforts to collect more and more reliable education data on the one hand and on the other, efforts focused on making better use of a much smaller number of indicators. Nor do the evaluations explore how the funding agencies might proceed if they based both their support programs and their evaluations on the limited, and not infrequently partial and inconsistent, data that aid-receiving education ministries use regularly to manage education systems.

The importance of institutional knowledge and learning among funding agencies

The evaluations we have reviewed provide strong support for a familiar recommendation: the need for substantial institutional knowledge and learning among funding and technical assistance agencies. Often, however, that important theme is developed as a supplementary observation, rather than incorporated as a major focus for systematic and critical evaluation.

As defined by Berg, organizational learning is "concerned with how new knowledge is translated into operational reality" (Berg 2000: 2). Formal evaluation should be a tool of organizational learning, response, and ultimately change, providing ideas and insights drawn

from projects and programs. Yet evaluation has not performed these functions well, particularly in regard to strengthening the capacities of the funding and technical assistance agencies (Berg, 2000). In practice, the priorities of aid and allocations trump the accountability, learning, and dialogue objectives that can be achieved through effective evaluations.

On this, both research and many years of evaluation are clear. Structural, process-related, and cultural factors continue to impede efficacy in aid administration (Forss et al., 1998). As Forss et al. stress, the major challenge in improving aid effectiveness is not in acquiring or documenting knowledge, but in enabling and encouraging organizations to act on existing knowledge.

The current dominant model of knowledge-based aid advocates that development agencies (1) implement strategies for internal knowledge management and organizational learning; (2) develop partnership mechanisms for the transfer of knowledge and learning to the partner countries; and (3) support development of partner country capacity to absorb, apply and provide knowledge (Ramalingam, 2005, in Krohwinkel-Karlsson, 2007). Despite increasingly rigorous feedback systems, development agencies continue to be criticized for their inability to incorporate past experiences, for learning too little too slowly, and for learning the wrong things from the wrong sources (Krohwinkel-Karlsson, 2007). Moreover, across the aid delivery system there is limited study of organizational learning, power structures, and differing incentives in development cooperation (Krohwinkel-Karlsson, 2007). The challenge for practitioners is to expand the view of learning from an internal perspective to a systemic perspective.
The evaluations we have reviewed address institutional learning from several perspectives. The preceding section addressed the first, data collection and analysis.

A second perspective focuses on information sharing and networks. To our surprise, in the evaluations we reviewed we did not find significant analysis of knowledge sharing among networks or inter-organizational partnerships.

Institutional analysis is particularly useful as aid is increasingly funnelled through networks, providing opportunities for funding agencies to learn from and ideally reinforce each other. To be useful, impact assessment requires attention to complex processes of change, including institutional learning (Norad, 2009).

A third perspective emphasises explicit attention to the organizations involved in aid-funded education activities, focused on both aid recipients and their environments and on the funding and technical assistance agencies. That may be especially important where NGOs play a prominent role in receiving funding and managing education activities, yet perhaps impossible to capture in an impact assessment.

A Norad report assessing support through and to umbrella and network organisations found that non-governmental organizations are expected to contribute to change processes with broad social objectives, not only education, but also poverty alleviation, democratisation and protection of human rights (Norad 2004). Their roles include service delivery, advocacy, and social mobilization. Accordingly, assessing their role in supporting education activities requires a systematic and critical understanding of their contributions to institutional learning, with attention to formal policies and informal practices that delimit their role and to the contexts of particular activities.

A fourth perspective asserts the benefits of collaborative evaluations. The Bénin case study points to the utility of joint evaluation in generating well informed institutional analysis (AFD/DANIDA/MCPD, 2012). The two Northern-based funding agencies reported organizational learning as well as positive exchanges with the third evaluation partner, a research institute in Bénin.

Effective institutional analysis also requires funding and technical assistance agencies to be more self-reflective and more self-critical. Most often, evaluations fail to address that systematically. Frequently, evaluations present a positive characterization of the funding agency that has commissioned the evaluation. Regularly evaluations review the agency positively for supporting projects that are considered relevant (they are deemed to address directly the development needs of the country) and for achieving effective results (still with limited attention to whether or not these results are actually attributable to aid). Criticism in these evaluations tends to highlight deficiencies among recipient country governments, typically in the form of "lack of capacity for monitoring and evaluation," or "limited experience in quantitative data analysis among government functionaries" (Eval: Asian Development Bank 2010). Similarly, the IDB evaluation of aid to secondary education, funded and conducted in house, focused mainly on the challenges facing the region/governments, rather than the role of the aid agency itself (Eval: IDB 2013).

4. Education, aid, and evaluations

Having reviewed several major findings on education issues and on challenges to evaluators and funding agencies, we turn now to what we can learn from the set of evaluations about the aid relationship and about evaluations and the evaluation process.

The Aid Relationship

Our central concern here is what can we learn from evaluations of aid-supported education activities. Our focus is thus at the intersection of education, foreign aid, and evaluation. While a detailed analysis of foreign aid, or even of foreign aid to education, is far beyond the scope of this synthesis, it is essential to explore briefly how the fact of aid and the aid relationship themselves shape education outcomes.

For the purposes of this discussion, we understand aid as the provision of resources in several forms, including technical assistance and advice, to education systems in low income countries. Since we are concerned with the aid relationship and its consequences, we do not seek to measure precisely the magnitude of foreign aid. Nor do we address here the evidence that in at least some circumstances, foreign aid may both facilitate and mask a net outflow of resources from less to more affluent countries.

Research and commentary on foreign aid, both in general and specific to particular countries and organizations, are extensive and readily accessible.

The starting point here is the recognition that the foreign aid provided by countries is first and foremost a foreign policy tool to promote those countries' national interests. The better known and less well known funding agencies are all responsible to their governments, regularly reporting on their activities and justifying their disbursements. In itself, that is neither undesirable nor problematic. Citizens expect their governments to promote and advance their interests. Foreign aid is one means for doing that. Potentially problematic, however, is losing sight of that purpose.

Foreign aid as we see it today is a relatively recent arrangement. The creation of the League of Nations Mandate system, succeeded by United Nations Trusteeship, institutionalized the idea that higher income countries had a formal responsibility to provide

development assistance to those countries, mostly former colonies of defeated countries. Foreign aid took on new force in the period following World War II, especially in the tension and competition between the Soviet Union and the United States. Demands for self-determination and decolonization required reconstructing the relationships between metropolitan countries and their former colonies, often including the provision of some form of on-going support, generally in exchange for continuing preferential trade and other links. Education became a prime focus for foreign aid.

Most foreign aid is provided by countries, either directly (national resources allocated to recipient governments) or indirectly (for example, as contributions to UNESCO or UNICEF, or in funds managed by an international organization). Philanthropic foundations have also been active, generally with a more narrowly defined focus than national foreign aid and with delimited private rather than public accountability.

In what ways have foreign aid and its provision shaped education outcomes?

From Support for Education Innovation to Aid Dependence

For many years, external support to education in low income countries was focused on specific projects intended to expand and improve education. Formally, aid was to support development expenditures (the capital budget), not the on-going costs of the education system (the recurrent budget). In that role, foreign aid was a very small part of total spending on education, perhaps 1-3%. Though its volume was limited, that aid had tremendous leverage. Where national governments struggle to pay teachers, produce textbooks, and supply pencils, innovation and reform seem beyond reach. Foreign aid could close that gap. Where new initiatives were deemed possible only with external funds, even very limited aid carried powerful force. National education officials increasingly framed their agendas to fit into foreign funders' priorities.

Most recently, especially in the world's poorest countries, that situation has changed. Both directly and indirectly through national budget support, foreign aid agencies are doing what previously they said they would not do: supporting the recurrent budget. Since the wage bill is the major portion of total education spending, in some countries, effectively the aid providers are paying the teachers. While that arrangement seems unsustainable, to date there has been little

discussion of a strategy for shifting to self-reliant education spending. Indeed, the education for all campaign has presumed substantial and increased provision of education aid.

Notwithstanding periodic promises of increased education assistance and notwithstanding the professional rewards to operations officers for dispersing funding, the most recent trend has been in the opposite direction. Globally, aid to basic education has stagnated or declined. That has not, however, reduced its influence.

Initially largely managed through agreements between aid-providing and aid-receiving countries, the implementation of external support to education now has an international character. Complaints about and frustrations with the aid process have led to successive international agreements intended to specify codes of aid conduct, to transfer some control from providers to recipients, to promote coordination among providers, and to standardize and accelerate the flow of aid. The dominant terminology has shifted from charity to partnership. While some aid-receiving countries have improved their ability to secure and direct external support, for nearly all, the dominating influence of the aid providers has become more solidly entrenched. Partnership is the rhetoric. Dependence is the practice.

The most visible form of that influence are the conditions attached to foreign aid. Not infrequently, even where the aid is directed to education programs, the accompanying conditions specify changes in macroeconomic policy and exchange regulations. Another powerful form of that influence is direct participation in making education policy. Several decades ago, the participants in national deliberations to propose, review, and adopt education policy were educators and education ministry officials. Today, both the foreign ministry, which administers foreign aid, and representatives of the funding agencies sit at the national policy table. As well, even as the funding and technical assistance agencies affirm that their decisions are guided by national education policy, they specify the form in which that policy must be drafted, the indicators deemed essential to assess progress, and even how the relevant data are to be collected and analysed. Though the term partnership has political value for both providers and recipients, there is little evidence of the mutually beneficial exchange that the notion of education partnership suggests.

What makes this schematic overview relevant to this synthesis is that evaluations of aid-funded education activities are situated squarely in the aid relationship.

Mismatched Time Horizons

Foreign aid has a clear cycle and time horizon. Funding agency operations staff work with aid recipients to develop support programs, earlier mostly projects, now including some sector and budget support. With appropriate foundational analysis, those programs are incorporated in annual budgets and approved by governments. Procurement and disbursement have their own pace, sometimes far slower than initially anticipated. Since most appropriations are annual, aid-providing governments find it difficult, or are legally unable, to assure long-term support.

Education initiatives, however, generally have time horizons that extend beyond one year, or even the three-to-five year cycle that some funding agencies can manage. Improved pedagogies, for example, may take years to develop, then time to implement, then further time to refine. Improved teacher education requires experimentation and practice to become improved teaching, which then requires more time to become visible as enhanced learning.

From the perspective of the longer time horizon of education systems, especially problematic is the relatively short job cycle of funding agency officials. As the aid literature regularly notes, funding agency staff are rewarded for the projects they oversee, especially for the volume of assistance they manage. Only rarely are those officials evaluated in terms of the success of those projects. For longer time-horizon education projects, it is common for the funding agency official who oversaw funding the project's creation to have moved to a new post before the project reaches its completion. Her successor, to be evaluated in terms of the projects she manages, has limited incentive to devote major energy to her predecessor's projects, or even to know much about them.

As well, a major consequence of the push toward out-sourcing and privatization is the transformation of the role of the funding agency's field staff, who are more likely to be contract managers than education experts and advisers. The aid and education horizons are thus sharply mismatched.

That mismatch has powerful consequences for evaluation. The short aid cycle requires near-term evaluations, often well before the intended outcomes can become clearly visible. Not surprisingly, evaluations are often correspondingly superficial, attentive to what can be measured quickly (how many teachers participated in the

workshop? were the books delivered?) rather than whether or not teaching and learning improved. Where evaluations do address longer-term consequences, they may be presented to officials not involved in the activity's creation, who may have very different interests and priorities.

Attribution Challenges

We noted earlier what is commonly termed the attribution problem. Only rarely do education initiatives and reforms yield instant benefits. When positive outcomes can later be measured, it is difficult to determine what were the major causes.

A simple example makes the point. A funding agency sponsors teachers resource centres, where teachers in resource-limited environments get together periodically to assist each other. Sharing experience, they learn to paint maps on walls, or to use rain puddles for science experiments, or to integrate debating and poetry into language instruction. A promising and cost-effective initiative. If those exchanges work well, those involved improve their teaching strategies. If that works well, learning is enriched. And if that works well, eventually that increased learning will be reflected in measures of student achievement. Evaluators could then use those measures to compare the results of students in classrooms where the teachers participated in the resource centres with the results of students in other classrooms. Problematic, however, is the time lapse between the participation in the innovation and the measured result. By the time students take achievement examinations, many factors will influence their results, in addition to the ordinary confounding conditions. It is simply not possible to control for all plausible alternative causal explanations, nor to assume that possible causal factors are randomly distributed across students, schools, teachers, and communities. Direct attribution cannot be confidently confirmed.

Yet, most often the funding agencies seek that confirmation, even when they participate in budget support which combines the aid of several agencies. In private discussion, funding agency officials agree that it is not possible, and may not be desirable, to establish attribution. They also explain that they must be able to report to parent agencies of governments what *their* funds have accomplished. The agencies' logos and markers are ubiquitous, reminding all involved of the source of the funds. It is common to speak of Japanese or Danish schools, meaning Zambian or Eritrean or Mozambican schools built with Japanese or Danish foreign assistance.

In part, funding agencies insist on confirming attribution for these national reasons. In part, that attribution is central to the effort to determine what works. How should an aid agency decide among competing claims for its support? Well, it should allocate resources to education initiatives that work best. How to know that? Use evaluations to determine what works. But doing so requires confirming attribution.

Thus a conundrum for evaluators. *Establishing attribution is simultaneously necessary, problematic, and perhaps impossible*. The aid system creates strong incentives for proceeding as if it were possible to establish clear attribution and then to report that on the basis of available evidence, attribution has been confirmed.

Evaluating Swedish Aid to Education

As we have noted, our focus for this synthesis is the set of evaluations of support to education activities. While we refer to particular evaluations throughout our report, we have not sought to assess the funding, technical assistance, or evaluations of individual agencies. However, since this synthesis is intended to assist Sweden in reflecting on recent Swedish development assistance and in shaping development assistance policy and practice in the future, it is useful to comment briefly here on Swedish aid to education. Since a systematic review of the content, forms, and modalities of Swedish foreign aid was beyond our mandate, we rely here on information available in the evaluations reviewed and on informal discussions with colleagues currently or formerly involved in Swedish development assistance.

Sweden has developed and periodically revised its policies and guidelines for its foreign aid. The Swedish Aid Policy Framework (2013) and the Swedish Policy for Global Development (2002) emphasize a holistic approach to education, focusing on improved access to quality education, particularly among girls and children living in conflict or post-conflict societies. Swedish education support is channelled directly and through other agencies, both Swedish (the Swedish Trade Union Confederation and the Swedish Church) and international (UN agencies and multi-lateral initiatives, such as the Global Partnership for Education). Sweden has as well explicitly addressed its strategy for evaluating development assistance, most recently reflected in summary reviews of decentralised evaluations (Eval: Sida 2013b; Eval: Sida 2014b).

The Swedish International Development Cooperation Agency (Sida) is recognized internationally as one of the primary supporters of education development in low-income countries over an extended period. In 2006, the Global Partnership for Education ranked Sweden highest among funding agencies in education cooperation, primarily due to the Sida's focus on bottom-up approaches to development, leveraging local systems and contextualized points of departure (Eval: Sida 2007a).

Sida has a history of close collaboration with aid recipient governments, marked by an approach to development that emphasizes the strong role of recipient countries and organizations (Eval: Sida 2004; Eval: Sida 2007a). This approach has been challenged, however, by the transition to sector-wide support and donor coordination that began in the early 2000s (Eval: Sida 2004; Eval: Sida 2007a). In some cases, this shift has caused Sida to lose the close contacts and relationships of trust it once had with government partners (in Mozambique, for example) (Eval: Sida 2004). The evaluations reviewed in this synthesis suggest that the prevailing perception among Sida officials is that funding agency coordination, for example through sectoral approaches, is a time consuming but necessary process. Regularly, Sida-commissioned evaluations emphasize that Sweden can play an influential role in ensuring that country-level dialogue between funders and recipient governments remains focused on education issues—supporting the conditions that enable and sustain effective teaching and learning (Eval: Sida 2007a).

While generally positive and not entirely independent, since Sida's contracted evaluation agency is in part evaluating its own work, the summary reviews of Sida's evaluations noted above point to gaps between articulated objectives and observed practices. The 2013 report concludes that a lack of awareness of Sida's conceptual framework, among both partners and Sida itself, coupled with weak outcome monitoring, have made it difficult to judge results and learn how to improve performance. Concerned with mixed outcomes from results-based management, the 2014 report pays particular attention to the use of theories of change, the focus on poverty, and the efficiency of Sida supported projects and programmes. Notwithstanding Sida's overarching stated objective, to create preconditions for better living conditions for people living in poverty and under oppression, the summary review indicates that contexts and causes of poverty are often not well analysed in evaluations. Overall, very few evaluations reviewed systematically address causal mechanisms—both causes of

poverty and how external assistance can reduce poverty—and poverty reduction is often missing from indicators, outputs, and outcomes.

While evaluations of Swedish aid report that major outputs have been achieved and capacities developed, that may not be sufficient to achieve changes in attitudes, norms and practices. The evaluation synthesis found that the overall perspectives of the poor are very rarely highlighted or described, and that programs and evaluations are in practice very top-down. Few evaluations explicitly develop measures of impacts on the well-being of the poor. In "Reality Check: Bangladesh" the evaluation reports that information and explanations often do not reach people living in poverty and that the quantitative target bias in current Sida practices may reinforce rather than reduce discrimination. Notwithstanding the stated commitment to include stakeholders at all levels, the perspective of aid recipients in evaluations is minimal. While there has been success in addressing gender equality, a Sida high priority objective, results are very uneven. The synthesis commissioned by Sida criticizes the evaluations, characterizing them as uneven in quality, with several deemed to have insufficient evidence and analysis. Among the major observations: many of the evaluations reviewed failed to look critically at basic assumptions and the broader political context, which is particularly problematic in a context of state fragility.

In these respects Sida also faces the challenge discussed above common to many aid agencies: improving institutional learning and effective evaluative partnerships, as well as mechanisms to learn from experiences, and from evaluations in particular (Eval: Sida 2013; Eval: Sida 2014).

Recent evaluations and synthesis reviews of Sida evaluations highlight the tension between:

(a) Sida's commitment to local approaches to development and encouragement of political will and capacity development among aid recipient governments, and

(b) the international emphasis on results-based management of aid to education, and the resulting investments in monitoring and evaluation systems that are designed and developed according to funding agencies' standards, rather than local needs and capacities (see Eval: Sida 2007b, for example).

Our review does not enable us to assess the consequences of major institutional changes within Sida related to education. Earlier,

the professional education staff of Sida's education department, now integrated with other units, was several times the size of the education staff of its Nordic counterpart agencies. Currently, as we understand it, the professional staff focused on education is far smaller. Earlier, Sida's evaluation department was widely recognized as a pace-setter among its peers. Currently, as we understand it, Sida engages external firms for nearly all its evaluation work, including education. We note those changes because whatever their benefits, they may pose challenges for Sida's capacity for institutional learning.

In sum, while Sweden's education aid has a long and proud history, and while Sida's development assistance framework and approach are exemplary in their insistence on responsiveness to and involvement of aid recipients, recent evaluations, both broader and more narrowly focused, have found important gaps between stated objectives and observed practice.

Evaluations: For What? For Whom?

We turn now to the evaluations and the evaluation process. Evaluations themselves are rarely self-reflective or self-critical.

Déjà Vu All Over Again

For evaluations to be useful, they must be read, reviewed, digested, and their findings incorporated in policy and programs. Yet, sometimes evaluations disappear into a bottomless pit. Or perhaps a black hole, though without its intense energy. Evaluations of the provision of computers to address the shortage of skilled and experienced teachers provide clear examples. It is productive to follow that path from conception through implementation to evaluation. Though the case study details are lengthy, it is those details in this and the following section that are powerfully instructive as they illuminate the importance of context and complexity.

The starting point has been clear and consistent over many years, beginning well before small computers became so common. The development of education systems in poor countries is impeded by a persisting shortage of skilled teachers, either in general or in particular subjects, especially mathematics and science. How can that shortage be addressed, while teacher education is expanded? One proposed remedy has proved particularly attractive, indeed seductive. Use technology to enable a few very competent and experienced teachers

to reach a large number of learners. New technology can of course have other roles, but it is instructive to focus on this one.

The particular technology to be employed has changed over time. In the 1960s radio lessons were expected to reduce dramatically and quickly the number of adults unable to read and write comfortably. Many people had radios, even in remote rural areas. Listening groups could be organized to hear the broadcasts and work on assigned exercises. The local literacy tutors need not be very be very skilled or experienced, since their responsibilities were to convene and organize the groups, manage the radio, distribute and review exercises, and lead the follow up discussions.

The strategy was clear. Concentrate expertise at a central distribution point, in this case, a broadcast studio, usually in the capital. Deliver that expertise using available technology. Recruit less skilled and lower paid staff at the distant end to manage reception and follow up.

During the 1970s the focus shifted to television, with a major initiative in West Africa. Foreign aid supported selecting and training the experts, developing the infrastructure (broadcast facilities, power sources), and acquiring the hardware (television monitors, batteries, and charging stations). A decade later the focus shifted to computers, initially to be installed in school-based clusters. As prices declined, classrooms were to have computers. Most recently has emerged the prospect of one laptop per child.

Over several decades, the evaluations of this general approach have been consistent. With occasional exceptions, the major objectives have not been met. Radio lessons did not eliminate, or perhaps even reduce, illiteracy. Televised instruction did not improve achievement outcomes or perhaps even expand access. Nor have computers transformed education in the ways anticipated. Similar thinking in more affluent countries has followed a similar path.

Systematic evaluations commonly report initial enthusiasm and achievement of basic delivery and training objectives and then note the problems.

The problems are several. A full review is beyond the scope of this brief comment. Regularly, the technology proves to be more fragile and less reliable than anticipated. Radio batteries die at a critical moment. Electric power is unreliable and so uneven that it damages the hardware. There is little or no funding for maintenance and

replacement. Especially in the computer era, where there is no dedicated funding to replace software and hardware, either one or the other becomes a major impediment: the new hardware requires new software, or the new software will not run on the old hardware. If all the problems were technical, there might be some prospect that over a longer time they could be resolved. However, the problems are not primarily technical. More important, the premise is flawed. Technology, whether radios, televisions, computers, or duplicating machines, or books, is always a support for teachers and learners, not a substitute for competent teachers. Like books, computers can assist teachers in doing new things and in doing what they do better, but they cannot replace teachers or the interactive character of learning.

Each round of technology had and continues to have specific uses that are effective. But the general strategy of using technology to substitute for teachers and face to face instruction has proved frustratingly ineffective over several decades. Generations of evaluations report that explicitly.

Yet there seems to be little learning from experience.

Our primary concern here is not the role of technology in education. That will surely be researched and debated for many years to come. Nor is our primary concern here the use of computers or other technology to replace teachers or extend their tools. Rather, our focus is on learning from experience and on the roles of evaluations.

Among the evaluations we reviewed are several that assess the provision and use of information and communications technology. One example is Swedish support to the use of information and communications technology in teacher education in Tanzania (Eval: Sida 2014a). The evaluation was extensive, systematic, and detailed, including questionnaires and site visits. Evaluators found that the major objectives of the support had been achieved. Their supplementary observations, however, raised many of the concerns noted above. There was insufficient funding for maintenance and for training staff responsible for maintenance and upkeep. Some of the provided computers had failed, increasing the demand on the others. There was no funding for hardware replacement, either for the computers that failed or to retire those that reached the end of their productive lifespan. Though apparently well used for their specific tasks, since the computers were not well integrated into the programs and courses in the secondary schools to which the teachers were assigned, the teachers' own learning was more about computers and

using them than about incorporating computers in their instructional activities. Some teachers were assigned to schools that have no electricity. While the schools and education ministry were pleased to receive the computers, their budget did not provide for maintenance or replacement.

Déjà vu. Or déjà vu all over again.

The evaluators' observations might well have been written two decades ago (Grant Lewis and Samoff, 1992). Striking and puzzling. Both the projects and the evaluations reflect limited learning. First, the same projects are repeated and then repeated again, notwithstanding the persuasive weight of many years of evaluations that highlight the problems of the approach. Second, while the current evaluations are clear on those problems, they neither report on the earlier history not signal to the funding agencies that they have ignored their own experiences and earlier evaluations.

What do we learn here? First, the expected cumulation of knowledge and institutional learning often do not occur. Evaluations and well grounded knowledge prove less important in shaping funding agency behaviour than other influences that favour particular projects and allocations, notwithstanding the evidence of problems. Second, regularly both funding agency staff and evaluators pay little attention to relevant history, including systematic, detailed, and critical evaluations, and apparently have little incentive to do so.

Evaluations, notwithstanding good intent and hard work, disappear into a bottomless pit.

Ignoring Context and Complexity

"Evaluations are essential. We must learn from experience. Evaluations tell us what we have done well and what needs to be done better. The evaluation of that project told us what was successful and what were the problems. We used that information in the follow-up project. We learned how to do it better."

Asked about evaluations and their role, a senior education ministry official was enthusiastic and emphatic. They are important, she insisted, and we use them regularly. Her example was concrete. Problematic, however, was the timing. The evaluation that she said was important in developing the follow up project was completed three years after the follow up project began. While the follow up project may well have addressed problems in the earlier initiative, the

evaluation of that initiative simply could not have played the role she outlined.

Her director was equally enthusiastic.

"*I recall learning a lot from the evaluation, especially concerning the number of the female tutors in the teacher education program. There were too few female tutors. Perhaps some people thought that women could not become competent teachers about computers.*"

That observation too was problematic. In their report, the evaluators noted that the original project had no gender component. They go on to explain that since the original project had no gender component, they did not evaluate its gender dimensions. Hence, while there may well have been few female tutors, neither the Director nor anyone else in the education ministry could have learned that from the evaluation.

These comments, drawn from an effort to trace an evaluation (2014) from the aid providers to the aid recipients (2015), offer striking insights into the evaluation process. Everyone is clear that aid-funded activities must have formal evaluations, generally conducted by outsiders. Everyone can articulate the rationale: we learn from the observations of the evaluators and we then improve what we do. In practice, however, those most directly concerned, both aid providers and aid recipients, do not find evaluations critical to their work or perhaps even useful. Indeed, as was the case in this situation, not infrequently they are not aware of evaluation findings or recommendations. Though the evaluation had surely been sent to the ministry and was readily available, the education ministry officials whose work was directly affected by its findings neither had a copy nor, they said, knew where to find one.

Nor were the aid providers, either those working within the country or those at headquarters, well informed about the evaluation and its findings. As they discussed their on-going work, they were clear that evaluations were not primary inputs and that familiarity with evaluation findings was not a high priority in their work lives. The only people reasonably familiar with the evaluation were the evaluators, in this case a firm contracted by the funding agency. That is, *those best able to use the evaluation to shape policy and practice had no direct role in either policy or practice.*

What do we learn here?

A single case study, of course, is just that. Without further work, we cannot confidently generalize from that experience. Still, other research, interviews with people directly involved in aid and evaluations, and our review of evaluation documents confirm that this situation was not unique, and that it is instructive.

First, put sharply, evaluations, even where they are directly relevant to their work, do not feature prominently in the daily lives of educators in aid-receiving countries. When they develop new initiatives, educators do not turn to evaluations for information and guidance. Regularly, they have at best a dim recollection of potentially relevant evaluations and no direct access to their details. If evaluations influence subsequent action, it is not through a direct link between the report of evaluation findings and decisions on education programs.

Second, the important learning in this example was among those involved in the funded project and was not, it seems, stimulated by or captured in the evaluation. Those involved did learn from experience and did use that learning to shape their subsequent work. Not only did they not need that evaluation for their learning, but they did not regard that evaluation as *their* tool, responsive to *their* needs, readily appropriate and incorporated into *their* thinking and decisions.

To be clear, the educators were not excluded from the evaluation. Consulted as its terms of reference were drafted, the education ministry had opportunities for input throughout the evaluation. Education ministry staff participated in selecting sites, establishing contacts, and conducting interviews. Ministry officials received and commented on the draft evaluation. Even so, as they pursued their responsibilities, they largely ignored it. In part, that may reflect some changes in personnel. But that is not a satisfactory explanation, since personnel changes are an ordinary feature of schools and their administration. Evaluations whose utility depends on a single individual or two are unlikely to have much use.

A more powerful explanation is that throughout what appeared to be a participatory process, the education officials regarded the evaluation largely as an external event, a requirement of the aid process. We see clearly here that ownership matters, not only for aid-funded education activities, but also for their evaluations.

Third, perhaps most important, evaluations that limit their view to inputs and outputs, or that document process mechanically without exploring interconnections and interactions—that ignore complexity and context—are unable to produce findings that influence

subsequent behaviour. In this case, the evaluators reported on what was and was not done, but not to whom that mattered. The evaluators assessed progress on specified objectives but did not examine what we have termed ownership, that is, whether or not the education ministry regarded the aid-supported activities as its initiative, to be institutionalized, protected, funded, and maintained. The evaluators talked with education officials, but neither explored their interactions with those directly involved in the funded project nor developed an approach that included the officials as collaborators in the evaluation. The evaluators reviewed documents on national policy but did not refine their reading by exploring either the locus of interest in the aid-funded activity or the locus of authority for sustaining it.

Inattention to complexity and context sorely limited, indeed undermined, both the substantive quality of the evaluation and its utility.

It is useful here to return for a moment to the common assertion that randomized controlled trials are the most scientific, that is most valid and most reliable, strategy for evaluating education initiatives and reforms. That perspective presents RCTs as a methodology that renders complexity and context less important in explaining observed outcomes. These case studies help us understand why the general form of that claim is untenable.

First, that orientation seeks to ignore complexity and context by controlling for factors other than the inputs to be measured that might influence the outcomes. Practical constraints limit the number of factors that can be controlled. How, then, to determine which factors require high priority attention? It is a deeper understanding of the complex interactions in education that is necessary to select the factors to be controlled. Randomization is the alternative approach for addressing confounding influences. The assumption is that in a large population, factors other than the input to be measured are evenly distributed among learners who experience the new program and learners who do not. But how could we confirm that? A deeper understanding of context is required to determine whether or not alternative influences are randomly distributed across the population or found unevenly among learners.

Second, that orientation deals awkwardly at best with situations where maximization is not the highest priority. Where the infrastructure is weak, for example, educators may find it more

important to assure redundancy than to maximize the benefit of a particular input.

Third, educators may find statistical significance (the likelihood that an outcome will occur by chance) less important than analytic significance (the consequences of a program for the education system have higher priority than increased confidence in the causal chain).

Education is by design interactive. Nearly always, how an outcome is achieved is at least as important and perhaps more important than the outcome itself. Inattention to complexity and context undermines our ability to understand and explain that.

Formative and participatory evaluations

Throughout our review, we have noted challenges and problems in the most common evaluation approaches. Evaluations often do not address the deep-rooted and structured relationships that determine the effectiveness and sustainability of poverty reduction efforts (Ofir and Kumar, 2013). While a few funding agencies, among them Sida, stand out in their use of participatory approaches, few and far between are evaluations that include the voices of those most affected by limited education access and poor education quality. Since participatory evaluation can address some of those problems and at the same time is criticized as non-scientific and not objective, it is important here to comment on participatory evaluation and to explore its use in the evaluations we have reviewed.

Participatory approaches are widespread in international development, attracting increased interest as a response to the limits of top-down approaches in the 1970s and 1980s, especially where funding agency priorities sometimes seemed incompatible with the needs of intended beneficiaries. A key objective is to empower the community to conduct its own analysis of its needs and priorities, and organize these community-driven elements into a plan of action (Bamberger et al, 2015). Participatory approaches generally work through community groups rather than through individuals and often rely heavily on mapping and graphical techniques to structure participation and to include community members who may not be literate. Participatory evaluation encompasses a way to understand and include the needs of diverse constituencies and to understand the context of the aid delivery process. While participatory strategies can be used to evaluate participatory development approaches, they have

much broader utility. Involving recipients in assessing development assistance can not only deepen and strengthen observations and findings but can also substantially increase the use of evaluations that all too often are simply another document to be noted and filed.

Participatory approaches include participatory rural appraisal (PRA), participatory action research (PAR), and participatory learning and action (PLA) as well as asset-based community development. PRA comprises a family of approaches, methods, and behaviours to enable poor people to express and analyse their lives, and to plan, monitor and evaluate their actions (Chambers, 1994). PRA evolved out of Rapid Rural Appraisal (RRA), developed as an alternative to earlier top-down approaches and surveys based on questionnaires. PAR engages research design, methods, analyses, and findings with the participation of diverse institutions under study. The aim of the inquiry and the research questions develop out of the convergence of the perspective of science and the perspective of practice (Bergold and Thomas, 2012). PLA includes participatory and visual methods with natural interviewing techniques, to facilitate collective analysis and learning, moving beyond consultation and promoting community participation in issues of relevance to their own development (FAO, 2015). Participatory approaches and methods also include stakeholder analysis, storytelling, social mapping, causal-linkage and trend and change diagramming, scoring, and brainstorming on program outcomes (Chambers, 1994). Responding to a lack of conceptual clarity on what constitutes a participatory approach, and what makes an evaluation participatory, Cullen et al., (2011) propose a three-dimensional framework for classifying participatory evaluation approaches, examining which stakeholders participate, in what capacity, and during which evaluation phases.

Local capacity to generate and analyse information is often significantly greater than outsiders assume. Participatory approaches encourage evaluators to be facilitators who assume local capacity until proven otherwise (Chambers, 1994). Local individuals familiar with participatory rural approaches (PRA) have proven to be better facilitators than outsiders (Shah et al., 1991). Yet, the use of rapid assessment to address complex social issues risks superficiality. To ensure that the process element of social development is addressed systematically and critically, research teams can include social scientists with a strong conceptual background in poverty analysis (Norton et al., 2001: 28). Another added value to participatory evaluation approaches is that visual means of data collection (such as

maps, models, or diagrams) are often easier to triangulate than personal, individually collected information such as questionnaires. In shared diagrams or maps, triangulation occurs as participants crosscheck and create knowledge together (Chambers, 1994).

When participatory methods are well-designed and implemented, they are rigorous, and provide information that can address gaps in demographic and other quantitative data that may be otherwise overlooked, especially by external evaluators in the field for a short time. Indeed, participatory strategies contribute to a whole society approach to ownership and empowerment in the development process.

UNICEF's evaluation of support to basic education in the Democratic Republic of the Congo provides an instructive example of a participatory evaluation within a Real World Evaluation framework (Eval: UNICEF 2012). The evaluation examined the planning context, interventions, results, and impact of the program, with a particular focus on implementation gaps, constraints, weaknesses, and achievements as well as sustainability. To address validity and rigor, an external independent company reviewed and rated all evaluation reports.

Using a mixed-methods approach, the evaluation integrated information from diverse sources. The evaluation also addressed data limitations, reconstructing baseline data when there were gaps, using secondary data sources, key informants, focus groups, construct mapping, and PRA techniques. A major focus was to explore the consequences of increased local participation. Evaluators found that community involvement worked particularly well in the early childhood education program and in rapid assessments for IDPs in emergency situations, the extent of local ownership remained very uneven and local actors felt frustrated by what they saw as firm constraints on their involvement. Local project directors, it turned out, were not well versed in and perhaps hostile to inclusive approaches. This evaluation also shows that a participatory evaluation may generate more and more reliable quantitative data, particular in contexts of fragility, by involving local residents in the development of context-based indicators.

Yet, despite these effective uses of participatory evaluation approaches, they are regularly contested, as they are often difficult to implement consistently. Though they are widely employed, there is relatively limited empirical research on why and how participatory

evaluation approaches are used in international development (Cullen et al., 2011), especially how they differ in interpretation and practice.

Not surprisingly, the many variations of participatory evaluation and their sometimes sharp methodological differences fuel continuing contention about its strengths and limitations. Participation can be viewed either as a desired outcome or as a process by which to achieve an objective (Morra Imas and Rist, 2009). Participatory methods can be seen as an expansion of decision-making and at times, an opportunity to shift power dynamics and promote social change. Scholars of evaluation debate whether or not the purpose of evaluation is as expansive as shifting power dynamics and promoting social change. Critics of participatory approaches contest the inclusion of participants in evaluation, citing a threat to objectivity. For many evaluators, participation means that the objective of the evaluation becomes participation, losing sight of the initial objective for which the evaluation was commissioned.

Yet, when effectively implemented, participation yields substantial information that fills the gap left by sole reliance on other methods. Weaver and Cousins indicate three positive results from participatory approaches in evaluation: when stakeholders are included in the evaluation process, findings are more useful, there is more fairness, and inclusion of the unique perspectives of stakeholders improves validity and credibility (2004). Program stakeholders may share contextual considerations, particularly in situations where the evaluation is done externally and in a very limited time, which appears often to have been the case in the evaluations of aid to education we reviewed.

Including a broader range of stakeholders in the evaluation process may also increase the use of evaluation findings (Cullen et al., 2011; Brandon, 1998, 1999; Cousins, 2003; Patton, 2008; Ryan, Greene, Lincoln, Mathison, and Mertens, 1998; Weiss, 1986). As more diverse stakeholders are included, the evaluations will necessarily address a wider range of priorities, leading to an evaluation process that is more democratic, more sensitive, and more fair (Weaver and Cousins, 2004). That is particularly important in in settings of recent or current violent conflict. Tracing the evaluation from provider to recipient shows clearly that evaluation consumers are more likely to follow evaluation conclusions when staff actively participated in the process (Brandon, 1998) and more committed to acting on findings because they had a voice in the process (Weiss, 1986).

To be successful, an environment conducive to participation is key, including managing conflicts among stakeholders (Cullen et al., 2011). Expanding the pool of participants and thereby increasing the prospect of broad ownership of the evaluation and effective use of its findings requires evaluator initiative and flexibility. Usually, data collection is the evaluation phase with the greatest stakeholder participation, whereas data analysis has the least participation (Cullen et al., 2011). Yet, participatory approaches are necessary precisely because program design is typically the domain of technicians, distant from actual program beneficiaries.

Typically, a participatory evaluation begins by asking why the evaluation is being conducted, who are the intended beneficiaries, what outcomes are expected, and what approaches are to be employed. Essential questions also include: Who will be included, in what capacity, and in which evaluation phases? What will be participants' roles? Who will make decisions concerning the evaluation? (Cullen et al., 2011). Throughout the process, the language used must enable and foster participation and be accessible to diverse constituencies and across gender and social categories.

Participatory evaluations can be formative as well as summative. They can thus address what are often divergent evaluation objectives: providing grounded and timely feedback to aid recipients and facilitating the end-of-project assessments required by funding agencies.

Available research on participatory evaluations regularly confirms that substantive participation that increases project effectiveness goes beyond soliciting diverse constituencies as interviewees and data collectors. Active engagement in the evaluation process requires recipient participation from the outset, from conception and design through implementation and interpretation. Often, however, evaluators use the term participatory but treat participants as subjects of the evaluation rather than collaborative evaluators.

The GIZ Chad evaluation (2005), assessing the use of parent associations in local ownership and capacity building, particularly for girls' education, is one example of a participatory approach to evaluation and to development. The inclusive evaluation process enabled evaluators to learn of implementation challenges that were not readily apparent. The evaluation concluded that a local approach and coordination with targeted groups and intermediaries are

preconditions for successful implementation of the development support. Parent associations were eventually included in the national sectoral policy as a result of this project, both strengthening them and expanding participation in the policy process.

A second example of how participatory evaluation can provide rigorous evidence is USAID Benin (2005). This evaluation emphasized working on the local level, with parent and other local organizations. The assessment of outcomes, for example children's learning and increased parents' role in school management, could not have been made without participatory evaluation. This evaluation was also able to illuminate and document the roles of parent associations. Though at first glance they may seem more scientific, evaluations that do not include these locally grounded assessments are in practice not only less inclusive but also less rigorous.

By working directly with a local evaluation unit and providing capacity building for the education ministry the evaluation of support to education in Bénin developed a more penetrating understanding of the power struggles that framed the decentralization initiative (AFD/DANIDA/MCPD, 2012). In this setting, the participatory approached helped the funding agencies uncover how, why, and in what circumstances development assistance and development policy affected outcomes.

Participatory evaluations thus bring clear benefits at smaller and larger scales. Local voices are the most effective stewards of ensuring that methods do not determine outcomes. Participatory evaluations must also confront and manage important challenges. Some are practical. Participatory approaches may result in increased time and financial demands and difficulty addressing the needs of multiple constituencies. Participatory approaches demand participants' time and can raise participant expectations, itself a potential benefit (Norton, 2001: 16). Specific efforts are required to ensure that evaluations are inclusive across diverse socioeconomic groups. Some challenges are theoretical or methodological. Critics insist that including stakeholders in evaluations heightens the risk that stakeholder bias may reduce the validity of the evaluation. Selection of stakeholders may also be contentious, as funding agencies may try to select only those aid recipients who have shown positive results. However, since evaluators and funding agencies, and not aid recipients, generally retain control of the evaluation process, claims

that participatory evaluation compromises objectivity and possibly validity due to stakeholder self-interest, are less persuasive.

Participatory evaluation approaches also offer the prospect that the evaluation itself can have positive development consequences. As they provide context, local participants generate new research questions and indicate areas where closer attention and detailed analysis are needed. Participatory evaluations can themselves empower local citizens to participate in policy and to have a voice in their communities.

Participatory evaluation approaches are neither unproblematic nor universally appropriate. They can, however, reduce three risks that have emerged sharply in our review of evaluations. First, by their nature participatory evaluation approaches require the attention to context and complexity that is essential for understanding the roles and consequences of development assistance. Second, where they are designed to play a formative as well as summative role, participatory evaluations can be a generative input for aid recipients rather than an imposed burden that has no immediate relevance. Third, by broadening the ownership of the evaluation process, recipient participation substantially increases the likelihood that evaluation findings and recommendations will be used, by funders as well as recipients.

Too many evaluations have too little use

Our review found limited evidence that evaluations are used for one of their intended purposes: to improve the quality of aid-funded education projects. With some exceptions, the majority of the evaluations we reviewed did not summarize or note findings from previous evaluations, contrary to the notion that evaluations are integral components of evidence-based policy. Case study analyses support this observation: while our respondents consistently emphasized the importance of evaluations in general, few could provide concrete examples of evaluation-induced changes in policies or practices.

We have highlighted multiple reasons for this. Decontextualized evaluation approaches, superficial or weakly supported analyses and recommendations, mismatched time-horizons, and attribution challenges mean that evaluations rarely provide actionable results that feed directly into project design and implementation. Professional priorities, institutional reward systems,

sharply constrained institutional learning, and over-stretched demands on their time make evaluations both required and at the same time of limited direct utility to funding agency education staff. Narrow ownership of the evaluation process regularly makes evaluations a periodic intrusion rather than a constructive contribution for funding agency and recipient country educators.

Where required evaluations go far beyond what educators deem useful and regularly overwhelm capacity, they are likely to become formalistic exercises, completed when necessary and ignored as soon as possible. Not infrequently, it turns out, evaluations are technically sound, extensive, perhaps expensive, and largely ignored. More evaluations, less use.

Together, these findings support the conclusion that different purposes require different types of evaluations. Funding agencies are interested in ensuring that their funds are used as intended, and in determining who and what to fund. Governments want to ensure their education policies align with national priorities and political objectives. Implementing organizations want to improve their operations in order to attract continued support. Teachers, families, and communities want to know how to support children's learning. No single type of evaluation will meet all of these objectives.

Aid Agencies' Data Demands

Periodic voices note that funding and technical assistance agencies could draw on the measures that education officials use to manage their education system. Occasionally an agency official argues just that. Currently, however, funding agencies require measurement and data collection that far exceed the needs of day-to-day education management in high income as well as low income countries. Education managers in, say, Tanzania, are expected to collect, analyse, and report on many more measures and much more data than are used by education managers in a European or U.S. city with a much larger education budget. As well, since most funding agencies insist that aid recipients use the provider's recording and accounting systems, countries like Tanzania must prepare thousands of reports each year on the aid they receive and host numerous funding agency visits to monitor programs and negotiate new support. Even though the funding agencies have supported the establishment of education management information systems, regularly the demand for education

data and analysis overwhelms the capacity of the aid receiving countries. Put sharply, the incessant demand that low income countries collect, manage, and analyse ever more data diverts experience and expertise from the education activities that the aid is intended to support. In the aid relationship, *aid management becomes an obstacle to aid effectiveness.*

5. Re-thinking evaluations and their role

What do we learn about evaluations from our review of evaluations of aid-funded education activities? How to balance evaluation complexity, cost, and utility?

Earlier in this report we highlighted major findings concerning education initiatives and especially the ways in which the aid process has been more or less effective in supporting education innovation and reform. It is fruitful here to focus critical attention on evaluations and the evaluation process.

With occasional exceptions, more and more complex evaluations are unlikely to improve education or increase aid effectiveness. Especially where there is little local generative participation in the evaluation process, there is likely to be little local ownership of evaluations, little local engagement in their elaboration and implementation, and little local attention to their findings. In the absence of broader attention to their roles, better evaluation design and increased scientific rigor cannot solve these problems.

For funding agencies, the implications are several.

Where evaluations are needed to confirm that aid funds were used as intended, limit the evaluations to that role. For that purpose, evaluations can be much simpler, less costly, and less time consuming for both providers and recipients.

Where evaluations are intended to serve other purposes, say increasing local transparency and accountability for aid flows, they can be designed and managed for those purposes.

Complex and expensive evaluations by detached outsiders can serve occasional narrowly defined objectives but have limited general utility. While their findings are presented as definitive, often so too are sharply divergent findings generated through a similar approach. Far

more cost-effective and more likely to be used are evaluations that achieve reliability, validity, and legitimacy through the systematic inclusion of aid recipients from conception through implementation to interpretation and that incorporate both formative and summative objectives. Lagging is the development of evaluation strategies that recognize that data collection and analysis are no longer the exclusive domain of experts.

Though evaluators regularly note that their assignment leaves no time to address broader questions, carefully designed evaluations can review relevant history, extract and synthesize findings and interpretations helpful in the current task, and thereby contribute to institutional learning. Funding agencies can encourage that by recognizing that they are both the initiator and an important subject of evaluations. Drawing on the evaluations they commission and on the work of their professional staff, funding agencies can become productively more self-reflective.

Evaluations can themselves become part of development assistance. Where they incorporate significant recipient participation, and especially where they are well integrated into aid-supported activities and provide formative results, evaluations can be empowering. They can as well structure accountability to aid recipients, unusual but important to a healthy aid relationship. As we have noted, that orientation can generate otherwise difficult-to-secure information and can strengthen an evaluation's reliability and validity. In many circumstances the benefits of this orientation will outweigh the advantages of an evaluation undertaken by detached outsiders.

Regularly, funding agencies take risks in supporting innovation in education. A parallel willingness to take risks in evaluation will encourage the development of innovative approaches to understanding the consequences (intended and unintended) and impacts (desired and problematic) of both education reform and external support.

Rather than a standard evaluation approach to be used broadly, funding agencies and supported education systems can develop a portfolio of evaluation sorts and types, appropriate to different circumstances. Both aid providers and aid recipients will find it useful to increase the proportion of evaluations that are formative, rather than summative. Focusing on educators' evaluation needs and uses is more likely to improve education outcomes than the common focus on aid providers' monitoring requirements.

Ensuring local ownership of evaluations does not exclude the possibility of conducting experimental or quasi-experimental impact evaluations. Where there is local demand, RCTs and quasi-experimental methods can be used as instruments to explore specific, locally defined evaluation questions. Baseline data, for example, can be shared with implementers in order to develop tailored implementation strategies that match students' specific educational needs. Likewise, if the purpose of the evaluation is to learn, rather than to monitor or to supervise, end-line impact estimates can be used to identify mediators (such as increased attendance, improved teacher morale, greater access to print material) and moderators (such as gender and ethnicity) of program effects. When accompanied with process evaluations and qualitative assessments, these types of impact estimates can be used to answer *why*, *how*, and *in what circumstances* evaluation questions.

Rather than the generally unachievable objective of determining what works or what works best, evaluations can be designed to examine how things work in specified circumstances and then used to improve both the education and the aid process.

Funding agencies can learn from the research on public policy. Evaluations that are good enough may be far more useful and far more used than evaluations that seek unimpeachable accuracy and validity. In the often disorderly and regularly chaotic arena of education, evaluations that are satisfactory and sufficient may do more to improve education and aid effectiveness than evaluations that claim to be rational, linear, and optimal. If so, then most evaluations can be more modest, not more but less complex.

Since local ownership of evaluations matters as much as local ownership of education reform, evaluations can be designed with local ownership as a primary priority. That will require not only assuring deep local participation from the outset, far beyond formal consultation, but also transferring major responsibility for evaluations to those expected to use their results. Will that shared control encounter other problems? Certainly. Still, that will support education better than evaluations that are resisted, tolerated, and ignored.

While evaluation by detached outsiders, or teams led and managed by detached outsiders, will strengthen some evaluations, that approach, as we have seen, renders other evaluations less useful. Both education and aid will benefit from evaluations and evaluators rooted within the activities to be assessed and from encouraging

administrators, teachers, and learners to incorporate reflection and evaluation in their daily work.

6. References

Angrist, J., & Lavy, V. (1999). Using Maimonides Rule to Estimate the Effect of Class Size on Scholastic Achievement. The Quarterly Journal of Economics, 533–575.

Bamberger, M. J., Rugh, J. and Mabry, L. S. (2012). RealWorld evaluation: Working under budget, time, data, and political constraints. 2nd ed. Thousand Oaks, CA: SAGE Publications.

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2005). Remedying education: Evidence from two randomized experiments in India (No. w11904). National Bureau of Economic Research.

Banerjee, Abhijit, and Esther Duflo. (2011). Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty. Public Affairs.

Berg, E. (2000). Why Aren't Aid Organizations Better Learners? Expert Group on Development Issues (24 August). http://www.observatoritercersector.org/pdf/centre_recursos/2_14_why_00406.pdf [2015.11.15].

Bergold, J. and Thomas, S. (2010). 'Partizipative Forschung', Handbuch Qualitative Forschung in der Psychologie, pp. 333–344. doi: 10.1007/978-3-531-92052-8_23.

Bernard, Tanguy, Jocelyne Delarue, and Jean-David Naudet. (2012). Impact Evaluations: A Tool for Accountability? Lessons from Experience at Agence Française de Développement. Journal of Development Effectiveness 4 (2): 314–27. doi:10.1080/19439342.2012.686047.

Bradstock, Alastair, and Steve Bass. (2013). "Evaluating Sustainable Development." In Donaldson et al. (2013).

Brandon, P. R. (1998). Stakeholder participation for the purpose of helping ensure evaluation validity: Bridging the gap between collaborative and non-collaborative evaluations. American Journal of Evaluation 19(3): 325–337. doi: 10.1177/109821409801900305.

Brandon, P. R. (1999). Involving program stakeholders in reviews of evaluators' recommendations for program revisions. Evaluation and Program Planning 22(3): 363–372. doi: 10.1016/s0149-7189(99)00030-0.

Carden, Fred, and Colleen Duggan. "Evaluating Policy Influence." (2013). In Donaldson et al. (2013).

Carlsson, Jerker, and Lennart Wohlgemuth. (Eds.) (2000). Learning in Development Co-operation. EDGI Study 2000:2. Stockholm: Almqvist & Wiksell International.

Chambers, R. (1994). The origins and practice of participatory rural appraisal. World Development 22(7), 953–969. doi: 10.1016/0305-750x(94)90141-4.

Chapman, David W., and Jessica Jester Quijada. (2009). An Analysis of USAID Assistance to Basic Education in the Developing World, 1990–2005. International Journal of Educational Development 29 (3): 268–80. doi:10.1016/j.ijedudev.2008.08.005.

Chapman, David W., and Audrey Schuh Moore. (2010). A meta-look at meta-studies of the effectiveness of development assistance to education. International Review of Education 56: 546-565.

Clements, Paul, Thomaz Chianca, and Ryoh Sasaki. (2008). Reducing World Poverty by Improving Evaluation of Development Aid. American Journal of Evaluation 29 (2): 195–214. doi:10.1177/1098214008318657.

Cousins, B. J. (2003). Utilization effects of participatory evaluation. International Handbook of Educational Evaluation, 245–265. doi: 10.1007/978-94-010-0309-4_16.

Culbertson, Michael J., Daniel McCole, and Paul E. McNamara. (2014). Practical Challenges and Strategies for Randomised Control Trials in Agricultural Extension and Other Development Programmes. Journal of Development Effectiveness 6 (3): 284–99. doi:10.1080/19439342.2014.919339.

Cullen, A. E., Coryn, C. L. S. and Rugh, J. (2011). The politics and consequences of including Stakeholders in international development evaluation. American Journal of Evaluation 32(3), 345–361. doi: 10.1177/1098214010396076.

Dahlman, C. (2008). Innovation strategies of the BRICKS: Brazil, Russia, India, China, and Korea. Presentation at OECD-World Bank Conference on Innovation and Sustainable Growth in a Globalised World.

Deaton, Angus S. (2009). Instruments of development: Randomization in the tropics, and the search for the elusive keys to

economic development. Cambridge: National Bureau of Economic Research, NBER Working Paper 14690.

Deaton, Angus S. (2010). Instruments, randomization, and learning about development. Journal of Economic Literature, 48 (June): 424–455.

Donaldson, Stewart I., Tarek Azzam, and Ross F. Conner (Eds). (2013). Emerging Practices in International Development Evaluation. Charlotte, NC: Information Age Publishing.

Duflo, E., Dupas, P., and Kremer, M. (2009). Additional resources versus organizational changes in education: Experimental evidence from Kenya (Unpublished manuscript.). Cambridge, Mass.: Massachusetts Institute of Technology: Abdul Latif Jameel Poverty Action Lab.

Duflo, E. and Banerjee, A. V. (2011). Poor economics: A radical rethinking of the way to fight global poverty. New York: Public Affairs Press.

Evans, David, and Anna Popova. (2015). What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews. WPS7203. The World Bank.
http://documents.worldbank.org/curated/en/2015/02/24060240/really-works-improve-learning-developing-countries-analysis-divergent-findings-systematic-reviews.

Fives, Allyn, Daniel W. Russell, John Canavan, Rena Lyons, Patricia Eaton, Carmel Devaney, Norean Kearns, and Aoife O'Brien. (2015). The Ethics of Randomized Controlled Trials in Social Settings: Can Social Trials Be Scientifically Promising and Must There Be Equipoise? International Journal of Research & Method in Education 38 (1): 56–71. doi:10.1080/1743727X.2014.908338.

Food and Agriculture Association (FAO). (2015). Introducing Participatory Approaches, Methods and Tools. http://www.fao.org/docrep/006/ad424e/ad424e03.htm [2015.11.15].

Forss, Kim, Basil Cracknell, and Nelly P. Stromquist. (1998). Organisational learning in development co-operation: How knowledge is generated and used. EDGI Working Paper 1998:3. Stockholm: Ministry for Foreign Affairs.

Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor. (2014). Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica. Science 344 (6187): 998–1001. doi:10.1126/science.1251178.

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. (2010). Impact Evaluation in Practice. The World Bank. http://elibrary.worldbank.org/doi/book/10.1596/978-0-8213-8541-8.

Grant Lewis, Suzanne M., and Joel Samoff, editors. (1992). Microcomputers in African Development: Critical Perspectives. Boulder: Westview Press.

Grantham-McGregor, Sally, Yin Bun Cheung, Santiago Cueto, Paul Glewwe, Linda Richter, and Barbara Strupp. (2007). Developmental Potential in the First 5 Years for Children in Developing Countries. The Lancet 369 (9555): 60–70. doi:10.1016/S0140-6736(07)60032-4.

Greenberg, D., and Shroder, M. (2004). The Digest of Social Experiments (Third Edition). Washington: Urban Institute Press.

Greenhalgh, T., Wong, G., Westhorp, G., & Pawson, R. (2011). Protocol—realist and meta-narrative evidence synthesis: Evolving Standards (RAMESES). BMC Medical Research Methodology, 11(1), 115. http://doi.org/10.1186/1471-2288-11-115

Greenwood, Royston, Amalia Magan Diaz, Stan Xiao Li, and José Céspedes Lorende. (2010). The multiplicity of institutional logics and the heterogeneity of organizational responses. Organization Science 21(2): 521-539.

Grindle, M. (2010). Social Policy in Development: Coherence and Cooperation in the Real World. Working Paper 98. United Nations: Department of Economic & Social Affairs.

Jerven, Morten. (2013). Poor Numbers: How We Are Misled by African Development Statistics and What to Do About It. Ithaca: Cornell University Press.

Jerven, Morten. (2015). Africa: Why Economists Get It Wrong. London: Zed Books.

Kloos, Karina, Oberg, Achim, Oelberger, J. Carrie, and Powell, Walter. (2014). Measuring Missions: The Distribution of Discourse on Evaluation for the Nonprofit Sector. ISTR 10th International Conference, Universita Degli Studi Di Siena, Siena, Italy. http://citation.allacademic.com/meta/p_mla_apa_research_citation/5/5/3/0/2/p553024_index.html [12 Dec 2015].

Krishnaratne, Shari, Howard White, and Ella Carpenter. (2013). Quality education for all children? What works in education in developing countries. International Initiative for Impact Evaluation Working Paper 20 (September).

Krohwinkel-Karlsson, Anna. (2007). Knowledge and Learning in Aid Organizations. Karlstad, Sweden: Swedish Agency for Development Evaluation (SADEV).

Krueger, A. (1997). Experimental estimates of education production functions. Working Paper No. w6051. National Bureau of Economic Research.

Lindblom, Charles E. (1959). The science of 'muddling through.' Public Administration Review 19, 2 (Spring): 79-88.

Lindblom, Charles E. (1979). Still Muddling, Not Yet Through. Public Administration Review 39, 6 (November-December): 517-526.

Lloyd, Rob, Derek Poate, and Espen Villanger. (2014). Results Measurement and Evaluability: A Comparative Analysis. Journal of Development Effectiveness 6 (4): 378–91. doi:10.1080/19439342.2014.966455.

Lloyd, Rob, and Espen Villanger. (2014). Assessing Aid Impacts Revisited: Results Measurement in Norwegian Aid. Journal of Development Effectiveness 6 (4): 461–79. doi:10.1080/19439342.2014.963883.

Maclure, Richard. (1993). School Reform in Burkina Faso: The Limited Prospects of Decentralization and Local Participation. Canadian and International Education 22 (2): 69-87.

MacPherson, Nancy. (2013). "Preface." In Donaldson et al. (2013): xi-xiii.

Masino, Serena, and Miguel Niño-Zarazúa. (2015) What works to improve the quality of student learning in developing countries?

(Tokyo: United Nations University, World Institute for Development Economics Research).

McEwan, P. J. (2015). Improving learning in primary schools of developing countries. A meta-analysis of randomized experiments. Review of Educational Research 85,3: 353-394 (September).

Miguel, E., and Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. Econometrica, 72(1), 159–217.

Morgan, Peter. (2013). "Evaluating Capacity Development." In Donaldson et al. (2013).

Morra-Imas, L. G. and Rist, R. C. (2009). The road to results: Designing and conducting effective development evaluations. Washington: World Bank.

Nkansa, Grace Akukwe, and David W. Chapman. (2006). Sustaining community participation: What remains after the money ends? International Review of Education 52,6: 509-532.

Norad. (2004). Study of the impact of Norwegian NGOs on civil society: FORUT (Sri Lanka) and Save the Children Norway (Ethiopia).

Norad, (2014). Added costs. Added value? Evaluation of Norwegian support through and to umbrella and network organisations in civil society, Report 5/2014.

Norton, A., Bird, B., Brock, K., Kakande, M. and Turk, C. (2015). A rough guide to PPAs: Participatory poverty assessment - an introduction to theory and practice. Overseas Development Institute.

Ofir, Zenda and A. K. Shiva Kumar. (2013). "Evaluation in Developing Countries: What Makes it Different?" In Donaldson et al. (2013).

Olofsgard, Anders. (2014). Randomized Controlled Trials: Strengths, Weaknesses and Policy Relevance. 2014:1. Stockholm, Sweden: Expert Group for Aid Studies (EBA).

Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J. and Seller, R. (2013). Mixed methods appraisal Tool—2011 version. PsycTESTS Dataset. doi: 10.1037/t21090-000.

Pache, Anne-Claire, and Filipe Santos. (2010). When worlds collide: the internal dynamics of organizational responses to conflicting institutional demands. Academy of Management Review (35)3, 455-476.

Patton, M. Q. (2008). Utilization-focused evaluation. 4th ed. Thousand Oaks: Sage Publications.

Paul T., S. (2004). School subsidies for the poor: Evaluating the Mexican Progresa poverty program. Journal of Development Economics 74(1), 199–250. doi: 10.1016/j.jdeveco.2003.12.009.

Pawson, Ray. (2002). Evidence-based Policy: The Promise of 'Realist Synthesis'. Evaluation, 8(3), 340–358. http://doi.org/10.1177/135638902401462448.

Pawson, Ray, and Nick Tilley. (1997). Realistic Evaluation. Thousand Oaks: Sage.

Pitman, George Keith, Osvaldo Néstor Feinstein, and Gregory K. Ingram. (2005). Evaluating Development Effectiveness. Transaction Publishers.

Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'Cathain, A., Griffiths, F., Rousseau, M. C. (2011). Proposal: A mixed methods appraisal tool for systematic mixed studies reviews. http://mixedmethodsappraisaltoolpublic.pbworks.com [2015.03.30]

Pritchett, Lant. (2009) Is India a failing state? Detours on the Four Lane Highway to Modernization. Cambridge: Harvard University John F. Kennedy School of Government, Working Paper Series 09-013.

Pritchett, Lant. (2013). RCTs in Development, Lessons from the Hype Cycle. Center For Global Development. November 14. http://www.cgdev.org/blog/rcts-development-lessons-hype-cycle.

Pritchett, Lant. (2015). "Using 'Random' Right: New Insights from IDinsight Team," Center for Global Development Blog: http://www.cgdev.org/blog/using-"random"-right-new-insights-idinsight-team [2016.01.21].

Pritchett, Lant, and J. Sandefur. (2013). Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix (Working Paper No. 336). Center for Global Development.

Ramalingam, B. (2005). Implementing Knowledge Strategies: Lessons from international development agencies. http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/166.pdf [2015.11.15].

Reddy, Sanjay G. (2009) The emperor's new suit: Global poverty estimates reappraised. New York: United Nations, UN-DESA Working Paper 12.

Reddy, Sanjay G. (2012). Randomise This! On Poor Economics. Review of Agrarian Studies 2 (2): 60–73.

Riddell, Abby. (2012). The Effectiveness of Foreign Aid to Education: What Can Be Learned? 2012/75. WIDER Working Paper. http://www.econstor.eu/handle/10419/80938.

Rochlin, Steve, and Sasha Radovich. (2013). Future Directions for Improving International Development Evaluations. In Donaldson et al. (2013).

Ryan, K. (1998). Advantages and challenges of using inclusive evaluation approaches in evaluation practice. American Journal of Evaluation 19(1), 101–122. doi: 10.1177/109821409801900111.

Samoff, Joel, with Bidemi Carrol. (2013). "Education for All in Africa: Not Catching Up but Setting the Pace," in Robert F. Arnove, Carlos Alberto Torres, and Stephen Franz, editors, Comparative Education: The Dialectic of the Global and the Local. Lanham, MD: Rowman & Littlefield, Fourth Edition, 2013, 403-443.

Samoff, Joel, Martial Dembélé and E. Molapi Sebatane. (2011). 'Going to Scale': Nurturing the Local Roots of Education Innovation in Africa. Bristol: University of Bristol, EdQual Working Paper 28. http://www.edqual.org/publications/workingpaper/edqualwp28

Samoff, Joel, Martial Dembélé and E. Molapi Sebatane. (2012). "Scaling Up by Focusing Down: Creating Space and Capacity to Extend Education Reform in Africa," in Leon Tikly and Angeline M. Barrett, editors, Education Quality and Social Justice in the South: Challenges for policy, practice and research (London: Routledge).

Schultz, T. Paul. 2001. School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program. Working Paper 834. Economic Growth Center, Yale University. https://ideas.repec.org/p/egc/wpaper/834.html.

Shah, P., Bharadwal, G. and Ambastha, R. (1995). Farmers as analysts, facilitators and decisionmakers. Power and Participatory Development 83–94. doi: 10.3362/9781780445649.007.

Sida (2014). The Mid-Term Review of The West Africa Network for Peacebuilding (WANEP).

Simon, Herbert A. (1956). Rational choice and the structure of the environment. Psychological Review, 63: 129-138.

Simon, Herbert. (1982). Models of Bounded Rationality. Cambridge, MA: MIT Press.

Simon, Herbert. (1997). Administrative Behavior: A study of decision-making processes in administrative organization. New York: Free Press, Fourth Edition.

Stufflebeam, D. L. (2001). Evaluation models: New directions for evaluation. San Francisco: Jossey-Bass Inc.

Sturdy, Jennifer, Sixto Aquino, and Jack Molyneaux. (2014). Learning from Evaluation at the Millennium Challenge Corporation. Journal of Development Effectiveness 6 (4): 436–50. doi:10.1080/19439342.2014.975424.

Sweden. (2002). Shared Responsibility: Sweden's Policy for Global Development. http://www.government.se/contentassets/e9b903fda24f4a778cf7b06da7c10ef9/shared-responsibility-swedens-policy-for-global-development-government-bill-200203 [2015.11.15].

Sweden. (2013). Aid policy framework—the direction of Swedish aid. http://www.regeringen.se/contentassets/6eef64a9a36e48ff9a95e4d6ad97ce84/aid-policy-framework [2015.11.15].

Tikly, L. (2015). What works, for whom, and in what circumstances? Towards a critical realist understanding of learning in international and comparative education. International Journal of Educational Development, 40, 237–249. http://doi.org/10.1016/j.ijedudev.2014.11.008

Tilcsik, András and Marquis, Christopher (2013). Punctuated Generosity: How Mega-events and Natural Disasters Affect Corporate Philanthropy in U.S. Communities. Administrative Science Quarterly 2013, 58(1): 111-148.

UNESCO. (2015). Out-of-School Children. January 19. http://www.uis.unesco.org/Education/Pages/out-of-school-children.aspx.

Weaver, L., and J. B. Cousins. (2004). Unpacking the participatory process. Journal of Multidisciplinary Evaluation. 1: 19-40.

Weiss, C. H. (1983). The stakeholder approach to evaluation: Origins and promise. New Directions for Program Evaluation. London: Falconer Press.

Westhorp, G., Walker, B., & Rogers, P. (2012). Protocol - Under what circumstances does enhancing community accountability and empowerment improve education outcomes, particularly for the poor? http://r4d.dfid.gov.uk/Output/191758/Default.aspx

White, Howard. (2007). Evaluating Aid Impact. MPRA Paper 6716. University Library of Munich, Germany. https://ideas.repec.org/p/pra/mprapa/6716.html.

White, H. (2009). Some Reflections On Current Debates In Impact Evaluation (3ie Publications No. 2009-1). International Initiative for Impact Evaluation (3ie). https://ideas.repec.org/p/ris/iiierp/2009_001.html

World Bank. (1995). Priorities and Strategies for Education: A World Bank Review Washington: World Bank.

World Bank. (2011). Learning for All: Investing in People's Knowledge and Skills to Promote Development. World Bank Education Strategy 2020. Washington: World Bank.

7. Annexes: contents

A. List of evaluations reviewed

A	Agency	Title	Year Published	In-depth review?	In Text Citation
1	3ie	Quality education for all children? What works in education in developing countries	2013	YES	(Eval: 3ie 2013)
2	African Development Bank	Morocco: Evaluation of Bank Assistance to the Education Sector	2005	NO	(Eval: African Development Bank 2005)
3	Aga Khan Foundation	Educational Development and Improvement Programme	2013	NO	(Eval: Aga Khan 2013)
4	Agence Francaise de Developpement (AFD)	Case study of Aid to Education in Mauritania	2008	NO	(Eval: AFD 2008)
5	Agence Francaise de Developpement (AFD), Denmark Development Cooperation (DANIDA), and Benin Ministry of Development, Economis Analysis and Forecasting (MCPD)	Evaluation a mi-parcours du Plan decennal de developpement du secteur de l'education du Benin (PDDSE 2006-2015)	2012	CASE STUDY	(Eval: AFD/ DANIDA/MCPD 2012)

6	French Ministry of Foreign Affairs	La Cooperation Francaise face aux defis de l'education en Afrique: l'urgence d'une nouvelle dynamique	2007	YES	(Eval: French Ministry of Foreign Affairs 2007)
7	Agence Francaise de Developpement (AFD) and World Bank (WB)	L'enseignement post-primaire en Afrique subsaharienne Viabilite financiere des differentes options de developpement	2010	YES	(Eval: AFD & WB 2010)
8	Asian Development Bank	Education Sector in Bangladesh: What Worked Well and Why under the Sector-Wide Approach?	2008	NO	(Eval: Asian Development Bank 2008)
9	Asian Development Bank	Evaluation of Education Sector – Uzbekistan	2010	YES	(Eval: Asian Development Bank 2010)
10	Australian Agency for International Development (AusAID)	Independent Evaluation of Australia Indonesia Basic Education Program (AIBEP)	2010	NO	(Eval: AusAID 2010)
11	Australian Agency for International Development (AusAID) & UNICEF	AusAID Education Initiatives in Aceh, Papua and Papua Barat	2012	NO	(Eval: AusAid/UNICEF 2012)
12	Belgian Development Cooperation (BTC)	Thematic evaluation of Belgian development co-operation in the education sector	2007	YES	(Eval: BTC 2007)
13	CfBT Education Trust	The impact of sector-wide approaches: where from, where now and where to?	2011	YES	(Eval: CfBT 2011
14	Concern Worldwide	Education in Manica, Mozambique	2009	NO	(Eval: Concern 2009)

15	Concern Worldwide	Education support programme in Niger	2010	NO	(Eval: Concern 2010)
16	Concern Worldwide	Amader school project in Bangladesh: evaluation	2012	NO	(Eval: Concern 2012a)
17	Concern Worldwide	Burundi education programme: evaluation	2012	NO	(Eval: Concern 2012b)
18	Concern Worldwide	Promising practice in school-related gender-based violence prevention and response programming globally	2013	YES	(Eval: Concern 2012c)
19	Conn, K.	Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations	2014	YES	(Eval: Conn 2014)
20	Denmark Development Cooperation (DANIDA) (lead), Canadian International Development Agency (CIDA) and United Nations Children's Fund (UNICEF)	Multifaceted Challenges: A study on the barriers to girls' education - Zambezia Province	2005	NO	(Eval: DANIDA et al. 2005)
21	Department for International Development UK (DFID)	DFID's Education Programmes in Three East African Countries	2012	YES	(Eval: DFID 2012)
22	Department for International Development UK (DFID)	Evaluation of Results Based Aid in Rwandan Education - Year 1	2014	YES	(Eval: DFID 2014)

23	Department for International Development UK (DFID) & Institute of Education, University of London	A rigorous review of the political economy of education systems in developing countries	2014	YES	(Eval: DFID & IOE 2014)
24	Department for International Development UK (DFID) & University of Sussex	Pedagogy, Curriculum, Teaching Practices and Teacher Education in Developing Countries	2013	YES	(Eval: DFID & US 2013)
25	Department for International Development UK (DFID) and partners	What works to improve teacher attendance in developing countries? A systematic review on what works to improve teacher attendance in developing countries	2012	YES	(Eval: DFID 2012)
26	Department for International Development UK (DFID) and partners	Literacy, Foundation Learning and Assessment in Developing Countries	2014	YES	(Eval: DFID 2014a)
27	Department for International Development UK (DFID) and partners	The role and impact of private schools in developing countries	2014	YES	(Eval: DFID 2014b)

28	Department for International Development UK (DFID), Upper Quartile Consulting	Evaluation of Results Based Aid in Rwandan Education - Year 2 Report	2015	YES	(Eval: DFID 2015)
29	European Commission (EC)	Thematic global evaluation of European Commission support to the education sector in partner countries (including basic and secondary education)	2010	YES	(Eval: EC 2010)
30	German Technical Cooperation Agency (GIZ)	Promotion de l'éducation de base	2012	YES	(Eval: GIZ 2012)
31	German Technical Cooperation Agency (GIZ)	Cross Section Analysis of the Education Sector: Meta-Evaluation and Synthesis	2014	NO	(Eval: GIZ 2014)
32	German Technical Cooperation Agency (GIZ) & Reconstruction Credit Institute (KfW)	Ex-post evaluation 2012 – Brief Report Basic Education in Namibia	2012	NO	(Eval: GIZ & KfW 2012)
33	Inter-American Development Bank (IDB)	Review of IDB Support to Secondary Education: Improving Access, Quality and Institutions, 1995-2012	2013	YES	(Eval: IDB 2013)
34	Inter-American Development Bank (IDB) & Japan International Cooperation Agency (JICA)	Leading the Way to Math and Science Success: Challenges and Triumphs in Paraguay: New research from the Inter-American Development Bank on the promotion of critical thinking in preprimary and primary education	2005	NO	(Eval: IDB & JICA 2005)

35	IrishAID	Country Strategy Paper 2007-2010: Zambia (Evaluation)	2010	YES	(Eval: IrishAID 2010)
36	Korea International Cooperation Agency (KOICA)	Ex-Post Evaluation Report on the Project of the Construction and Extension for 18 Primary Schools in Nairobi, Nakuru, and Thika (Kenya)	2012	NO	(Eval: KOICA 2012a)
37	Korea International Cooperation Agency (KOICA) and World Friends Korea	Ex-Post Evaluation Report on the Two Primary and Secondary Education Projects in Palestine	2012	NO	(Eval: KOICA 2012b)
38	Mathematica, for USAID	Impact Evaluation of Burkina Faso's BRIGHT Program: Final Report	2009	YES	(Eval: Mathematica 2009)
39	McEwan, P.	Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments	2014	NO	(Eval: McEwan 2014)
40	MiET Africa, Swiss Agency for Development and Cooperation (SDC) and the Embassy of the Kingdom of the Netherlands (EKN)	Schools as Centres of Care and Support (SCCS) Responding to the Needs of Orphans and Other Vulnerable Children in Rural Areas	2010	NO	(Eval: MiET Africa, SDC, & EKN 2010)
41	Netherlands Ministry of Foreign Affairs	Primary Education in Zambia	2005	NO	(Eval: Netherlands Ministry of Foreign Affairs 2005)

42	Netherlands Ministry of Foreign Affairs	Primary Education in Uganda	2008	NO	(Eval: Netherlands Ministry of Foreign Affairs 2008)
43	Netherlands Ministry of Foreign Affairs	Policy review of the Dutch contribution to basic education 1999–2009	2011	NO	(Eval: Netherlands Ministry of Foreign Affairs 2011a)
44	Netherlands Ministry of Foreign Affairs	The two-pronged approach: Evaluation of Netherlands support to primary education in Bangladesh	2011	YES	(Eval: Netherlands Ministry of Foreign Affairs 2011b)
45	Norweigain Agency for Development Cooperation (Norad)	Joint Evaluation of Nepal's Education for All 2004-2009 Sector Programme	2009	CASE STUDY	(Eval: Norad 2009)
46	Norweigian Refugee Council (NRC), Save the Children, European Comission, Concern Worldwide	Accelerated Primary Education Support program in Somalia	2012	NO	(Eval: NRC 2012)

47	Organisation internationale de la Francophonie (OIF), (AFD) Ministère des affaires étrangères et européennes (MAEE) Agence universitaire de la Francophonie (AUF)	Les langues de scolarisation en Afrique francophone	2010	NO	(Eval: OIF et al 2010)
48	Riddell, A.	The Effectiveness of Foreign Aid to Education: What can be learned?	2012	NO	(Eval: Riddell 2012)
49	Room to Read	School Libraries Cross-National Evaluation	2015	YES	(Eval: Room to Read 2015)
50	RTI International	Implementing School-Based Management in Indonesia	2011	YES	(Eval: RTI 2011)
51	Save the Children	Literacy Boost Malawi: Year 2 Report	2011	NO	(Eval: SC 2011)
52	Save the Children (SC)	Mid-term evaluation of the Inclusive Quality Pre-Primary and Primary Education for Roma/Egyptian Children Project	2011	NO	(Eval: SC 2011)
53	Swedish International Development Cooperation (Sida)	Swedish Support to the Education Sector in Mozambique	2004	YES	(Eval: Sida 2004)

54	Swedish International Development Cooperation (Sida)	Evaluation and Monitoring of Poverty Reduction Strategies – 2005- Budgeting for Education: Bolivia, Honduras and Nicaragua	2005	YES	(Eval: Sida 2005)
55	Swedish International Development Cooperation (Sida)	Sida's contributions 2006: Progress in educational development	2007	YES	(Eval: Sida 2007a)
56	Swedish International Development Cooperation (Sida)	Swedish Support in the Education Sector in Zanzibar, 2002–2007	2007	YES	(Eval: Sida 2007b)
57	Swedish International Development Cooperation (Sida)	Are Sida Evaluations Good Enough? An Assessment of 34 Evaluation Reports	2008	NO	(Eval: Sida 2008a)
58	Swedish International Development Cooperation (Sida)	Policy Guidance and Results-Based Management of Sida's Educational Support	2008	YES	(Eval: Sida 2008b)
59	Swedish International Development Cooperation (Sida)	Gender equality in and through education	2010	NO	(Eval: Sida 2010)
60	Swedish International Development Cooperation (Sida)	Review of the Sida-funded Project Education for Sustainable Development in Action (ESDA) - Final Report	2012	NO	(Eval: Sida 2012)

61	Swedish International Development Cooperation (Sida)	Evaluation of the Barbro Johansson Model Girls' Secondary School in Tanzania	2013	NO	(Eval: Sida 2013a)
62	Swedish International Development Cooperation (Sida)	Swedish Development Cooperation in Transition? Lessons and Reflections from 71 Sida Decentralised Evaluations	2013	YES	(Eval: Sida 2013b)
63	Swedish International Development Cooperation (Sida)	Evaluation of Implementation of ICT in Teachers' Colleges Project in Tanzania - Final Report	2014	CASE STUDY	(Eval: Sida 2014a)
64	Swedish International Development Cooperation (Sida)	Lessons and Reflections from 84 Sida Decentralised Evaluations 2013 – a Synthesis Review	2014	YES	(Eval: Sida 2014b)
65	United Nations Children's Fund (UNICEF)	Child-Friendly Schools Programming: Global Report	2009	NO	(Eval: UNICEF 2009)
66	United Nations Children's Fund (UNICEF)	Government of Tanzania/ UNICEF 7 Learning Districts Strategy (2007-2011)	2010	YES	(Eval: UNICEF 2010)
67	United Nations Children's Fund (UNICEF)	2012 Democratic Republic of Congo: Evaluation du programme Ecole et Village Assainis	2011	YES	(Eval: UNICEF 2011)
68	United Nations Children's Fund (UNICEF)	2012 Gambia: Evaluation of the Girls Education Project of the Forum for African Women Educationalists - The Gambia (FAWEGAM)	2012	NO	(Eval: UNICEF 2012a)

69	United Nations Children's Fund (UNICEF)	2012 Sierra Leone: Evaluation of UNICEF role as a Lead Partner in Education	2012	YES	(Eval: UNICEF 2012b)
70	United Nations Children's Fund (UNICEF)	Evaluation of the Girls Education Project of the Forum for African Women Educationalists – The Gambia (FAWEGAM)	2012	NO	(Eval: UNICEF 2012c)
71	United Nations Children's Fund (UNICEF)	External Evaluation of the "Child-Friendly School" Initiative (2007-2011) in the Republic of Croatia	2012	NO	(Eval: UNICEF 2012d)
72	United Nations Children's Fund (UNICEF)	Independent Evaluation of Program:Improving Access to Quality Basic Education in Myanmar (2006-2010)	2012	YES	(Eval: UNICEF 2012e)
73	United Nations Children's Fund (UNICEF) and partners	Process and Impact Evaluation of the Basic Education Assistance Module (BEAM) in Zimbabwe	2013	NO	(Eval: UNICEF 2013)
74	United Nations Children's Fund (UNICEF), Save the Children (SC), Department for International Development UK (DFID)	Developing a Local Model for The Delivery of Primary Education in Karkaar Region, Puntland	2012	NO	(Eval: UNICEF, SC & DFID 2012)
75	United States Agency for International Development (USAID)	Assessment of the USAID Assistance Program to the Reform of the Benin Primary Education System	2005	YES	(Eval: USAID 2005)

76	United States Agency for International Development (USAID)	Program Evaluation for USAID - Guinea Basic Education Program Portfolio	2006	YES	(Eval: USAID 2006)
77	United States Agency for International Development (USAID) & World Learning	Action Communautaire pour l'education des filles: Evaluation finale (2001-2005)	2006	YES	(Eval: USAID & World Learning 2006)
78	World Bank (WB)	An Unfinished Agenda: An Evaluation of World Bank Support to Primary Education	2005	YES	(Eval: WB 2005)
79	World Bank (WB)	Bangladesh Education Sector Review Seeding Fertile Ground: Education That Works for Bangladesh	2013	NO	(Eval: WB 2013)
80	World Bank (WB)	What Really Works to Improve Learning in Developing Countries?	2015	YES	(Eval: WB 2015)

B. On evaluations

Here, we present an overview of the 80 evaluations included in our synthesis according to the following dimensions: agency/author type, country (or countries), approach, and activities evaluated.

Agency/author type

This refers to the agency or organization that funded the project under evaluation. We distinguish between the following types of aid providers: bilateral, multilateral, non-profit/foundation, and UNICEF. We also include evaluations conducted by bi-laterals or multi-laterals in partnership with each other or with non-profit organizations, such as the evaluation of aid to basic education in Indonesia led by AusAID and UNICEF (Eval: AusAid & UNICEF 2012). In most cases, the author of these evaluations is an external consultant (or group of consultants), rather than the aid agency itself. In addition, we include 6 studies that are academic in nature, either conducted by a research institute or an individual.

Agency/author type	
Multi-lateral	6
Bi-lateral	33
Non-profit/foundation	8
UNICEF	8
Partnership	19
Academic researcher /research institute	6
TOTAL	80

Country/countries

The majority of the evaluations we reviewed focused on a single country (61 percent). 30 percent evaluate education programs across multiple countries. Of the evaluations that focus on a single country, most are in the African continent, which is not surprising considering that this is the region that receives the greatest proportion of foreign aid to education (27 percent of bilateral and multilateral commitments in 2012-13) (OECD 2015).

	Country
Multi-country (global)	24
Multi-country (African continent)	5
Multi-country (Latin American and Caribbean)	2
Multi-country (Asian continent)	0
Albania	1
Bangladesh	4
Benin	1
Benin	2
Burkina Faso	1
Burundi	1
Chad	1
Croatia	1
Democratic Republic of Congo	1
Ethiopia	1
Guinea	1
Indonesia	3
Kenya	1
Malawi	1
Mauritania	1
Morocco	1
Mozambique	3
Myanmar	1
Namibia	1
Nepal	1
Niger	1
Pakistan	1
Palestine	1
Paraguay	1
Rwanda	2
Sierra Leone	1
Somalia	3
South Africa	1
Tanzania	2
The Gambia	1
Uganda	1
Ukraine	1
Uzbekistan	1
Zambia	2
Zanzibar	1

Zimbabwe	1
TOTAL	80

Approach

The vast majority (63 percent) of the evaluations reviewed are descriptive in nature, consisting of a desk review of policy and project documents, interviews with key actors (aid officials, Ministry of Education officials, district and local level education officials, and, in some cases, teachers, families, and students), classroom observations, analyses of administrative data (trends in enrollment rates over time, for example), and in some cases, cross-sectional surveys of program participants (teachers, families, students). We reviewed two meta-analyses that seek to compare the pooled effect sizes of different projects, and two synthesis reviews of quantitative impact evaluations (experimental and quasi-experimental impact assessments). In addition, our review includes 12 syntheses that incorporate multiple qualitative and quantitative studies related to specific themes, such as literacy development in low-income countries (Eval: DFID and partners 2014), or "what works" to keep teachers in classrooms, for example (Eval: DFID and partners 2012). Eight of the evaluations we reviewed are participatory evaluations (meaning that program participants played a leading role in the evaluation design and analysis, rather than just serving as interview subjects). Six of the evaluations we reviewed are impact evaluations (e.g., evaluations that use an experimental or quasi-experimental method to estimate a quantitative impact on educational outcomes). The fact that the majority of the evaluations we reviewed are descriptive is not surprising, given our focus on evaluations, rather than academic literature.

	Approach
Meta-analysis (quantitative)	2
Synthesis (quantitative)	2
Synthesis (quantitative and qualitative)	12
Descriptive (quantitative and/or qualitative)	50
Participatory	8
Impact evaluation	6
TOTAL	80

Activities evaluated

The majority of the studies in our synthesis are either evaluations of sector-wide support to education in a single country, such as the Asian Development Bank's support to education in Uzbekistan, for example (Eval: ADB 2010), or evaluations of an individual aid-funded education project, such as Sida's support to the development of ICT in Teachers' Colleges in Tanzania (Eval: Sida 2014a). Seven studies focus on aid management, evaluating donor collaboration for education in Sector Wide Approaches (Eval: CfBT 2011), for example, or reviewing Sida's evaluation practices (Eval: Sida 2013b; Eval: Sida 2014b). Studies that focus on topics related to educational development in low-income countries, rather than evaluations *per se*, are classified as "other." These include a study on the barriers to girls' education in Mozambique, for example (Eval: Danida and partners 2005), an review of the political economy of education systems (Eval: DFID and IOE 2014), and an overview of pedagogy, curriculum and teaching process in developing countries (Eval: DFID and US 2013), among others.

Activities evaluated	
Sector-wide support to education in a single country	27
An individual aid-funded project	20
An individual agency's support to education globally	6
Multiple projects across multiple countries	12
Aid management	7
Other	8
TOTAL	80

C. Selection strategy

We seek to to build on the reviews and syntheses completed to date through a cross-disciplinary, multi-modal, and multi-layered approach. To do so, rather than relying on academic research, we focus on evaluations conducted by and for those who are directly involved in the aid relationship, since these are the evaluations that are expected to be directly linked to changes in practices and policymaking. From this, it follows that the ultimate value of an evaluation depends on the extent to which it enables funding agencies, governments, education officials and educators to improve their practices. Thus, where possible, we explore how different constituencies, from funding agencies, to implementing organizations, and aid recipients, *use* evaluations.

Our selection strategy begin with a comprehensive search of evaluations of education activities commissioned by the following international and national agencies and organizations:

Agency/Organization

3ie

Aga Khan Foundation

Agence Francaise de Developpement (AFD)

Asian Development Bank

African Development Bank

Australian Agency for International Development (AusAID)

Belgian Development Cooperation

CfBT Education Trust

Canadian International Development Agency (CIDA)

Concern World Wide

Danish International Development Agency (Danida)

Department for International Development UK (DFID)

Education for Change

Embassy of the Kingdom of the Netherlands (EKN)

European Commission

French Ministry of Foreign Affairs

German Technical Cooperation Agency (GIZ)

Inter-American Development Bank

IrishAID

Japan International Cooperation Agency (JICA)

Korea International Cooperation Agency (KOICA)

Mathematica

MiET Africa

Netherlands Ministry of Foreign Affairs

Norwegian Agency for Development Cooperation (Norad)

Norweigian Refugee Council (NRC)

OECD Development Directorate

Organisation Internationale de la Francophonie (OIF)

Room to Read

RTI International

Save the Children

South Research

Swedish Agency for International Develoment (Sida)

Swiss Agency for Development and Cooperation (SDC)

United Nations Children's Fund (UNICEF)

United States Agency for International Development (USAID)

World Bank

World Friends Korea

World Learning

Cognizant of the need to be as inclusive as possible across a wide range of aid providers, no exclusion criteria were applied at this phase beyond the requirements that the evaluation focus on education activities that were at least in part aid-funded, that the evaluation be published after 2005, and that the evaluation be written in English, French, or Spanish, and that the full digital report be publically accessible. This resulted in an initial list of 80 evaluations.

Our selection and review process is closely aligned with realist synthesis, a methodology designed to explore complex and varied programs applied across multiple contexts (Greenhalgh, Wong, Westhorp, & Pawson, 2011; Pawson, 2002; Westhorp, Walker, & Rogers, 2012). The objective of a realist synthesis is to achieve depth of understanding, exploring context, mechanisms, and processes that lead to outcomes and impact, rather than producing a verdict on a program's effectiveness. To do so, realist synthesis draw from a diverse group of purposively selected studies, selected based on two main criteria: (1) relevance (to the theories or concepts under exploration), and (2) rigor. Importantly, rigor refers to the adequacy and appropriateness of the methods used in relation to the context, interactions, and processes under study, rather than to the evaluation's internal or external validity, *per se*. Our synthesis draws on these criteria and adds a third: diversity. We therefore modified and added to the sub-set of evaluations selected for in-depth review in order to ensure that the studies that we gave most attention to adequately reflect the diversity of funders, implementers, programs evaluated, contexts, and methodological approaches present across the 90 evaluations we initially identified.

In order to select the sub-group of evaluations for in-depth review, we applied a common list of dimensions and assessment criteria that we used to select and classify evaluations. We classified evaluations as strong/moderate/weak across these dimensions and used these classifications to guide our selection process, not as strict inclusion/exclusion criteria.

The template is as follows:

Item	Description	Strong/Moderate/Weak
1. Relevance	Attention to the relationship between aid funded basic education programs and educational processes or outcomes Identifiable theory of change	
2. Program description	Program objective Activities NOTE: The rating in this case refers to the quality of the description of the the program objective and activities (strong, moderate, weak)	
3. Evaluation objective	Study identifies an evaluation objective or research question	
4. Approach	Study identifies a research design and method(s) including (but not limited to) one or more of the following: - Quantitative (descriptive, experimental, quasi-experimental) - Qualitative (document analysis, interviews, observation, focus groups) - Utility focused evaluation - Participatory evaluation - Meta-analysis - Narrative synthesis	
5. Rigor	Methods, measures, and analysis are appropriate for the relationship(s) or causal mechanism(s) under study Limitations are acknowledged	
6. Target audience	Does the evaluation identify a target audience(s)/constituency(ies)? - If yes, who?	
7. Participatory evaluation	Were there participants other than the formal evaluators? - If yes, who? What role did they play in the evaluation?	
8. Explicit assessment of process	Attention to program implementation, including internal and external factors influencing program participants and activities	
9. Explicit assessment of outcomes	Attention to program outcomes (short-term, medium-term, and/or long-term, related to participation, learning, teaching, and/or institutional change)	

10. External quality measure	Mixed Method Appraisal Tool (Pluye 2011) The MMAT is designed to classify and select sources for systematic reviews that include qualitative, quantitative and mixed methods.	
11. Activities evaluated	Study includes a description of program objectives and activities evaluated, geographic location, and intended aid recipients/program participants	
12. Lessons learned re: education, aid to education, evaluations	What do we learn from this evaluation regarding (1) education, (2) aid to education, and/or (3) evaluations	
13. Utility	In what other ways might this evaluation be useful (even if it is not useful for our synthesis)?	
14. Additional aspects that make the study worthy of inclusion		

As we used common selection criteria to determine which evaluations warranted fuller examination, we were at the same time attentive to the ways in which the selection process itself may constrain or specify the eventual findings. To that end, the evaluations selected for detailed review score high on most of the quality dimensions described above, but not necessarily all. This enables us to explore directly the complexity of the relationship among education, aid, and evaluation. 40 evaluations were selected for in-depth review.

Finally, from among the evaluations selected for in-depth review, we identified 3 studies for case study analyses, based primarily on feasibility for case study analysis. That is, we chose evaluations for which we were confident in our ability to establish direct contact with the aid agencies, implementing partners, and aid recipients involved.

D. Summary reviews

Title	Quality education for all children?	
Author/Agency	Krishnarane, S., White, H., Carpenter, E. 3ie	
Date published	September 2013	
Item	Description	Strong/Moderate/Weak
1. Relevance	Meta-analysis - systematic review – although focused on specific programs, not the aid relationship	Strong
2. Program description	Five areas: (1) reducing costs, (2) increasing preparedness, (3) providing information, (4) supply-side interventions	Strong
3. Evaluation objective	Identify "what works" in getting children into school in developing countries, keeping them there, and ensuring they learn whilst there	Strong
4. Approach	- Selection process based on studies w/RCT or quasi-experimental causal inference, with quantifiable outcome measures, from 1990 – 2009	Strong
5. Rigor	Rigor of the studies chosen was strong, harder to discern the rigor of the systematic review	Moderate
6. Target audience	Researchers, policy makers	Moderate
7. Participatory evaluation	No	Weak
8. Explicit assessment of process	No	Weak
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure		Strong
11. Activities evaluated	Multiple studies from abovementioned themes	Strong
12. Lessons learned re: education, aid to education, evaluations	Provides broad claims based on RCTs and quasi-experimental studies regarding certain types of educational initiatives. Is similar to a standard meta-analysis	
13. Utility		
14. Additional aspects that make the study worthy of inclusion	I think it's worth mentioning "what works" according to prominent meta-analyses, systematic reviews, such as this one –and then contrasting our synthesis approach and findings.	

Title	Morocco: Evaluation of Bank Assistance to the Education Sector	
Author/Agency		African Development Bank
Date published		2005
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of aid-supported sector-wide education support	Strong
2. Program description	The Bank supported "3 priority pillars" – basic education, a skilled labor force, and institutional development.	Strong
3. Evaluation objective	Review the ADB's assistance to Morocco's education sector, focusing on the consistency of the Bank's policies and strategies with those of the Moroccan authorities.	Strong
4. Approach	Mostly document analysis – covering interventions from 1994-2004, also includes interviews with Moroccan authorities and school visits.	Moderate
5. Rigor	Methods are not thoroughly described, not replicable, limitations not acknowledged. It's not clear how conclusions are reached	Weak
6. Target audience	Policy makers, Bank staff	Strong
7. Participatory evaluation	No	Weak
8. Explicit assessment of process	Not really – some discussion of implementation challenges	Moderate
9. Explicit assessment of outcomes	Not really – attribution challenges not clearly acknowledged	Weal
10. External quality measure		Weak
11. Activities evaluated	School construction, some technical assistance for institutional development	
12. Additional aspects that make the study worthy of inclusion		

Title	Educational Development and Improvement Programme	
Author/Agency		Aga Khan Foundation
Date published		February 2013
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of 3 year project – funded by AusAid, implemented by Agha Khan: aims to enhance access, equity, and quality w/increased gender parity through a Whole School Improvement Approach	Strong
2. Program description	Very comprehensive program – includes infrastructure investments, community mobilization for girls and children with disabilities, government capacity development, activities are not clearly described	Moderate
3. Evaluation objective	Evaluate project across the following criteria: relevance, effectiveness, efficiency, gender equality, monitoring and evaluation, sustainability	Moderate
4. Approach	Primarily qualitative – meetings with stakeholders, programme staff, review of school registers, lesson plans, student notebooks, SMC meeting minutes	Moderate
5. Rigor	Limited discussion of methods – findings not clearly linked to data analyzed	Weak
6. Target audience	Agha Khan	
7. Participatory evaluation	"Stake holder interviews"	Weak
8. Explicit assessment of process	Yes – although still mostly inputs/outputs – with some discussion of teachers' and students', officials' opinions/perceptions, changes in attitudes	Moderate
9. Explicit assessment of outcomes	Yes – but no attempt to address attribution issue	Weak
10. External quality measure (MMAT)		Weak
11. Activities evaluated	Unclear	Weak
12. Lessons learned re: education, aid to education, evaluations		

13. Utility	
14. Additional aspects that make the study worthy of inclusion	May be worth including because Agha Khan is a growing player in the field of "non-traditional donors" – and Pakistan is also a priority country

Title	**Case study of Aid to Education in Mauritania**	
Author/Agency		AFD
Date published		July 2008
Item	Description	Strong/Moderate/Weak
1. Relevance	very quick overview	Strong
2. Program description	Aid to education in Mauritania (construction, equipment, training, capacity building, and evaluation)	Moderate
3. Evaluation objective	Case study to see if another evaluation is needed	Weak
4. Approach	Descriptive, qualitative literature review	Weak
5. Rigor	weak	weak
6. Target audience	policymakers	weak
7. Participatory evaluation	no	weak
8. Explicit assessment of process	N/A	weak
9. Explicit assessment of outcomes	Somewhat—"progress still needs to be made", need an overall evaluation as follow-up and improvements on mutual accountability	weak
10. External quality measure		weak
11. Activities evaluated	listed above	weak
12. Additional aspects that make the study worthy of inclusion	not sufficient for inclusion –just an overview and description of education aid to Mauritania and ways forward, rather short document.	

Title	L'enseignement post-primaire en Afrique subsaharienne: Viabilite financiere des differentes options de developpement[3]	
Author/Agency	AFD/ World Bank	
Date published	2010	
Item	Description	Strong/Moderate/Weak
1. Relevance	a comparative analysis of post-primary education in 33 low-income Sub-Saharan countries, also including examples from middle-income countries in other regions	Strong
2. Program description	The AFD and the World Bank worked on this study together to acquire a comparative perspective	Strong
3. Evaluation objective	Examining ways to finance post-primary education in Africa	Strong
4. Approach	Utility focused evaluation	Moderate
5. Rigor	more descriptive, less rigorous, but adaptation to each national context given the diversity of countries involved	Moderate
6. Target audience	policymakers	
7. Participatory evaluation	simulation model directed towards national leaders and their development partners, to help influence policy decisions particularly for post-primary education	Weak
8. Explicit assessment of process		Moderate
9. Explicit assessment of outcomes		Strong
10. External quality measure	CBA, includes simulations	Weak
11. Activities evaluated	CBA post-primary education	Moderate
12. Additional aspects that make the study worthy of inclusion	quantitative study that emphasizes adaptation to national contexts	

[3] All direct quotes and translations from documents originally in French are my translation, here and elsewhere in this report.

Title	La Cooperation Française face aux defis de l'éducation en Afrique: l'urgence d'une nouvelle dynamique	
Author/Agency	AFD/ French Ministry of Foreign Affairs	
Date published	2005	
Item	Description	Strong/Moderate/Weak
1. Relevance	To consider various strategy options for more efficiency and coherence in French aid to education.	Strong
2. Program description	The overall context of French aid to education	Strong
3. Evaluation objective	Evaluating French aid to education and calling for a renewed approach	Strong
4. Approach	Policy document	Moderate
5. Rigor		Moderate/weak
6. Target audience	Policymakers	
7. Participatory evaluation	(national stakeholders but does not mention local), however this is proposed	Weak
8. Explicit assessment of process		Strong
9. Explicit assessment of outcomes	The evaluation found that decentralized approaches and the Pole de Dakar (sectoral analysis center) do not quite address educational quality, and therefore proposes the creation of a "Pole Qualité" or basically, a center focused on quality that will be a home base for resources, a place for exchange and collaboration, in teaching and learning, providing teaching training and addressing the school environment. The evaluation maintains that the "Quality Center" will facilitate South-South cooperation and take into account national differences between each country, and will diffuse and share tools, experiences and best practices.	Strong
10. External quality measure	interesting document in terms of strategy, not much by way of methods, CBA	Moderate
11. Activities evaluated	French education cooperation strategy	Strong
12. Additional aspects that make the study worthy of inclusion	The evaluators maintain that there is a « new dynamic in international engagement » but still it is too little in terms of the education sector.	

Title	**Education Sector in Bangladesh: What Worked Well and Why under the Sector-Wide Approach**	
Author/Agency	Operations Evaluation Department – Asian Development Bank	
Date published	December, 2008	
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of development cooperation in Bangladesh education sector	Strong
2. Program description		
3. Evaluation objective	Objective: assess combined performance of ADB, DFID, World Bank, and JICA in SWAp –explain what worked well, what did not, and why, to inform future education development cooperation strategy (1989 – 2007 period)	Strong
4. Approach	Top-down: strategic and institutional, and Bottom-up: operational and implementation	Strong
5. Rigor	Attention to historical/institutional analysis	Strong
6. Target audience	ADB and other multi-lateral /bi-lateral agencies	Strong
7. Participatory evaluation		Weak
8. Explicit assessment of process	More attention to overall policy change, planning and coordination, alignment with national development	Moderate
9. Explicit assessment of outcomes		Weak
10. External quality measure		Weak
11. Activities evaluated	Sector wide – report includes description of the project components of all loans from period under study	Strong
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		

14. Additional aspects that make the study worthy of inclusion	

Title	**Uzbekistan: Education Sector Assistance Program Evaluation**	
Author/Agency	Independent Evaluation Department – Asian Development Bank	
Date published	September 2010	
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of aid to education (sector wide)	Strong
2. Program description	Objective: improve access to quality of basic education in Uzbekistan	Weak
3. Evaluation objective	Objective: assess performance of ADB assistance in the education sector of Uzbekistan from 1997 – 2009, identify factors affecting performance, draw lessons and recommendations to feed preparation of future programming	Strong
4. Approach	Top-down: strategic and institutional, and Bottom-up: operational – Both mostly through project documents, evaluation reports, and "information generated by fieldwork," evaluation culminated in a workshop	Moderate
5. Rigor	Attention to historical/institutional analysis	Moderate
6. Target audience	Aid officials – ADB	Moderate
7. Participatory evaluation	Limited – draws on previous evaluations that included	Weak

	focus group and interviews	
8. Explicit assessment of process		Strong
9. Explicit assessment of outcomes		Weak
10. External quality measure		Weak
11. Activities evaluated	Evaluation covers all ADB funded programs in the country: curriculum development, planning and coordination, school management & community participation, support for NGO provision of education, among others.	
12. Lessons learned re: education, aid to education, evaluations	Includes sections on project evaluation and technical assistance evaluation	
13. Utility		
14. Additional aspects that make the study worthy of inclusion	Well organized and written, diversity in terms of funding agency and aid recipient country. Would provide nice contrast to other sector-wide evaluations.	

Title	Australia Indonesia Basic Education Program (AIBEP)	
Author/Agency		AusAID
Date published		2010
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of Australia's aid to basic education in Indonesia from 2006 – 2010, objective is to "support the Government of Indonesia in improving equitable access to higher quality and better governed basic education services in targeted, disadvantaged areas"	Strong
2. Program description	Loans and grants for school construction, district capacity development, policy advice and institutional/organizational development	Strong
3. Evaluation objective	Independent completion report – assess AusAid's educational support to Indonesia against effectiveness, efficiency, impact and sustainability.	Strong
4. Approach	Qualitative - Literature review, analysis of primary/secondary data (program reports, trends in GER/NER), semi-structured interviews with stakeholders, field visits to schools and district offices	Moderate
5. Rigor	Limited – it is a completion report, not an ex-post evaluation. No comparisons, no longitudinal, limited link from findings to conclusions.	Weak
6. Target audience	AusAid, government of Indonesia.	Strong
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	Somewhat – description of perceptions and attitudes, challenges encountered in implementation	Moderate
9. Explicit assessment of outcomes	Yes, but attribution is not addressed	Weak
10. External quality measure	Limitations briefly discussed, no clear link between data sources and findings/recommendations	Weak
11. Activities evaluated	School construction, capacity development projects	Moderate
12. Lessons learned re: education, aid to education,		

evaluations	
13. Utility	Mostly for AusAid, not necessarily governmet
14. Additional aspects that make the study worthy of inclusion	Diversity – AusAid and Indonesia

Title	AusAID Education Initiatives in Aceh, Papua and Papua Barat	
Author/Agency		AusAid /UNICEF
Date published		March 2012
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of assistance to education technical support to Papua (Indonesia)	Strong
2. Program description	Goal is to improve quality of primary education – through strengthened education planning, teaching practices and school management	Strong
3. Evaluation objective	Evaluate the program against effectiveness, efficiency and sustainability	Moderate
4. Approach	Participatory and formative (which in practice, in this case, means focus groups and interviews)	Moderate
5. Rigor	Data and selection methods clearly described, limitations defined	Moderate
6. Target audience	AusAID primarily, Government of Indonesia (national and district)	Strong
7. Participatory evaluation	In name, not necessarily practice	Moderate
8. Explicit assessment of process		Moderate
9. Explicit assessment of outcomes		Moderate
10. External quality measure (MMAT)		Moderate
11. Activities evaluated	Assistance to education offices (district) in strategic planning, support to improve teaching practices and school management	Strong
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	Thematic evaluation of Belgian development co-operation in the education sector	
Author/Agency	Education for Change/South Research	
Date published	August 2007	
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of aid to education – covers Belgian federally funded education and training programs between 2002 and 2006	Strong
2. Program description	Overview of Belgium's contributions to aid to education – case studies in Benin, Burundi, DR Congo, Ecuador, Tanzania, Vietnam	Moderate
3. Evaluation objective	Objective: improve the relevance of the Belgian Directorate General of Development Cooperation (DGDC)'s actions, inform new education strategy note	Strong
4. Approach	Document review, interviews w/policy actors, case studies	Strong
5. Rigor	Solid approach, but could be more information about strategies for case studies	Moderate
6. Target audience	BDGDC	Strong
7. Participatory evaluation	Yes – interviews w/aid officials and government officials in recipient countries	Moderate
8. Explicit assessment of process	Yes – in particular, assessment of Belgian cooperation's technical assistance, "managing by results," alignment, coordination, information transparency	Strong
9. Explicit assessment of outcomes	No	Weak
10. External quality measure		Weak
11. Activities evaluated	Overall Belgian development cooperation in education evaluated – direct bilateral aid, indirect aid	Moderate
12. Lessons learned re: education, aid to education, evaluations	Discussion of the separation of policy work form implementation responsibilities – see p. 14 Discussion of program and project implementation – in general VERY USEFUL evaluation. Should be read in depth.	
13. Utility		
14. Additional aspects that make the study worthy of inclusion	Describes the "policy architecture" within which Belgium aid to education operates – attention to multiple constituencies and multiple aspects of aid to education (e.g., donor coordination, methods of evaluation/monitoring)	

Title	The impact of sector-wide approaches: where from, where now and where to?	
Author/Agency	Boak, E., Ndaruhuts, S. for CfBT Education Trust	
Date published	2011	
Item	Description	Strong/Moderate/Weak
1. Relevance	Assessment of the role of sector wide approaches to education	Strong
2. Program description	Sector-wide approaches (SWAPs) to aid to education	Strong
3. Evaluation objective	Analyze the evolution of SWAps and their relationship with: (1) aid effectiveness, (2) planning and financing, (3) education outcomes, (4) fragility	Strong
4. Approach	Qualitative – literature review, interviews (telephone), written responses to questionnaires, some face to face interviews	Strong
5. Rigor		Strong
6. Target audience	Aid policy decision-makers, researchers	Strong
7. Participatory evaluation	Yes – aid practitioners, but not aid recipients	Moderate
8. Explicit assessment of process	Yes	Strong
9. Explicit assessment of outcomes	Yes	Moderate
10. External quality measure	Limited description of methods	Weak
11. Activities evaluated	SWAPs – lots of detailed information of particular countries' and agencies' experiences	Strong
12. Lessons learned re: education, aid to education, evaluations	- Report deals extensively with aid effectiveness, planning, financing, and outcomes. - Builds on previous evaluations of SWAPs	
13. Utility	- Aid policy decision makers, researchers, politicians in low-income countries	
14. Additional aspects that make the study worthy of inclusion		

Title	Education in Manica, Mozambique	
Author/Agency		Concern
Date published		2009
Item	Description	Strong/Mode-rate/Weak
1. Relevance		
2. Program description	The Concern Manica Education Project (2004 – 2008) aimed at improving access, quality and equity in primary education with a focus on girls and vulnerable children.	Strong
3. Evaluation objective	Did the project achieve its objective, and was it effective?	Strong
4. Approach	The method of the evaluation was participatory. Methods used for data collection were focus group discussions, semi-structured interviews and 'draw-and-write'. Secondary data was also gathered and data was triangulated across sources and methods.	Strong
5. Rigor	only one page	weak
6. Target audience	Capturing the learning from the evaluation to feed into the design of new National Education Programme was another central aim of the evaluation.	moderate
7. Participatory evaluation	Consulted project staff from three implementing partners (Magariro, ANDA and Concern) throughout the evaluation, in the design, the data collection and the analysis. The evaluation notes that two workshops were held: the first for evaluation design, and the second, for data collection and analysis. Consulted beneficiaries: school councils, teachers, school principals, parents, children, and district education authorities.	Strong
8. Explicit assessment of process		Moderate
9. Explicit assessment of outcomes	Additionally, took into account impact, relevance, efficiency and sustainability; and measured the extent to which intersectoral approaches of gender and HIV existed in the design and their implementation.	Moderate
10. External quality measure		Weak
11. Activities evaluated	access, quality and equality of primary education.	Strong
12. Additional aspects that make the study worthy of inclusion	strong focus on participatory methods	

Title	Education support programme in Niger	
Author/Agency		Concern
Date published		2010
Item	Description	Strong/Moderate/Weak
1. Relevance		Strong
2. Program description	A five-year project addressing quality primary education access, in a region in Niger (partial funding from the Human Dignity Foundation).	Moderate
3. Evaluation objective	program success	Strong
4. Approach	mixed methods	
5. Rigor		Weak
6. Target audience	Communities, NGOs, National partners	
7. Participatory evaluation	Highly participatory	Strong
8. Explicit assessment of process		Moderate
9. Explicit assessment of outcomes		Moderate
10. External quality measure		Weak
11. Activities evaluated	increasing access to quality primary education, with a focus on girls' participation, engaging communities in the management and development of the education system, and improving institutional capacity	Moderate
12. Additional aspects that make the study worthy of inclusion	very short evaluation, briefly discussed participation of women in community meetings and involvement of technical discussion leaders.	

Title	Amader school project in Bangladesh: evaluation	
Author/Agency		Concern
Date published		2012
Item	Description	Strong/Moderate/Weak
1. Relevance		
2. Program description	The Amader school project focused on primary school completion for extremely poor and excluded children, working with partners to create and encourage participation of local school based community groups (PTAs, mothers' groups, etc).	strong
3. Evaluation objective		strong
4. Approach	mixed methods	strong
5. Rigor		weak
6. Target audience		
7. Participatory evaluation		strong
8. Explicit assessment of process	evaluation too short to be explicit	moderate
9. Explicit assessment of outcomes		moderate
10. External quality measure		moderate/weak
11. Activities evaluated		strong
12. Additional aspects that make the study worthy of inclusion	Very short evaluation but interesting approach. Only 3 pages	

Title	Burundi Education Programme: Evaluation	
Author/Agency		Concern
Date published		February 2012
Item	Description	Strong/Moderate/Weak
1. Relevance	clear presentation, relevant	strong
2. Program description	very short description of education program	moderate
3. Evaluation objective	to assess outcomes on achievement	strong
4. Approach	results based approach and integrated key (DAC) indicators	strong
5. Rigor		strong
6. Target audience		
7. Participatory evaluation	participatory evaluation with 6 team members from CWB Burundi's education team, collected qualitative and quantitative data at school, commune and provincial level.	very strong
8. Explicit assessment of process		moderate
9. Explicit assessment of outcomes		moderate
10. External quality measure		moderate
11. Activities evaluated	First, the evaluation examined government initiatives at community engagement in education management. Second, the evaluation investigated the access of the poorest and most marginalized to quality education. Third, the evaluation assessed capacity- building of government institutions	strong
12. Lessons learned re: education, aid to education, evaluations		
13. Utility	interesting b/c it is an NGO and for participatory methods in evaluation, but very short evaluation.	
14. Additional aspects that make the study worthy of inclusion	The evaluation noted that the high extent of the implementing NGOs partnership with national, provincial and commune level officials was evident throughout the entire evaluation process.	

Title	Promising practice in school-related gender-based violence prevention and response programming globally	
Author/Agency		Concern
Date published		2013
Item	Description	Strong/Moderate/Weak
1. Relevance	might be interesting given Sida's gender/conflict priorities	Strong
2. Program description	a thematic approach: Concern's multi-level gender-based approach to school-related violence was adopted by the following agencies: Actionaid, USAID and Plan International. Other agencies adopted the Concern approach (Save the Children, UNICEF, International Rescue Committee) yet identified separate categories of violence.	strong
3. Evaluation objective	regarding school-related gender-based violence, the evaluation reviews best practices and effective interventions.	strong
4. Approach	desk review	weak
5. Rigor	Indicated a huge absence of objective data recording behaviour change in terms of reduced violence in schools and communities. Data was largely self-reported and involved checking off boxes, and interviews, when conducted, were structured interviews.	weak
6. Target audience	Concern, IrishAid, University of Sussex	
7. Participatory evaluation	Challenges listed below	Weak
8. Explicit assessment of process		Strong
9. Explicit assessment of outcomes	goal was to monitor the approach/methodology, therefore outcomes were defined in terms of adoption of the approach but not in terms of concrete results of the approach.	Moderate
10. External quality measure	There was little or no triangulation from outside sources, nor routine observations conducted.	Weak
11. Activities evaluated		Moderate
12. Additional aspects that make the study worthy of inclusion	The evaluation notes a huge gap in the literature on how to evaluate sexual and gender-based violence in schools. The challenge going forward is two-fold: finding a methodology most adapted for interviewing children, finding a methodology suited to monitor progress towards achieiving outcomes, as well as impacts on behavior.	

Title	**Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations**	
Author/Agency	Katherine Conn (dissertation – Columbia University)	
Date published	2014	
Item	Description	Strong/Moderate/Weak
1. Relevance	Meta-analysis of 12 types of interventions in SS-Africa	Strong
2. Program description	12 types: pedagogical, class size, instructional time, school supplies, abolishment of school fees, cash transfers, infrastructure, information/accountability, school-based management/decentralization, school meals, health treatments, student incentives, teacher incentives	Moderate
3. Evaluation objective	Present relative effectiveness but also understand why certain interventions seem to be more effective than others	Strong
4. Approach	Meta-analysis (pooled effect sizes of 12 interventions) - limited capacity to explore "why"	Moderate
5. Rigor		Strong
6. Target audience	Academic audience	Strong
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	No	N/A
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure (MMAT)		Strong
11. Activities evaluated	Specific activities are not explained in depth (given the nature of a meta-analysis)	Moderate
12. Lessons learned re: education, aid to education, evaluations	Interventions in pedagogical methods have higher pooled effect size on achievement outcomes than all other intervention types, adaptive instruction and teacher coaching techniques in particular. Health treatments have large pooled effect size on cognitive assessments (but smallest effect on achievement assessments)	
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	**Multifaceted Challenges – A study on the barriers to girls' education: Province – Mozambique**	
Author/Agency	DANIDA, CIDA, UNICEF	
Date published	2005	
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of barriers to education – not of aid to education	Weak
2. Program description	N/A – no program evaluated	N/A
3. Evaluation objective	Identify the supply and demand side barriers to girls' education	N/A
4. Approach	No description of methods	Weak
5. Rigor		Weak
6. Target audience	Not identified	Weak
7. Participatory evaluation		Weak
8. Explicit assessment of process	N/A	N/A
9. Explicit assessment of outcomes	N/A	N/A
10. External quality measure		Weak
12. Lessons learned re: education, aid to education, evaluations	Supply-side barriers (quality, institutional capacity, location and condition of schools, teachers, costs, impact of HIV/AIDS), and demand side barriers (poverty, perceptions of schooling, impact of HIV/AIDS)	
13. Utility	More useful to researchers and/or program designers – identification of barriers to education, not evaluation of efforts to improve access/quality.	
14. Additional aspects that make the study worthy of inclusion	None – not an evaluation	
12. Lessons learned re: education, aid to education, evaluations		

Title	Evaluation a mi-parcours du Plan decennal de developpement du secteur de l'education du Benin (PDDSE 2006-2015)	
Author/Agency	DANIDA, AFD, Bénin Ministry of Development Economics, Analysis and Forecasting (MCPD)	
Date published	February 2012 (note: in English at the end)	
Item	Description	Strong/Mode-rate/Weak
1. Relevance		strong
2. Program description	pre-primary education, primary education, secondary education, vocational education, higher education, and research, as well as adult education.	strong
3. Evaluation objective	The evaluation aims to measure to what extent objectives have been achieved in terms of decentralization of the educational sector in Benin, the actual situatlon, results achieved and lessons learned. before beginning the third phase of the program.	strong
4. Approach	The evaluation was initiated by Beninese authorities represented by « l'Observatoire du changement social (OCS) », in partnership with DANIDA and AFD. The study was conducted by an Independent team of four consultants. The evaluation was centered around 3 themes : a summary and analysis of policies and strategies undertaken and the results obtained, management and initiation of sectorial dialogue, and sector financing.	strong
5. Rigor	very descriptive (seems mostly qualitative with some numbers, but no mention of methods) limitations acknowledged: information system not capable of informing policy due to limited/unavailable data in the education sector. The evaluators also note the high degree of centralization involved in data collection and dissemination, and « significant « delays » in the production of annual statistics…ministries do not really use the indicators and performace reports are not rigorous enough to be credible, » (AFD/DANIDA/MCPD, 2012 : 48).	Strong/Moderate
6. Target audience	Danida et l'Agence Franc ¸aise de Développement (AFD), l'Observatoire du changement social (OCS) du Bénin et l'ambassade du Danemark au Bénin	
7. Participatory evaluation	The evaluation team received commentary and advice from an evaluation management committee, and there was also a local reference group which was comprised of « all » stakeholders, ministry resprentatives, unions, parent teacher organizations, civil society organizations that are active in the sector, which helped facilitate information flows. The evaluation noted that these consultations helped improve content and the form of the	Strong

	evaluation.	
8. Explicit assessment of process		Strong
9. Explicit assessment of outcomes		Strong
10. External quality measure		Strong
11. Activities evaluated	decentralization	Strong
12. Lessons learned re: education, aid to education, evaluations	The evaluation notes that decentralization appears to have had much more success in terms of the water and health sectors than the education sector in Benin, and that « decision-making remains highly centralized with limited delegation of responsibilities…the ministries in charge of education are not inclined to significantly transfer competencies to the commune level, » (AFD/DANIDA/MCPD, 2012 : 48). It notes that « when services are decentralized, there are limited resources to accompany their management »…this is particularly notable in educational quality, equity, and delivery (AFD/DANIDA/MCPD, 2012 : 48).	
13. Utility	The evaluation shows downstream pressure of EFA on secondary education, also that gaps in girls access to education persist despite measures.	
14. Additional aspects that make the study worthy of inclusion	Very clear and well-structured document. Strong perhaps in terms of lessons learnt.	

Title	DFID's Education Programmes in Three East African Countries	
Author/Agency		DFID
Date published		May 2012
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of bilateral aid to primary education in three East African countries: Ethiopia, Rwanda, and Tanzania	Strong
2. Program description	provision of basic education	Strong
3. Evaluation objective	to produce pilot studies to inform a transition from traditional aid approaches to results-based aid, investigating progress towards DFID objectives of quality education, cost-effectiveness and sustainability.	Strong
4. Approach	review of evidence on education program effectiveness review of DFID policy documents and guidance materials, analysis of spending patterns and interviews with London-based DFID staff for each country case study, reviewed DFID program design documents, performance frameworks, national education strategic plans and related reviews and evaluations; and conducted country visits over two months (DFID staff, other development partners, ministry of education officials, district education officers, head teachers, teaching staff, parents and civil society experts), and announced and unannounced school visits.	Strong
5. Rigor	Mixed Methods	Strong
6. Target audience	DFID	Moderate
7. Participatory evaluation	Important finding: notes more and more evidence that the more decisionmaking and accountability at the local level, the more learning outcomes improve. The evaluators were highly impressed by the high level of commited engagement by parent-teach association representatives, who also had a high range of responsibility (“from signing off school accounts to dealing with instances of bullying and dropping out”).	Strong

8. Explicit assessment of process		Weak
9. Explicit assessment of outcomes	Indicates that there is a lack of focus on learning outcomes	Moderate
10. External quality measure	Limitations discussed	Moderate
11. Activities evaluated	Implementation of a wide-ranging strategic plan in each country.	Moderate
12. Lessons learned re: education, aid to education, evaluations	Objectives were overly ambitious, with "competing objectives and insufficient prioritisation." "Following the MDGs, DFID has tended to define its objectives in terms of national averages. These mask major differences within and between regions in each of the three case study countries….leading to a missed opportunity to identify localized interventions." (p. 7) Calls for "focusing more broadly on public financial management and aid effectiveness, rather than the sector-specific questions of management systems and organisational change, such as the links between inputs, outputs and learning outcomes in education," (p. 10). "The platform for dialogue which the budget support monitoring process provides is only as good as the quality of input that development partners bring to it. Annual sector review processes can easily become routine, without a sufficient level of challenge, especially if they focus on national averages as the key targets," (p. 10).	
13. Utility	Mostly for DFID	
14. Additional aspects that make the study worthy of inclusion	Indicates that "too little attention is paid to issues of institutional change, the requirements of decentralised management or the need to make difficult choices in an environment of scarce resources, in contrast to the World Bank or USAID education strategies, which are more explicit about the need for institutional change and systems development," (p. 9). As budget support has not yet addressed institutional bottlenecks, the evaluators recommend that DFID work more closely with recipient countries to resolve complex reform challenges, and should be complemented by other forms of aid (such as project-based aid to build institional capacity, innovation funds, and targeted projects for specific issues like girls' education and parent organizations).	

Title	Evaluation of Results-Based Aid in Rwandan Education – 2013 Evaluation Report	
Author/Agency	Commissioned by DFID, conducted by Upper Quartile in association with the Institute of Policy Analysis and Research-Rwanda	
Date published	August 2014	
Item	Description	Strong/Moderate/Weak
1. Relevance	Mixed-methods process and impact evaluation of the results-based aid (RBA) pilot in Rwandan education (2012-2014)	Strong
2. Program description	RBA pilot in the Rwandan education sector	Strong
3. Evaluation objective	Impact of RBA on increasing school completion rates, and on teachers becoming more fluent in instructing in English.	Strong
4. Approach	Econometric modeling exercise (two models, using public data to examine completion rates). The evaluation takes into account the perspective of the recipient and other key actors to RBA; the influence of various interrelated factors influencing outcomes; identifies 'lessons learned' about how to improve the RBA pilot in Rwanda, about the effectiveness of RBA more generally as a funding mechanism and how RBA may be transferred to other contexts. The methodological approach adopted is that of 'realist evaluation'; "setting out to explore key questions about what works, for whom, in what circumstances and why," (p. 15).	Strong
5. Rigor	Mixed-Methods. The evaluation team undertook a context mapping and political economy analysis, utilizing national policy documents, existing research, and the evaluators' analysis of a housing survey. The two models were an internal check on each other in terms of validity of assumptions.	Strong
6. Target audience	DFID	Moderate
7. Participatory evaluation	"Qualitative fieldwork complements and helps 'unpack' the findings of the econometric modeling," (p. 19) and therefore undertaken at the national level, district level and school level, interviewing NGOs and national level	Strong

	government officials, district education officers, mayors, principals, sector education officials, groups of teachers, groups of parents, PTA chairpersons, and groups of students.	
8. Explicit assessment of process	Yes. Process-related questions	Strong
9. Explicit assessment of outcomes	Yes. Impact-related questions	Strong
10. External quality measure		Moderate
11. Activities evaluated	impact on school completion	Moderate
12. Lessons learned re: education, aid to education, evaluations	"Increases in the number of teachers have had a positive effect on completion, but attention is needed to improve teacher morale and attendance and their proficiency in English," (p. 7). Analysis of equity: disabled; specifically the mentally disabled, are less likely to attend/complete, likely due to lack of teachers trained in special education. Despite advancements in gender equity, might want to consider support for female learners who are at greater risk of non-completion in certain types of districts (lowest literacy rates, higher rates of povertiy).	
13. Utility	realist evalaution	
14. Additional aspects that make the study worthy of inclusion	Interesting approach: divided into impact-related questions and process-related questions. "In an agreed departure from the TOR, the framework for research and analysis is provided by a set of seven macro-evaluation questions developed and agreed by key members of the Upper Quartile evaluation team, the DFID Rwanda Education Adviser, and the DFID Lead on Payment by Results (PBR) Approaches," (p. 5).	

Title	A rigorous review of the political economy of education systems in developing countries	
Author/Agency	Commissioned by DFID, completed by the IOE	
Date published	April 2014	
Item	Description	Strong/Moderate/Weak
1. Relevance		Strong
2. Program description		N/A
3. Evaluation objective	Review literature from various disciplinary and interdisciplinary traditions; provide a conceptual framework to situate the analysis of political economy issues in education research; and identify research gaps	Strong
4. Approach	Many actions of teachers and schools – and the school outcomes that they are accountable for – are influenced by incentives and constraints operating outside the schooling system, in the external environment. All of these environmental factors influence education reform and its implementation ("whether policy design, financing, implementation or evaluation," p. 1). Despite the importance of these power relations in influencing teaching and learning, there is limited literature on power relations and their role in education's external environment to guide policymaking. Therefore, the evaluation calls for an interdisciplinary approach particularly for education, which may not be served by a single disciplinary lens (p. 1).	Moderate
5. Rigor	"Stringent inclusion and exclusion criteria were agreed for screening the evidence base. Included studies were characterized on the basis of features such as geographical region/country (giving some preference to DFID priorities) appropriateness of data collection, and data analysis and study design (qualitative or quantitative), etc," (p. 9).	Weak
6. Target audience	Policymakers/DFID	
7. Participatory evaluation		N/A
8. Explicit assessment of process		Strong
9. Explicit	Strengths: Authors' own expertise in research;	Moderate

assessment of outcomes	heterogeneity of sources consulted and research designs. Yet, difficult to draw strong comparisons from studies with different methodologies and examining different phenomena, also contextual factors challenge comparison.	
10. External quality measure	"Followed typical series of steps for a systematic review, yet acknowledge that a rigorous literature review requires adopting more flexible standards than in a systematic review," (p. 9). Each individual study was assessed by at least two review members under each of DFID's six principles of high quality studies (Eval: DFID 2013, p.10). These six principles are: 1. Conceptual framing; 2. Openness and transparency; 3. Appropriateness and rigour; 4. Validity; 5. Reliability; 6. Cogency," (Eval: DFID 2013, p.10).	Moderate
11. Activities evaluated	"These six principles were applied to each study in a consistent and comprehensive manner. For example, a hierarchy of evidence was used to evaluate the validity of quantitative studies ranging from randomised controlled trials (RCTs) (high quality) to less rigorous methodologies such as simple descriptive statistics that do not allow causal interpretations (such as comparison of means)," (Eval: DFID 2013, p.10).	Weak
12. Lessons learned re: education, aid to education, evaluations	Teacher unions exert great influence on the shaping of education policies, among all stakeholder groups.	
13. Utility	Funding agencies	
14. Additional aspects that make the study worthy of inclusion	Interesting perspective in terms of teacher unions (not much discussion of teacher unions in our synthesis); rent-seeking and patronage politics; decision-making and the process of influence; implementation issues; and driving forces. Several of our other evaluations address the need for political economy analysis and this review provides it. Also, interesting model for assessing the quality of evidence (six principles mentioned above). Useful for how to do a synthesis.	

Title	**Pedagogy, Curriculum, Teaching Practices and Teacher Education in Developing Countries**	
Author/Agency	DFID, University of Sussex	
Date published	2013	
Item	Description	Strong/Moderate/Weak
1. Relevance	Synthesis of (mostly) academic research, not focused on aid, but still relevant	Moderate
2. Program description	N/A – meta-analysis of lots of studies	N/A
3. Evaluation objective	What pedagogical practices are used by teachers in developing countries, what evidence is there of the effectiveness of these pedagogical practices, how can teacher education, school curriculum and guidance materials support effective pedagogy?	Strong
4. Approach	Systematic mapping of studies and then in-depth review of select group – quantitative and qualitative studies	Strong
5. Rigor		Strong
6. Target audience	Academic audience, practitioners	Strong
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	Yes	Strong
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure (MMAT)		Strong
11. Activities evaluated	N/A	
12. Lessons learned re: education, aid to education, evaluations		
13. Utility	Very useful for planning specific interventions – those designed to improve pedagogy	
14. Additional aspects that make the study worthy of inclusion		

Title	What works to improve teacher attendance in developing countries? A systematic review	
Author/Agency	DFID and partners	
Date published	October, 2012	
Item	Description	Strong/Moderate/Weak
1. Relevance		
2. Program description		
3. Evaluation objective	To review current research on the effectiveness of interventions aimed at increasing teacher attendance in developing countries, measured by teacher attendance.	Strong
4. Approach	Systematic review of quantitative studies using experimental or quasi-experimental designs	Strong
5. Rigor		Strong
6. Target audience		Strong
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	No	N/A
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure (MMAT)		Strong
11. Activities evaluated	Programs aimed at improving teacher absenteeism (directly or indirectly)	
12. Lessons learned re: education, aid to education, evaluations	Findings (from 9 studies that meet the methodological criteria established) suggest that direct interventions coupled with incentives to implement and use monitoring systems, and community involvement, can positively impact on teacher attendance, but more is needed to improve achievement	
13. Utility	Review on interventions	
14. Additional aspects that make the study worthy of inclusion	For diversity – few other evaluations focus on teacher attendance, an important challenge in educational development	

Title	Literacy, Foundation Learning and Assessment in Developing Countries	
Author/Agency	DFID and partners	
Date published	2014	
Item	Description	Strong/Moderate/Weak
1. Relevance	Review of studies addressing literacy and foundation learning in developing countries (methodologically diverse)	strong
2. Program description	very focused on the process of teaching and learning	strong
3. Evaluation objective	address issues pertaining to foundation learning and literacy.	
4. Approach	qualitative (no interviews, just documents). Some of the studies reviewed were ethnographic studies.	moderate
5. Rigor	"Considered within-child factors, including cognitive and language skills, and contextual factors including home language and literacy environment, community practices and quality of opportunity as well as the social stratifiers and economic drivers that influence non-enrolment, poor attendance, and dropout," (p. 1) then reviewed various interventions.	moderate
6. Target audience	not clear, DFID?	weak
7. Participatory evaluation		weak
8. Explicit assessment of process	Both child-level and school-level factors affect attainments, but the relative impact of the two sources of variability is difficult to quantify.	moderate/strong
9. Explicit assessment of outcomes		very strong
10. External quality measure	addresses limitations	moderate

11. Activities evaluated	literacy	moderate
12. Lessons learned re: education, aid to education, evaluations		moderate
13. Utility		
14. Additional aspects that make the study worthy of inclusion	more on the process of learning than outcomes, review of interventions	

Title	The role and impact of private schools in developing countries	
Author/Agency	Ashley, Mcloughlin, Kingdon, Nicolai, Rose	
Date published	2014	
Item	Description	Strong/Moderate/Weak
1. Relevance	Focused on the role of private schools, not the relationship between aid and private schools	Moderate
2. Program description	Private schools	
3. Evaluation objective	Research question: can private schools improve education for children in developing countries?	
4. Approach	Systematic review	Strong
5. Rigor		Strong
6. Target audience	Researchers, policy makers	Strong
7. Participatory evaluation	No	Weak
8. Explicit assessment of process	No	Weak
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure		Strong
11. Activities evaluated	Private schools in developing countries	Moderate
12. Lessons learned re: education, aid to education, evaluations	Lessons regarding education	
13. Utility		
14. Additional aspects that make the study worthy of inclusion	Worth including the role of private schools in educational development according to prominent meta-analyses	

Title	Evaluation of Results Based Aid in Rwandan Education – Year Two	
Author/Agency	DFID – conducted by Upper Quartile	
Date published	2015	
Item	Description	Strong/Moderate/Weak
1. Relevance	Very relevant – evaluation of aid to basic education sector in Rwanda	Strong
2. Program description	RBA – sector aid is tied to improving completion rates (primary and secondary)	Strong
3. Evaluation objective	Identify contribution of results-based aid (RBA) pilot to (1) increased school completion (primary and secondary), (2) increased use of English in instruction by teachers, and (3) the response of the government and other actors to RBA	Strong
4. Approach	Mixed methods – "realist" – explores what works, for whom, in what circumstances, and why. Impact evaluation consists of difference-in-differences model, process evaluation, to identify role of results-based component, and Value for Money approach	Strong
5. Rigor	Explicit attention to limitations, description and analysis are directly linked to methods, methods are replicable, and appropriate for context	Strong
6. Target audience	Rwandan education officials, DFID	Strong
7. Participatory evaluation	Incorporates qualitative methods, participatory in the process and VfM parts	Moderate
8. Explicit assessment of process	Yes	Strong
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure		Strong
11. Activities evaluated	Private schools in developing countries	Moderate
12. Lessons learned re:	Lots of information about the value of results-based aid, role of country ownership in delivering aid, excellent example of a mixed-methods,	

education, aid to education, evaluations	contextualized, yet also quantitatively (relatively) rigorous evaluation
13. Utility	
14. Additional aspects that make the study worthy of inclusion	

Title	Thematic global evaluation of European Commission support to the education sector in partner countries (including basic and secondary education)	
Author/Agency		European Commission
Date published		December 2010
Item	Description	Strong/Moderate/Weak
1. Relevance	European Commission and support to education sector	strong
2. Program description	Strategy evaluation: Between 2000-2007 substantial organisational changes in EC external aid: Merging directorates, creation of EuropeAid, deconcentration. Staffing: Heavy dependence on contractors (education sector). Staffing still limited, despite deconcentration (EC study 2009). Emphasis now on policy analysis and dialogue, leading to increased workload	strong
3. Evaluation objective	to assess impact and efficiency of EC aid (primary and secondary education), extent of donor complementarity and coordination, coherence with EC policies and partner Governments' priorities and activities, and with international commitments.	strong
4. Approach	Qualitative	strong
5. Rigor	Limitations acknowledged: attribution, difficulties in producing a comprehensive inventory of EC funding, and access to, and availability of, information because of the lack of institutional memory at EC HQ and field levels. Access to data and stakeholders was also sometimes constrained during field visits. However, the evaluation team compensated for this to a certain extent by cross-checking and combining information from different sources.	strong
6. Target audience	EC and other European development agencies	moderate
7. Participatory evaluation	focus groups with individuals across agencies.	weak
8. Explicit assessment of process	methodology applied: methodology utilized by the Joint Evaluation Unit. Began with an overview and typology of European Commission Aid, developed methodological framework, intervention logic, selected 23 countries receiving EC support and then a subset of 6 countries for desk study/field visits, then	strong

	developed evaluation questions, used broadly agreed-upon international indicators	
9. Explicit assessment of outcomes	To better understand the dynamics in different contexts and to extract lessons.	strong
10. External quality measure	Reference group, dissemination seminar at completion.	strong
11. Activities evaluated	the main objective of the field phase was to complete data collection and contribute to answering the evaluation questions, as well as to address specific issues more in-depth – however, in bold – "the field phase was not intended to conduct an in-depth assessment of the implementation of specific EC interventions," (p. 35). The emphasis was on processes and achievements that "could not be fully covered by the tools of the desk analysis," (p. 35).	moderate
12. Lessons learned re: education, aid to education, evaluations	Nothing new in terms of lesson learned	
13. Utility	European Commission evaluation –possibly of interest to Sida	
14. Additional aspects that make the study worthy of inclusion		

Title	Promotion de l'éducation de base, Tchad	
Author/Agency		GIZ
Date published		July 2012
Item	Description	Strong/Moderate/Weak
1. Relevance		Strong
2. Program description	The project operated outside of the Chadian Education Administration; it was executed by GTZ, but implicated other organizations such as KfW and the World Bank, within the framework of the national education program.	Strong
3. Evaluation objective	To evaluate innovative approaches ameliorating basic education access and quality, in particular for girls in 3 regions, integrated within the national policy.	
4. Approach	Qualitative	Strong
5. Rigor	The evaluation team (an international expert and a national expert) conducted 36 individual interviews and 19 group interviews in the capital and in one region and surveyed 220 students and 33 teachers through a standardized questionnaire in a second region.	Strong
6. Target audience	funding agencies and the Ministry of Education	
7. Participatory evaluation	participatory evaluation - interviews with parent associations, community-based associations, etc, as well as partner organizations	strong
8. Explicit assessment of process		Strong
9. Explicit assessment of outcomes	Parent associations were eventually included in the national sectoral policy as a result of the project, and therefore strengthened at the institutional level.	Strong
10. External quality measure		Strong

11. Activities evaluated	Project was mostly active in providing support to community schools, ie by parent associations, native language education in primary schools, the promotion of girls education and by improving knowledge, attitudes and practices related to HIV / AIDS. Activities also included innovative approaches, and providing different textbooks and trainings for parent, teacher, and student associations). The evaluation looked at social interventions such as the formation of networks between parent associations, and new pedagogical approaches.	Strong
12. Utility	Targeted groups were primary school-age girls, their parents, and parent associations.	
13. Lessons learned re: education, aid to education, evaluations		
14. Additional aspects that make the study worthy of inclusion		

Title	Cross Section Analysis of Education Sector: Meta-Evaluation and Synthesis	
Author/Agency		GIZ
Date published		2014
Item	Description	Strong/Moderate/Weak
1. Relevance	Synthesis of reviews of projects conducted by GIZ in the education sector (note: only found the executive summary, full document not available)	Strong
2. Program description	N/A – meta-analysis	N/A
3. Evaluation objective	Summarize and aggregate findings from project reviews/evaluations in order to make recommendations	Strong
4. Approach	Developed grid based on asking questions from TORs – used to thematic clusters	Strong
5. Rigor	Hard to say w/out full report	N/A
6. Target audience	GIZ officials, policy makers	Strong
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	Yes – but hard to assess w/out full report	N/A
9. Explicit assessment of outcomes	Yes – but hard to assess w/out full report	N/A
10. External quality measure (MMAT)	Hard to assess w/out full report	N/A
11. Activities evaluated	GIZ support to education sector – broadly	N/A
12. Lessons learned re: education, aid to education, evaluations	Lack of "proof of impact" – direct and indirect impacts are usually not demonstrated across reviewed projects, quality is often not assessed,	
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	Ex-post evaluation 2012 – Education in Namibia	
Author/Agency		GIZ
Date published		
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of GIZ support to Namibia education sector – 4 years	Strong
2. Program description	Technical assistance to the MoEd, improvement of decentralized education management, improvement of access to quality instruction in mother tongue language, revision and introduction of new curriculum for primary schools	Strong
3. Evaluation objective	Ex post evaluation performed 5 years after program ended.	Moderate
4. Approach	Document analysis, 70 structured interviews with aid officials, government officials, teachers, school boards, with 9 focus group discussions – mostly done remotely	Moderate
5. Rigor	Hard to say w/out full report	N/A
6. Target audience	GIZ officials, policy makers	Strong
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	Yes – but hard to assess w/out full report	N/A
9. Explicit assessment of outcomes	Yes – but hard to assess w/out full report	N/A
10. External quality measure (MMAT)	Hard to assess w/out full report	N/A
11. Activities evaluated	GIZ support to education sector – broadly	N/A
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	**Review of IDB Support to Secondary Education: Improving Access, Quality and Institutions, 1995-2012**	
Author/Agency	Office of Evaluation and Oversight – Inter-American Development Bank	
Date published	October 2013	
Item	Description	Strong/Moderate/Weak
1. Relevance	Overall review (mostly descriptive) of Bank funding of secondary education projects	Strong
2. Program description	Evaluation of multiple strategies – formal and non-formal education sector, demand and supply side - none described in detail	Moderate
3. Evaluation objective	Objective: examine Bank support for secondary education to identify lessons and provide recommendations to strengthen future Bank performance, specifically regarding: Equitable access to secondary education Secondary education quality Reforms of education institutions to improve management capacity	Strong
4. Approach	Desk-based review of 58 projects, 9 country case studies, and a literature review	Strong
5. Rigor		Moderate
6. Target audience	IDB officials	Strong
7. Participatory evaluation		Weak
8. Explicit assessment of process	Limited critical analysis of process or outcomes and the role of multiple constituencies	Weak
9. Explicit assessment of outcomes		Weak
10. External quality measure		Weak
11. Activities evaluated	"Access interventions," "quality interventions" and "institutional reforms'	Strong

12. Lessons learned re: education, aid to education, evaluations	
13. Utility	
14. Additional aspects that make the study worthy of inclusion	For diversity and attention to broad strategy/educational reforms.

Title	Ex-Post Evaluation Report on the Project of the Construction and Extension for 18 Primary Schools in Nairobi, Nakuru, and Thika (Kenya)	
Author/Agency	KOICA/ World Friends	
Date published	2012	
Item	Description	Strong/Moderate/Weak
1. Relevance	Aid funded education project	Strong
2. Program description	Objective: "improve access to primary education and educational environment" – through construction and extension of 10 primary schools	Moderate
3. Evaluation objective	Evaluate these programs through OECD-DAC requirements: relevance, efficiency, effectiveness, impact, sustainability	
4. Approach	Literature review, "in-depth interviews with stakeholders in Korea and in Kenya," focus groups, field survey	
5. Rigor	Focus groups, in-depth interviews, surveys	Moderate
6. Target audience		Moderate
7. Participatory evaluation		Weak
8. Explicit assessment of process	Yes – but report is poorly written and organized	Weak
9. Explicit assessment of outcomes	No	Weak
10. External quality measure	No	Weak
11. Activities evaluated		Weak
12. Lessons learned re: education, aid to education, evaluations		

13. Utility	Example of an evaluation that may be useful to the project coordinator due to specific findings (in bulleted list) regarding technical and operational challenges, but has very limited value in terms of providing insights into education, aid, or evaluation
14. Additional aspects that make the study worthy of inclusion	Could be worth including just for diversity sake (diversity in bi-lateral agency, and quality of evaluation)

Title	Ex-post Evaluation Report on the Two Primary and Secondary Education Projects in Palestine	
Author/Agency	KOICA/World Friends Korea	
Date published	2012	
Item	Description	Strong/Moderate/Weak
1. Relevance	Aid-funded education program	Strong
2. Program description	Objective: "Human resources development initiative" – build school facilities for technical education, improve training environments, build girls school, donate computers, provide nutritional supplements to students	Moderate
3. Evaluation objective	Evaluate these programs through OECD-DAC requirements: relevance, efficiency, effectiveness, impact, sustainability	Weak
4. Approach	Focus groups, in-depth interviews, surveys	Moderate
5. Rigor		Moderate
6. Target audience		Weak
7. Participatory evaluation	Yes – but report is poorly written and organized	Weak
8. Explicit assessment of process	No	Weak
9. Explicit assessment of outcomes	No	Weak
10. External quality measure		Weak
11. Activities evaluated	Construction of schools, donation of computers	Weak
12. Lessons learned re: education, aid to education, evaluations		

13. Utility	Example of an evaluation that may be useful to the project coordinator due to specific findings (in bulleted list) regarding technical and operational challenges, but has very limited value in terms of providing insights into education, aid, or evaluation
14. Additional aspects that make the study worthy of inclusion	Could be worth including just for diversity sake (diversity in bi-lateral agency, and quality of evaluation)

Title	Impact Evaluation of Burkina Faso's BRIGHT Program	
Author/Agency	Mathematica (Dan Levy, Matt Sloan, Leigh Linden, Harounan Kazianga, for USAID)	
Date published	2009	
Item	Description	Strong/Moderate/Weak
1. Relevance	Focus on improving educational outcomes, among girls in particular, - primary school construction, canteens, take-home rations, textbooks, mobilization campaign, literacy training, capacity building.	Strong
2. Program description	Good description of all components. Specific attention paid to the 10 provinces where girls' enrollment rates were lowest.	Strong
3. Evaluation objective	Evaluate the impact of the program on school enrollment, test scores, and assess heterogeneous impacts (boys versus girls).	Strong
4. Approach	Quantitative – quasi experimental (regression discontinuity design). The impact evaluation examined program impact on school enrollment and on test scores, and whether there were gender differences. Previous two reports were just assessments.	Strong
5. Rigor	Causality (attribution) addressed through statistically viable comparison group, with limitations, which the authors discuss	Strong
6. Target audience	USAID and partners	
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	No	N/A
9. Explicit assessment of outcomes	Yes-- even in the absence of BRIGHT, it is likely that enrollment would have increased in the 132 villages in which it was implemented.	Strong
10. External quality measure (MMAT)		Strong
11. Activities evaluated	primary schools construction, implementation of complementary interventions to increase girls' enrollment (separate latrines for boys and girls; canteens; take-home rations and textbooks; as well as a mobilization campaign, literacy training, and capacity	Strong

	building among local partners).	
12. Lessons learned re: education, aid to education, evaluations	Limitations discussed: Challenges to external validity: need to pay attention to different contexts if they want to implement the intervention elsewhere. In terms of comparing BRIGHT with other recently evaluated education interventions, the evaluators point out that in many places schools already exist—in this evaluation, schools did not exist before; thus the BRIGHT context might be specific to these policy instruments. Calls for cost-effective analysis: for example, would building a less expensive school have the same effects? Questions asked about program sustainability (after intervention) and long-term outcomes.	
13. Utility	Lessons learned in terms of limitations of impact evaluations. The evaluation does not however tell us anything new, but was generally well-conducted overall. Implemented by a consortium of NGOs—Plan International, Catholic Relief Services (CRS), Tin Tua, and the Forum for African Women Educationalists (FAWE)—supervised by USAID, and funded by Millennium Challenge Corporation (MCC).	
14. Additional aspects that make the study worthy of inclusion	Used regression discontinuity. Evaluation approach: assessed how children receiving the intervention fared relative to how they would have fared without the intervention.	

Title	Improving learning in primary schools of developing countries: a meta-analysis of randomized experiments	
Author/Agency	McEwan, P.	
Date published	2014	
Item	Description	Strong/Moderate/Weak
1. Relevance	Meta-analysis of 77 randomized experiments evaluating effects of "school-based interventions" on learning in low-income countries (focus is learning, not participation)	Moderate
2. Program description	77 programs – non are described in detail	Weak
3. Evaluation objective	Quantitative effect estimate	Strong
4. Approach	Meta-analysis	Moderate
5. Rigor		Moderate
6. Target audience	Researchers – economists	Moderate
7. Participatory evaluation	No	Weak
8. Explicit assessment of process	No	Weak
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure		Strong
11. Activities evaluated	Range of education programs – grants, deworming, nutritional treatments, disseminating information,	
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		
14. Additional aspects that make the study worthy of inclusion	I think it's worth mentioning "what works" according to prominent meta-analyses, systematic reviews, such as this one –and then contrasting our synthesis approach and findings.	

1Title	**Schools as Centres of Care and Support (SCCS) : Responding to the Needs of Orphans and Other Vulnerable Children in Rural Areas**	
Author/Agency	MiET Africa, Swiss Agency for Development and Cooperation (SDC) and the Embassy of the Kingdom of the Netherlands (EKN)	
Date published	November 2009	
Item	Description	Strong/Moderate/Weak
1. Relevance		Moderate
2. Program description	The SCCS programme, an example of a school-based response to the increase of orphans and vulnerable children, focuses on multi-sectoral partnerships to address poverty and health. Quality education is viewed as a means to strengthen schools, which are also to "function as hubs of integrated service delivery for children" (p. 3), to increase access to health, social services and education.	Strong
3. Evaluation objective	This case study presents an outline of the SCCS model; from its evolution, to implementation.	Strong
4. Approach	Case study	Moderate
5. Rigor		Moderate/Weak
6. Target audience	African policymakers	Moderate
7. Participatory evaluation	Schools are arranged in clusters to benefit from support for instructional quality as well as from partnerships with parents, the community, NGOs and governmental institutions offering social services, including health, nutrition, security and fund-raising assistance.	Moderate
8. Explicit assessment of process		Strong
9. Explicit assessment of outcomes		Strong
10. External quality measure		Weak

11. Activities evaluated	Presents an outline of the SCCS model, discusses implementation as well as extension of the pilot.	Strong
12. Additional aspects that make the study worthy of inclusion	Interesting in terms of participatory approach, the organization is an African NGO, and interesting in terms of multi-sectoral collaboration	
12. Lessons learned re: education, aid to education, evaluations	Cluster approach for community support to education.	
13. Utility	NGO evaluation in partnership with national funding agencies.	

Title	Primary Education in Zambia	
Author/Agency	Netherlands Ministry of Foreign Affairs: Policy and Operations Evaluation Department (IOB)	
Date published	April 2008	
Item	Description	Strong/Moderate/Weak
1. Relevance	Analysis of sector-wide support to education in Zambia	Strong
2. Program description	Sector-wide support: school construction, teacher training, school infrastructure support, reducing pupil/teacher ratio, management capacity development	Strong
3. Evaluation objective	Improve insight into effectiveness of education programmes, improve understanding of factors influencing education outcomes, improve investments in education, help MoE to use existing databases more effectively	Moderate
4. Approach	Multivariate regression analyses (education production function) to assess the association between learning, participation and background characteristics/specific interventions. Propensity score matching techniques was used to create (ex-post) control groups.	Strong
5. Rigor	Limitations clearly discussed, methods and data well described	Strong
6. Target audience	Netherlands Ministry of Foreign Affairs, MoE	Strong
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	No	N/A
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure (MMAT)		Strong
11. Activities	School and classroom construction, provision of	

evaluated	teaching and learning materials	
12. Lessons learned re: education, aid to education, evaluations	Major lesson from this evaluation – in line with what we see across evaluations – trade off between increasing educational access, and increasing educational quality.	
13. Utility		
14. Additional aspects that make the study worthy of inclusion	One limitation – the role of aid is not addressed	

Title	Primary Education in Uganda: Impact Evaluation	
Author/Agency	Netherlands Ministry of Foreign Affairs: Policy and Operations Evaluation Department (IOB)	
Date published	April 2008	
Item	Description	Strong/Moderate/Weak
1. Relevance	Analysis of the impact of education strategies under the Ugandan SWAp supported by the Netherlands.	Strong
2. Program description	Sector-wide support	Strong
3. Evaluation objective	Examine whether school attendance and enrollment increased, and which aid interventions had the most impact and were the most cost-effective in meeting these objectives.	Strong
4. Approach	Multivariate regression analyses of the association between various interventions (pupil/teacher ratio reduction, classroom availability, toilet availability, teacher education, teacher training, head teacher qualifications, distance to primary school) and access and learning.	Moderate
5. Rigor	Attribution not directly addressed, weaknesses with data dealt with somewhat	Moderate
6. Target audience	Policy makers at the Netherlands Ministry of Foreign Affairs, Ugandan government	
7. Participatory evaluation	No	Weak
8. Explicit assessment of process	No	Moderate
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure (MMAT)	Study methods are replicable, but limitations and approach are discussed in very general terms	Moderate

11. Activities evaluated	School construction, teacher training, school infrastructure support, reducing pupil/teacher ratio, management capacity development	Strong
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		Impact evaluation
14. Additional aspects that make the study worthy of inclusion		

Title	Policy Review of the Dutch Contribution to Basic Education	
Author/Agency	Netherlands Ministry of Foreign Affairs	
Date published	2009	
Item	Description	Strong/Moderate/Weak
1. Relevance	Dutch contribution to basic education 1999-2009	Strong
2. Program description	Dutch support to basic education via bilateral aid an also via multilateral agencies and Dutch NGOs (4th largest contributor to aid to education).	Strong
3. Evaluation objective	Policy review of the Dutch contribution to basic education 1999-2009	Strong
4. Approach	Policy review included (all focusing on basic education) (1) evaluations in four partner countries (2) a literature review on investment impact, (3) a review of external evaluations by six Dutch NGOs (working on basic education with Dutch funding) and (4) an analysis of Dutch expenditure, (summary document p.2)	Moderate
5. Rigor	Policy Review	Weak
6. Target audience	Ministry of Foreign Affairs of the Netherlands	
7. Participatory evaluation	No	Weak
8. Explicit assessment of process	No	Weak
9. Explicit assessment of outcomes	The evaluation notes that the Netherlands made a larger and more impactful contribution to basic education, and considered EFA and the MDGs and partner country priorities.	Moderate
10. External quality measure	No	Weak
11. Activities evaluated		Weak
12. Lessons learned re: education, aid to		

education, evaluations	
13. Utility	Useful as a review of the role of Dutch policy in aid to education, does not provide very much information into evaluation, basically an overview.
14. Additional aspects that make the study worthy of inclusion	Very brief overview, not very much detail.

Title	The two-pronged approach: Evaluation of Netherlands support to primary education in Bangladesh	
Author/Agency	Netherlands Ministry of Foreign Affairs	
Date published	August 2011	
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of Netherlands support to primary education in Bangladesh	Strong
2. Program description	support to basic education (1999-2009): formal and informal	Strong
3. Evaluation objective	Country study evaluating the "relevance, efficiency, effectiveness, and sustainability of the Netherlands contribution to basic education in Bangladesh," (p. 14).	Strong
4. Approach	Case study, mixed-methods. Bangladesh selected since evaluation notes that it was among the largest beneficiaries of aid to education. Evaluation states it aimed to examine the effectiveness of these channels in achieving EFA and the MDGs.	Strong
5. Rigor	"Extensive literature review, an analysis of quantitative data of the education sector, interviews with key players in the education sector in Dhaka and a qualitative field study that was conducted in two districts among local education officials, different types of primary schools, and teacher training institutes. No primary quantitative data collection was done for the purpose of the impact evaluation," (p. 14).	Strong
6. Target audience	Netherlands Ministry of Foreign Affairs	Moderate
7. Participatory evaluation	Key informant interviews with stakeholders engaged in education.	Moderate
8. Explicit assessment of process	Support to education was conducted through two separate channels – for non-formal primary education through BRAC, a major NGO player in Bangladesh, and for formal primary education through the Government of Bangladesh.	Strong
9. Explicit assessment of outcomes	Overall, the regression analysis confirms earlier findings assessing improvements in learning, but the evaluators not that the evaluation is inconclusive in terms of the input sdetermining these outcomes.	Strong

10. External quality measure	Some discussion of limitations	Moderate
11. Activities evaluated	Netherlands support to primary education (formal) and non-formal primary education (channeled through BRAC) in Bangladesh.	Strong
12. Lessons learned re: education, aid to education, evaluations	The evaluation notes that non-formal education through NGOs can cost less, take less time, and yield good learning outcomes, yet there is always the question of sustainability of aid for non-formal education.	
13. Utility	Interesting conceptual approach: two-pronged approach (more difersified than previous support).	
14. Additional aspects that make the study worthy of inclusion	also examines support to informal education and aid channeled through NGOs.	

Title	Joint Evaluation of Nepal's Education for All 2004 – 2009 Sector Programme	
Author/Agency	Cambridge Education Ltd, METCON Consultants, for Norad	
Date published		
Item	Description	Strong/Mode rate/Weak
1. Relevance	Evaluation of bi-lateral and multi-lateral support for EFA in Nepal	Strong
2. Program description		
3. Evaluation objective	Objective: provide information about outcomes of EFA program 2004 – 2009	Strong
4. Approach	Examination of national trends and district variations, primary data is qualitative – represents perceptions of actors from students to development partners	Strong
5. Rigor	Detailed description of ethnographic methods, "illuminative evaluation", limitations mentioned	Strong
6. Target audience	Ministry of Education and Sports (Nepal) and bi-lateral/multi-lateral aid community	Strong
7. Participatory evaluation	Yes – attention of how "program was received by the ultimate beneficiaries – students, parents, various actors at local level"	Strong
8. Explicit assessment of process	Yes	Strong
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure		Strong
11. Activities evaluated	Focus is on progress towards EFA goals, not specific activity.	Weak
12. Lessons learned re: education, aid to education, evaluations	Attention to inclusion of children with disabilities/special needs, bilingual and linguistic diversity	
13. Utility		
14. Additional aspects that make the study worthy of inclusion	Explicit attention not to just incorporating multiple constituencies, but even to techniques to encourage "different voices" – e.g. "talking to quieter children informally during the big group meeting"	

Title	Accelerated Primary Education Support (APES) Project in Somalia	
Author/Agency	Norwegian Refugee Council, European Commission, Save the Children, Concern World Wide	
Date published	2012	
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of aid-supported education project in Somalia	Strong
2. Program description	APES was implemented in 13 regions, development of "cohesive education system" - included infrastructure development, national campaigns to enroll students, improve curriculum, promote inclusive/gender responsive practices in communities and schools, improve management capacity – relatively detailed description of all	Strong
3. Evaluation objective	Measure output delivery and outcome achievement	Moderate
4. Approach	Literature review, interviews and focus groups with key actors in the field, field observations and photography, semi-structured school questionnaires.	Moderate
5. Rigor		Weak
6. Target audience	Key actors – local international aid officials, government officials	Strong
7. Participatory evaluation	No – except for focus group and interviews	Weak
8. Explicit assessment of process	Yes	Moderate
9. Explicit assessment of outcomes	Yes	Weak
10. External quality measure		Weak
11. Activities evaluated	Comprehensive support to Somalia education - see 2 above	Moderate
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	Les langues de scolarisation en Afrique francophone	
Author/Agency	Organisation internationale de la Francophonie (OIF), (AFD) Ministère des affaires étrangères et européennes (MAEE) Agence universitaire de la Francophonie (AUF)	
Date published	June 2010	
Item	Description	Strong/Moderate/Weak
1. Relevance	Not very relevant, since we are not doing a close look at instruction in indigenous vs colonial languages.	weak
2. Program description	Joint project between Affaires étrangères et européennes (MAEE), l'Agence franc ¸aise de développement (AFD), l'Organisation internationale de la Francophonie (OIF) and l'Agence universitaire de la Francophonie (AUF), aimed at improving pedagogical approaches in multilingual African contexts	strong
3. Evaluation objective	To evaluate language policy, didactic models, pedagogical prectices, methodological tools, curricula, teacher training, and evaluation of teachers involved with language-learning in Francophone Africa.	strong
4. Approach	qualitative	moderate
5. Rigor	6 country case studies and a culminating research event in Paris	strong
6. Target audience	Ministry of Foreign Affairs, Agence Française de développement.	moderate
7. Participatory evaluation	calls for participatory approaches (need to include various actors, including leaders, parents, students, teachers) in order to implement successful bilingual and multilingual educational systems.	Strong
8. Explicit assessment of process		Strong
9. Explicit assessment of outcomes	Pedagogical models are being used, as well as African languages.	Strong
10. External quality measure		Strong

11. Activities evaluated	French language instruction as well as choice of language in the classroom.	Moderate
12. Additional aspects that make the study worthy of inclusion	Focus on participatory approaches, language learning policy.	
13. Lessons learned re: education, aid to education, evaluations		
14. Utility		

Title	The Effectiveness of Foreign Aid to Education: What can be learned?	
Author/Agency		none (Abby Riddell)
Date published		2012
Item	Description	Strong/Moderate/Weak
1. Relevance		N/A
2. Program description		N/A
3. Evaluation objective	a paper rather than a formal evaluation. Useful for lit review	N/A
4. Approach		N/A
5. Rigor		N/A
6. Target audience		N/A
7. Participatory evaluation		N/A
8. Explicit assessment of process		N/A
9. Explicit assessment of outcomes		N/A
10. External quality measure		N/A
11. Activities evaluated		N/A
12. Lessons learned re: education, aid to education, evaluations	This review shows that "many of the lessons of what works in foreign aid to education are known, but they are not implemented,"(p. 37), and advocates focusing on the sector as a whole rather than sub-sectors.	
13. Utility	Good for literature review	
14. Additional aspects that make the study worthy of inclusion	This review demonstrates the difficulties of focusing only on inputs and outputs, particularly in terms of sustainability.	

Title	Implementing School-Based Management in Indonesia	
Author/Agency	RTI International	
Date published	September 2011	
Item	Description	Strong/Moderate/Weak
1. Relevance	Focused evaluation on school-based management, funded by USAID and implemented by RTI	Strong
2. Program description	This project operated in 50 districts to improve basic education management and governance, covering about 10% of Indonesia's population	Strong
3. Evaluation objective	Evaluate the effectiveness of the decentralization/school-based management tools by assessing various aspects of project performance and impact of the interventions	Strong
4. Approach	Routine project monitoring data, comparison of baseline/endline achievement data, qualitative field surveys, two annual quantitative surveys implemented in target schools, studies of school funding (at the school level), interviews with principals	Strong
5. Rigor		Strong
6. Target audience	USAID, Indonesian government, education officials (central and local)	Strong
7. Participatory evaluation		Moderate
8. Explicit assessment of process		Yes
9. Explicit assessment of outcomes		Yes
10. External quality measure (MMAT)		Strong
11. Activities evaluated	Project as a whole- consisted of training school supervisors, support to develop 4 year school development plans, support and mentoring	Strong
12. Lessons learned re: education, aid to education, evaluations	Excellent example of a mixed-methods approach –	
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	Literacy Boost Malawi: Year 2 Report	
Author/Agency	Save the Children / Amy Jo Dowd & Francis Mabeti	
Date published	2011	
Item	Description	Strong/Moderate/Weak
1. Relevance	Impact evaluation of literacy program run by Save the Children in Malawi	Strong
2. Program description	Literacy Boost – teacher training program to strengthen pedagogical methods, community action activities (reading camps, reading buddy programs, community literacy festivals). However, these activities are not described in any detail	Weak
3. Evaluation objective	Estimate the impact of a literacy program on reading skills, teacher outcomes (lesson planning, delivery) and community outcomes (project activity, support for education)	Strong
4. Approach	Quantitative – baseline and end line comparison, with comparison group (difference-in-differences)	Strong
5. Rigor	Probably strong – but methods section does not describe in detail	Moderate
6. Target audience	Save the Children practitioners, policy makers	Strong
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	No	N/A
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure (MMAT)		Moderate
11. Activities evaluated	teacher training, community activities to support literacy development (not described in detail)	Moderate
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	Mid-term evaluation of the Inclusive Quality Pre-Primary and Primary Education for Roma/Egyptian Children Project	
Author/Agency	Save the Children Albania	
Date published	2011	
Item	Description	Strong/Moderate/ Weak
1. Relevance	Somewhat relevant – but focus on pre-primary, not basic education	Moderate
2. Program description	This project aims to encourage the Roma/Egyptian students in schools and kindergartens to match peers achievement levels – through in class instruction in culture/identity, tutoring, recreational activities, literacy classes outside of school, parental sessions, child friendly environment development in schools	Moderate
3. Evaluation objective	Assess progress and make recommendations	Strong
4. Approach	Qualitative – focus groups, interviews, quantitative – questionnaire to 100 parents	Moderate
5. Rigor		Moderate
6. Target audience	Save the Children practitioners, policy makers	Strong
7. Participatory evaluation	Lots of emphasis on perceptions, barriers, among actors	Moderate
8. Explicit assessment of process	Yes	Strong
9. Explicit assessment of outcomes	Yes	Weak
10. External quality measure (MMAT)	Lacks discussion of limitations, lacks internal/external validity discussion, not necessarily replicable	Weak
11. Activities evaluated		
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	Swedish Support to the Education Sector in Mozambique	
Author/Agency		Sida
Date published		2004
Item	Description	Strong/Moderate/ Weak
1. Relevance	Relevant for post-conflict, study is a bit dated. Very descriptive but perhaps useful.	moderate
2. Program description	Broad overview	moderate
3. Evaluation objective	Seeks a long-term perspective to consolidate the results achieved	moderate
4. Approach	Overview of Swedish support to the education sector in Mozambique from 1976 to 2004, focused on 1994-2004. Interviews of indivduals involved with Swedish aid to education in Sweden and Mozambique; examined archives; documentary analsyis	moderate
5. Rigor	very descriptive	moderate
6. Target audience	Policymakers	moderate
7. Participatory evaluation	7 local stakeholders interviewed	moderate
8. Explicit assessment of process	limited	weak
9. Explicit assessment of outcomes	Swedish aid has contributed to the outcome that Mozambican primary school students have textbooks in all subjects	moderate
10. External quality measure		moderate
11. Activities evaluated	The evaluation looks at Swedish support to education, through (1) jointly funded pool for sector plan implementation (with other donors); (2) providing textbooks to primary schools; (3) improving sector management and administration, particularly in terms of decentralization.	moderate
12. Additional aspects that make the study worthy of inclusion		post-conflct situation
13. Lessons learned re: education, aid to education, evaluations		
14. Utility		Sida study

Title	Evaluation and Monitoring of Poverty Reduction Strategies – 2005- Budgeting for Education: Bolivia, Honduras and Nicaragua	
Author/Agency		Sida
Date published		2005
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation and monitoring of budgeting for education	Strong
2. Program description	alignment of poverty reduction strategy to achieve education MDGs; dpolicymakers etermined the best way to achieve this is via output-oriented budgets.	Strong
3. Evaluation objective	Needs assessment ("human, physical, and financial resources") to estimate "cost of achieving MDGs", (p. 5); measures current education sector achievements and conducts a cost-effectiveness analysis.	Strong
4. Approach	CBA that empirically "treats school enrolment as a function of educational costs and of various schooling inputs," (p. 11).	Strong
5. Rigor		Strong
6. Target audience	policymakers	Moderate
7. Participatory evaluation	stock-taking of local actors through visits to several municipalities in the three countries.	Moderate
8. Explicit assessment of process	"Household survey data and appropriate econometric methods were used to estimate the empirical model and to identify the effect of school costs and of schooling inputs," (p.11).	Strong
9. Explicit assessment of outcomes		Strong
10. External quality measure	simulation model and case studies, "The case studies on cost-effectiveness analysis and result-oriented budgeting presented in this report build on the methods and framework developed by Gertler and Van Der Gaag (1988), Gertler and Glewwe (1990) and applied, among others by Bedi and Marshall (1999), Bedi et al. (2004) and Vos and Ponce (2004)," (p. 11).	Moderate

11. Activities evaluated	monitor and evaluate the PRSP processes in the three Latin America countries eligible for debt relief: Bolivia, Honduras and Nicaragua. The study will be carried out over a period of 5 years, beginning in 2003	Moderate
12. Additional aspects that make the study worthy of inclusion	Interesting discussion on limitations of simulation model – indicates need to look at demand-side variables.	
13. Lessons learned re: education, aid to education, evaluations	Net primary school enrolment rates have increased, yet, need progress on quality. Cost-effectiveness analysis illustrates that reaching the MDG of 100% net primary enrolment in Bolivia, Honduras and Nicaragua, is impossible "using only one or more of the education policy instruments considered in the enrolment models estimated for these countries. This suggests that apparently one also has to look at demand-side variables - in particular the reduction of poverty - to reach the goal of universal primary education," (p. 5-6).	
14. Utility	Sida study	

Title	Sida's contributions 2006: Progress in educational development	
Author/Agency		Sida
Date published		2007
Item	Description	Strong/Moderate/Weak
1. Relevance	Very relevant – objective is to describe and analyse the results of Sida's cooperation in education in 2006.	strong
2. Program description	Sida cooperates bilaterally with 16 countries in education and also supports various UN agencies (UNICEF, UNESCO, UNGEI) and the World Bank's Fast Track Initiative (FTI).	strong/moderate (country by country
3. Evaluation objective	Review of Sida's progress in educational development	moderate
4. Approach	Primarily documentary analysis	moderate
5. Rigor		moderate
6. Target audience	Sida	
7. Participatory evaluation	No	weak
8. Explicit assessment of process	Yes	strong
9. Explicit assessment of outcomes	Yes	moderate
10. External quality measure	Yes	moderate
11. Activities evaluated	Swedish aid to education	strong
12. Additional aspects that make the study worthy of inclusion		
13. Lessons learned re: education, aid to education, evaluations	Points out the challenges of finding balance between support to various subsectors of education. Also indicates "The shift from project support to sector and budget support puts technical issues regarding aid modalities at the forefront," (p. 11).	
14. Utility	Good general overview of Swedish aid to education, but quite general.	

Title	Swedish Support in the Education Sector in Zanzibar, 2002 - 2007	
Author/Agency	Sida (Wort, M., Sumra, S., Schaik, P. Mbasha, E.)	
Date published	2007	
Item	Description	Strong/Moderate/ Weak
1. Relevance	Assessment of Sida's work in education in Zanzibar (SWAp)	Strong
2. Program description	"Program for emergency support" – classroom construction, refurbishing, "Capacity development – overseas training for education professionals"	Strong
3. Evaluation objective	Assess the relevance, effectiveness, efficiency and sustainability of the Swedish support to the Zanzibar education sector – make recommendations for continued Swedish support	Strong
4. Approach	Document review, interviews w/education officials, focus group	Moderate
5. Rigor		Moderate
6. Target audience	Sida officials	Strong
7. Participatory evaluation	Limited participation – via focus groups and interviews with "stakeholders"	Moderate
8. Explicit assessment of process	Description but with limited details	Weak
9. Explicit assessment of outcomes	Outputs – schools constructed, teachers trained, etc.	Weak
10. External quality measure		Weak
11. Activities evaluated	School construction, and Zanzibar Education Development Program (capacity development, monitoring and information systems)	Moderate
12. Lessons learned re: education, aid to education, evaluations		
13. Utility	Useful for Sida policy makers considering future funding to Zanzibar, less useful for our purposes.	
14. Additional aspects that make the study worthy of inclusion	Some mention of previous evaluations, some perspectives on capacity development (and the limits of sending education officials overseas to complete education, rather than focusing on in-country efforts)	

Title	Are Sida Evaluations Good Enough? An Assessment of 34 Evaluation Reports	
Author/Agency		Sida
Date published		2008
Item	Description	Strong/Moderate/ Weak
1. Relevance	Useful for our project, though not explicitly focused on education	moderate
2. Program description	transitioning to results based management	strong
3. Evaluation objective	to enhance the quality of Sida evaluations.	strong
4. Approach	Questions: "Do Sida evaluations produce information on processes and results that is comprehensive and detailed enough in view of Sida's management needs and reporting requirements? Are findings, conclusions and recommendations well supported by reported evidence? Do the evaluations produce lessons that are useful for learning and improvement beyond the evaluated projects and programmes?" (p. 5).	moderate
5. Rigor	"The assessment focuses on the following issues: • the quality of the Terms of Reference (TOR) for the evaluations and the extent to which the evaluation reports adequately responds to those TOR; • the quality of the design of the evaluation, including its data collection methods; • the quality of the information on results and implementation; • the quality of conclusions, recommendations and lessons learned," (p. 6).	strong
6. Target audience	Sida (no other stakeholders)	weak
7. Participatory evaluation	"this is a desk study and has nothing to say about the actual reception and use of the evaluation by its stakeholders. As use is an important quality criterion for evaluation processes, this is an important limitation," (p. 6).	weak
8. Explicit assessment of process	desk study, information about the actual evaluation processes is limited. The conclusions are based on the final reports and supplementary information about costs (all provided by Sida).	moderate

9. Explicit assessment of outcomes	Most of the evaluations in the sample addressed the questions raised in the TOR, though they did not necessarily provide satisfactory answers. The TOR were not always clearly formulated and focused, however.	strong
10. External quality measure	"For each of the issues addressed there was a set of quality criteria against which the reports could be systematically rated. The rating was done by the team of external evaluators and evaluation specialists who had also defined the criteria. Each of the reports was read by at least two of the team members and the results were discussed one report at a time in the wider group. The resulting assessments thus represent the reflected collective opinion of the rating team," (p. 6).	strong
11. Activities evaluated	Evaluations – this evaluation was supposed to be step one of a larger study that would also examine actual use of the evaluation instrument across countries; yet due to budget cuts and staff shortages, the second part of the study was eliminated.	
12. Additional aspects that make the study worthy of inclusion	" 'Learning' is one of the main purposes of evaluation. The 'lessons learned' section in an evaluation report is meant to present new insights that are relevant to a wider audience than the immediate stakeholders. Lessons learned are supposed to generalise and extend the findings from the intervention under study, either by considering it as an example of something more general or by connecting it to an ongoing discourse. This requires familiarity with both the international development debate and the discipline or sector under study and may not be possible or even necessary in all cases. The degree of generalisation may also vary from case to case. For all that, it is surprising that only 26% of the evaluation reports contain a section on lessons learned, and it is a cause for concern that the sections that where available are so weak. Only four reports were found to make strong contributions to the understanding and knowledge of development cooperation. ," (p. 9).	
13. Utility	useful reference document but maybe not for synthesis since it is not education-focused.	
14. Additional aspects that make the study worthy of inclusion	The report indicates that quality of Sida evaluations should be much better.	

Title	Policy Guidance and Results-Based Management of Sida's Educational Support	
Author/Agency		SIDA
Date published		2008
Item	Description	Strong/Moderate/Weak
1. Relevance	Very relevant – Results-Based Management and Sida's Educational Support	Strong
2. Program description	RBM for Sida's Educational Support	Strong
3. Evaluation objective	Examines strengths and weaknesses of the entire management process in the educational sector	Strong
4. Approach	documentary analysis, surveys, interviews, to assess steering instruments, results information from M&E, and evaluation instruments, as well as organisational conditions influencing actual use of information on policy and results.	Strong
5. Rigor		Strong
6. Target audience	Sida	Moderate
7. Participatory evaluation	Despite increased capacity support in M&E, limited links between information on results and the change in the design and implementation of programs etc. Also, lack of use of pilot study results.	Weak
8. Explicit assessment of process		Strong
9. Explicit assessment of outcomes	The ealuation notes that educational quality persists and is oftentimes insufficiently measured.	Moderate
10. External quality measure	"The methodology and approach included an attempted survey sending out questionnaire by e-mail. In the event, the response to questionnaires (only 7 returns – 20% response) was limited and the questionnaire findings were restricted to a collation and analysis of informed comments from the respondents…extensive consultations were undertaken in Stockholm and in selected case countries," (p. 19).	Moderate
11. Activities evaluated	"The strengths and potential limitations of Sida guidance instruments and results information flows for education sector cooperation, especially user relevance and	Strong

	assess how current organisational conditions, especially systems and processes, influence usefulness and effectiveness," (p. 9).	
12. Additional aspects that make the study worthy of inclusion	"Basic preconditions for results-based management are lacking in the educational sector. An overall conclusion is that management in the education sector is based on blueprint formats rather than a systematic use of policy instruments or information on results," Stefan Molund, Acting Director, Dept for Evlauation, p. iii.	
13. Utility	Useful discussion of evaluation at Sida	
14. Additional aspects that make the study worthy of inclusion		

Title	Gender equality in and through education	
Author/Agency	Sida/ Karlsson, P., Sjostedt, M., Johansson, C. Swedish Agency for Development Evaluation (SADEV)	
Date published	2010	
Item	Description	Strong/Mode-rate/Weak
1. Relevance	Gender focus is relevant – considering gender equity is a Sida priority – BUT the document does not explicitly evaluate a specific aid-funded program.	Strong
2. Program description	No specific program evaluated – rather, overall development cooperation and education systems in Afghanistan, Bolivia, Bangladesh, Cambodia, Ethiopia, Kosovo, Tanzania	Moderate
3. Evaluation objective	Objective: how can gender equality be promoted through Swedish bilateral support to education – what factors are important in promotion of gender equity, how does Sida use dialogue w/actors involved, how is gender equity promoted through capacity development	Strong
4. Approach	Document analysis, interviews (phone), and case studies (interviews w/aid officials, government officials)	Moderate
5. Rigor	Overview	Weak
6. Target audience	Presumably Sida – not defined	Moderate
7. Participatory evaluation	Yes – interviews with aid officials and education officials in recipient countries	Moderate
8. Explicit assessment of process	Limited assessment of the role of development cooperation in promoting gender equity	Moderate
9. Explicit assessment of outcomes	Yes – but not necessarily linked to aid funded education programs	Weak
10. External quality measure		Weak
11. Activities evaluated		Weak
12. Lessons learned re: education, aid to education, evaluations	Some information regarding the "dialogue dilemma" – the "delicate balance between promoting ownership and conducting a 2-way dialogue while still promoting Swedish specific priorities"	
13. Utility		
14. Additional aspects that make the study worthy of inclusion	Relevant because of discussion of aid dialogue/donor coordination, but analysis is weak and more provides general overall, less critical attention to link between aid to education and process/outcomes	

Title	Review of Sida-funded Project Education for Sustainable Development in Action (ESDA)	
Author/Agency	Sida/ Devine, V., Erikkson, R., Sida (InDevelop)	
Date published	May, 2012	
Item	Description	Strong/Moderate/Weak
1. Relevance	Aid funded evaluation program –but not focused on improving quality of basic education, the focus is on environmental education in Ukraine	Weak
2. Program description	Objective: introduction and dissemination of sustainable development into school curriculum	Moderate
3. Evaluation objective	Objectives: assess achievements of ESDA, success factors, weaknesses, recommendations for further funding/follow-up activities	Moderate
4. Approach	Document analysis, interviews, observations/participation in workshops and conferences	Moderate
5. Rigor		Moderate
6. Target audience	Sida and Sida partners	Strong
7. Participatory evaluation	Yes – program coordinators in country	Moderate
8. Explicit assessment of process	Brief discussion of institutional arrangements and management practices	Moderate
9. Explicit assessment of outcomes	Focus is on outputs and some intermediate outcomes – e.g. –decreased energy consumption in schools	Moderate
10. External quality measure		Weak
11. Activities evaluated		Moderate
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	Evaluation of the Barbro Johansson Model Girls' Secondary School in Tanzania	
Author/Agency	Dastgeer, A., Sumra, S., Cristoplos, I., Rothman, J. / Sida	
Date published	February 2013	
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of a single girls' boarding school in Dar es Salaam – funded by Government of Sweden.	Moderate
2. Program description	Boarding school w/40 teachers and 600 students – "a testimony to the friendship between Sweden and Tanzania"	Moderate
3. Evaluation objective	Examine the process the school has made towards its original objectives of providing high quality education for girls	Strong
4. Approach	Document review, interviews with officials in Dar es Salaam and Stockholm, phone interviews with students and their families, PTA, teachers, headmistress, etc.,	Strong
5. Rigor		Moderate
6. Target audience	Sida	Strong
7. Participatory evaluation	Yes - students and their families	Moderate
8. Explicit assessment of process	Some attention to educational processes at the school – teacher retention for example	Moderate
9. Explicit assessment of outcomes	- Mostly focused on evaluating the organizations trends in education assessments	Moderate
10. External quality measure		Moderate
11. Activities evaluated	Evaluation of one all girls' boarding school on Dar es Salaam	Moderate
12. Lessons learned re: education, aid to education, evaluations		
13. Utility	Would be useful to Sida in deciding whether or not to continue funding the particular boarding school – not terribly useful for our purposes	
14. Additional aspects that make the study worthy of inclusion		

Title	Swedish Development Cooperation in Transition? Lessons and Reflections from 71 Sida Decentralized Evaluations (April 2011 – April 2013)	
Author/Agency	Sida (conducted by InDevelop)	
Date published	2013	
Item	Description	Strong/Moderate/Weak
1. Relevance	Joint initiative between Sida and Indevelop to draw lessons from evaluations – relevant for Sida strategic decisions and operations	Strong – not education focused, but of relevance for aid/evaluation and Swedish context
2. Program description	N/A	N/A
3. Evaluation objective	Contribute to "evidence-based learning, improve Swedish development cooperation"	Strong
4. Approach	Reviews of evaluations, synthesis, some quantitative analyses, mostly qualitative.	Strong
5. Rigor	Methods are similar to our approach	Strong
6. Target audience	Sida, Sida's partners, development practitioners, international community	Strong
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	Yes	N/A
9. Explicit assessment of outcomes		Moderate
10. External quality measure		Moderate
11. Activities evaluated	Evaluations of Sida-funded aid projects	Strong
12. Lessons learned re: education, aid to education, evaluations	4 main success factors for achieving results: committed and engaged individuals and organizations, professionalism and high levels of competency w/in partner organizations, program developed through a political and economic needs/feasibility assessment, ownership and political will	
13. Utility	Very useful for our interest in evaluations and their use	
14. Additional aspects that make the study worthy of inclusion		

Title	Evaluation of Implementation of ICT in Teachers' Colleges Project in Tanzania	
Author/Agency	Anderson, B., Ngemera Nfuka, E., Sumra, S., Uimonen, P., Pain, A. / Sida InDevelop)	
Date published	May, 2014	
Item	Description	Strong/Moderate/Weak
1. Relevance	Objective: improve quality of education in teachers colleges by integrating ICT in teacher education	Strong
2. Program description	Teacher colleges provided with ICT equipment, internet connection, tutoring in ICT for tutors at teachers colleges	Moderate
3. Evaluation objective	"Determine what has been achieved, what lessons have been learned during program implementation, establish what can be approved in ongoing implementation" (p. 17)	Strong
4. Approach	Quantitative and qualitative – surveys and interviews, as well as administrative data	Strong
5. Rigor	Interviews, workshops, survey data – relatively limited in scope and analysis	Moderate
6. Target audience	Not identified, presumably Sida	Moderate
7. Participatory evaluation	Yes – interviews and surveys with teachers participating in the program	Strong
8. Explicit assessment of process	Yes	Strong
9. Explicit assessment of outcomes	Yes – short term outcomes (use of computers	Weak
10. External quality measure (MMAT)	No detailed information on methodology nor limitations	Weak
11. Activities evaluated	ICT teacher training program in teachers' colleges– could be described better	Moderate
12. Lessons learned re: education, aid to education,	Evaluation includes attention to the relevance of the program for multiple constituencies/sectors: Tanzanian gov't, Tanzanian education sector, ICT development in the country	

evaluations	
13. Utility	The issues identified primarily have to do with problems in internet connection, problems distributing computers, etc. This information could be useful to program coordinators/directors.
14. Additional aspects that make the study worthy of inclusion	This project came to be recognized as “best practice” in e-learning in Africa, according to the document – due to the projects’ recognition of the key role of teachers. Evaluation includes attention to sustainability and alignment with national and Sida goals

Title	**Lessons and Reflections from 84 Sida Decentralised Evaluations 2013 – a Synthesis Review**	
Author/Agency	Cristoplos, I., Hedqvist, A.L., Rothman, J. /Sida (InDevelop)	
Date published	2014	
Item	Description	Strong/Moderate/Weak
1. Relevance	Not about education	Weak
2. Program description	N/A – synthesis evaluation	N/A
3. Evaluation objective	Objective: Analyze and summarize conclusions drawn from Sida evaluations – in all sectors	Strong
4. Approach	Qualitative – document analysis	Strong
5. Rigor	Document analysis – no attention to how these evaluations were used	Moderate
6. Target audience	Primary intended user – Sida, secondary intended user – Sida's cooperation partners/development practitioners	Strong
7. Participatory evaluation	No	Weak
8. Explicit assessment of process	N/A	N/A
9. Explicit assessment of outcomes	N/A	N/A
10. External quality measure		Moderate
11. Activities evaluated	Evaluations of Sida programs – multi-sector	N/A
12. Lessons learned re: education, aid to education, evaluations	Explicit attention to "lessons learned in evaluation" – The recommendations in this report could be used as a lens through which to analyze other evaluations/program.	
13. Utility	Report is directed at aid officials and policy makers in donor countries	
14. Additional aspects that make the study worthy of inclusion	Worth including because of discussion of the weak interpretation and recommendations of "capacity development" findings from evaluations, critique of lack of coherence between poverty reduction/inclusion and evaluations, limits in assessing "effectiveness"	

Title	Child Friendly Schools Programming: Global Evaluation Report	
Author/Agency		UNICEF
Date published		2009
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of UNICEF's "child-friendly schools" programming strategy	Strong
2. Program description	Implementation of CFS globally (description of implementation is lacking)	Moderate
3. Evaluation objective	Assess how CFS models have been implemented in multiple contexts to improve education quality, assess extent of success in achieving CFS principles of child-centeredness, inclusiveness, and democratic participation	Strong
4. Approach	Desk review of CFS documents from all regions, site visits to 6 countries (surveys, observations, interviews, photos and videos, focus groups), online survey	Strong
5. Rigor	Methods are well aligned with approach – not impact evaluation – process evaluation	Strong
6. Target audience	UNICEF, Governments	Strong
7. Participatory evaluation	Not explicitly	N/A
8. Explicit assessment of process	Yes	Strong
9. Explicit assessment of outcomes	Not explicitly	Moderate
10. External quality measure (MMAT)	Methods are replicable, limitations discussed, findings linked to sources/methods	Strong
11. Activities evaluated	Varies by country – comprehensive approach to CFS	Moderate
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	Evaluation of Government of Tanzania and UNICEF Interventions in 7 Learning Districts	
Author/Agency	JIMAT Development Consultants, Ifakara Health Institute for UNICEF & Gov't of Tanzania	
Date published	July 2013	
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of UNICEF's country programming in Tanzania	Strong
2. Program description	Basic Education & Life skills component: increase primary and pre-primary schooling & transitions to secondary/post-primary: child friendly schooling, HIV/AIDS life-skills education, quality education through support & protection for vulnerable children, "accelerated primary education opportunity," "accelerated secondary education"	Strong
3. Evaluation objective	Assess effectiveness of area-based programming approach, the theoretical model, and draw lessons for future programming – focus on DAC criteria	Strong
4. Approach	HH survey in 7 participating districts, with matching survey in comparison districts – difference-in-difference model, plus focus groups w/students, community members	Strong
5. Rigor	Detailed description of methods –sampling procedure, limitations	Strong
6. Target audience	Policy officials and program directors	Strong
7. Participatory evaluation	Focus groups	Moderate
8. Explicit assessment of process		Weak
9. Explicit assessment of outcomes	Mostly short –term outcomes. Attribution issues exist and are acknowledged	Strong
10. External quality measure		Strong
11. Activities evaluated	Description of the program is not as strong	Weak
12. Lessons	Evaluation asks: how efficient was the coordination in fund	

learned re: education, aid to education, evaluations	disbursement, how did the use of national systems contribute to or hinder the objectives, how did field monitoring ensure quality and program delivery?
13. Utility	
14. Additional aspects that make the study worthy of inclusion	One of the few studies to use a quasi-experimental quantitative method.

Title	2012 Democratic Republic of Congo: Evaluation du programme Ecole et Village Assainis	
Author/Agency		UNICEF
Date published		December 2011
Item	Description	Strong/Moderate/Weak
1. Relevance	might be a little less interesting for this synthsis (education-focused) because of strong health component, but interesting to look at for intersectoral approaches	Weak
2. Program description	11 Provinces in Congo: to ensure child health and development by increasing access to potable water, and to improve sanitation and education in terms of hygiene practices.	Strong
3. Evaluation objective	To evaluate the activities, the processes, and results of the program. Additionally, the evaluation aims to provide UNICEF and the Congolese government recommendations for an eventual collaboration or program in the future, within the context of Basic Education.	Strong
4. Approach	mixed-methods: interviews, documentary analysis, field visits where questionnaires and semi-structured interviews were administrated in the provinces, a case study of one of the provinces, semi-directed interviews with various directors and program partners,	Strong
5. Rigor	Diverse sources were consulted and data was triangulated, with an on-the-ground perspective. Triangulation helped to verify and sort the most pertinent issues, with the specific expertise of the consultants, available documentation, and information provided by respondents.	Strong
6. Target audience	UNICEF, Congolese government, partner implementing agencies	Strong
7. Participatory evaluation	Yes—with actors at all levels.	Strong
8. Explicit assessment of process		Strong
9. Explicit assessment of outcomes		Strong
10. External quality measure		Strong
11. Activities evaluated	The planning context of the program and its results and impact, with special attention to	Strong

	the implementation context, (strengths, weaknesses, and constraints).	
12. Additional aspects that make the study worthy of inclusion	Recommendation of evaluation: need to quickly reinforce capacity of CSO leaders in terms of documentation and knowledge management tools; as well as involve school inspectors and principals in the implementation of the sub-program.	

Title	Evaluation of UNICEF's role as a Lead Partner in the education sector in Sierra Leone	
Author/Agency	UNICEF/Anna Haas (independent consultant)	
Date published	July 2012	
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of UNICEF's role as "Lead Partner" in the coordination of education sector – focus is on the aid relationship, not impact on education.	Strong
2. Program description	Education Sector Plan 2007 – led by UNICEF	Moderate
3. Evaluation objective	Formative evaluation – assess the performance of UNICEF as Lead Partner	Strong
4. Approach	Main source: interviews with 22 actors (Ministry of Education, multilateral agencies, NGOs), review of documents, observations from the 2012 education sector review	Moderate
5. Rigor		Strong
6. Target audience	UNICEF and other coordinators of the education sector, Government of Sierra Leone	Strong
7. Participatory evaluation	Not explicitly	N/A
8. Explicit assessment of process	Yes	Moderate
9. Explicit assessment of outcomes	No	N/A
10. External quality measure (MMAT)		Strong
11. Activities evaluated	Coordination – leading of educational development in Sierra Leone from 2007 – 12	Strong
12. Lessons learned re: education, aid to education, evaluations		
13. Utility	One of the few evaluations to focus on a careful assessment of the aid agency's role in educational development	
14. Additional aspects that make the study worthy of inclusion		

Title	Evaluation of the Girls Education Project of the Forum for African Women Educationalists – The Gambia (FAWEGAM)	
Author/Agency	UNICEF/Adelaide Sosseh (independent consultant)	
Date published	August 2012	
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of efforts to improve gender equity in educational outcomes	Strong
2. Program description	Not clear described – mix of advocacy and programmatic efforts	Moderate
3. Evaluation objective	Enable FAWEGAM to to "build on its strengths, minimize weaknesses, overcome constraints"	Moderate
4. Approach	Desk reviews, focus groups and interviews with actors – mothers and girl students, UNICEF and FAWEGAM officials, teachers, etc	Moderate
5. Rigor	Findings not necessarily linked to data/methods, weak analysis in parts	Weak
6. Target audience	UNICEF, FAWEGAM, education officials	Strong
7. Participatory evaluation	Not explicitly	N/A
8. Explicit assessment of process	Yes	Moderate
9. Explicit assessment of outcomes	Yes	Weak
10. External quality measure (MMAT)		Weak
11. Activities evaluated		
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	External evaluation of the "For Safe and Enabling School Environment" Project in Croatia	
Author/Agency	UNICEF / IVO Pilar Institute of Social Sciences	
Date published	2012	
Item	Description	Strong/ Moderate/Weak
1. Relevance	Evaluation of program designed to reduce school violence	Moderate
2. Program description	Public campaign to raise awareness of peer violence among boys and girls, school project to promote working /living conditions ins schools that nourish tolerance and respect, create protective network in communities	Strong
3. Evaluation objective	Ex-post evaluation of the program – assess implementation, evaluate the role and contribution of impact of program	Strong
4. Approach	Mixed methods – comparison and treatment groups of 10 schools (non-randomly selected) – completed questionnaire examining behavior, knowledge, skills, competencies, and qualitative study (community, parents, students,)	Moderate
5. Rigor	Major limitations – but these are acknowledged	Moderate
6. Target audience	UNICEF, education officials	Moderate
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	Yes	Moderate
9. Explicit assessment of outcomes	Yes	Weak
10. External quality measure (MMAT)		Weak
11. Activities evaluated	"Whole school" approach to promoting safe spaces, training teachers, raising awareness, training students in social and emotional skills, etc.	Moderate
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	Independent Evaluation of Program: Improving Access to Quality Basic Education in Myanmar (2006-2010)	
Author/Agency		UNICEF
Date published		
Item	Description	Strong/Moderate/Weak
1. Relevance		strong
2. Program description	Program aimed to scale up interventions already piloted and locally implemented (Childhood Development (ECD), Child Friendly Schools (CFS) and Life Skills Education (LSE)).	strong
3. Evaluation objective	To assess performance (relevance, efficiency, effectiveness) and suggest modifications if necessary.	strong
4. Approach	Qualitative, document review and a rapid situation analysis of the education sector.	strong/
5. Rigor	Meetings and focus group discussions were held in-country with key stakeholders including UNICEF staff in Yangon and in the field. Field visits were made to a selected sample of target beneficiaries, to observe, and conduct interviews/in-depth interviews/focus group discussions.	strong
6. Target audience	UNICEF, donor, and NGO stakeholders	moderate
7. Participatory evaluation	UNICEF, donor and NGO stakeholders—yet list of interviewees is in annex 2 which is not published online	moderate/strong
8. Explicit assessment of process		strong
9. Explicit assessment of outcomes	Evaluation indicates a lack of an exit strategy	strong
10. External quality measure	No mention of local stakeholders, but might be in annex 2 that is not online	moderate
11. Activities evaluated	educational management and various programs (scaling up)	

12. Lessons learned re: education, aid to education, evaluations	Here, more and more on educational management. The evaluation indicates that eighteen programme interventions had limited impact on educational quality. There has not been a capacity building needs assessment to address issues of quality (seems to be recommended by the evaluators). The evaluation maintains that programs for 0-3 year olds should not be part of education sector strategy but within a multisectoral response to child development (for instance, Ministry of Social Welfare).
13. Utility	Findings and conclusions: Implemented in a context without a comprehensive sector plan. Recommendations regarding M&E limitations: Though M&E very acknowledged in original proposal, there was a lack of consistant indicators. Multiple steps to address M&E challenges during implementation but was "too complicated" for UNICEF to manage and as a result a great deal of the data obtained was not analysed. Overreliance on "a large-scale survey to measure changes in school practices without any triangulation using qualitative research methods," (p. 3).
14. Additional aspects that make the study worthy of inclusion	

Title	Process and Impact Evaluation of the Basic Education Assistance Module (BEAM) in Zimbabwe	
Author/Agency	Smith, H., Chroro, P., Musker, P. /CfBT Education Trust, Impact Research International & Paul Musker and Associates /UNICEF	
Date published		2013
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation covers the BEAM strategy implemented by the Gov't of Zimbabwe – not an evaluation of aid funded program, although BEAM received some aid funding through the Child Protection Fund of the National Action Plan (NAP)	Moderate
2. Program description	BEAM expands access to primary and secondary school by paying tuition, levies, examination fees, and boarding fees - grants given to School Development Committees	Strong
3. Evaluation objective	Identify implementation gaps and inform future programming	Strong
4. Approach	Mixed methods – survey questionnaires in 352 schools, focus groups and interviews in 40 schools	Strong
5. Rigor		Strong
6. Target audience	Gov't of Zimbabwe and BEAM donors	
7. Participatory evaluation	Yes – school administrators, teachers, parents, students	Strong
8. Explicit assessment of process	Process assessment mostly includes perceptions (e.g. – "percent who say education access improved 'a lot,' 'a little,'" etc.) from the survey, limited in-depth analysis	Weak
9. Explicit assessment of outcomes	"" ""	Weak
10. External quality measure		Weak
11. Activities evaluated	School block grants to cover tuition, exam fees, and levies for vulnerable children– beneficiaries are identified by a school committee	Moderate

12. Lessons learned re: education, aid to education, evaluations	
13. Utility	
14. Additional aspects that make the study worthy of inclusion	Strong methods description, but analysis is relatively weak.

Title	Developing a local model for the delivery of primary education in Karkaar Region, (Somalia)	
Author/Agency	UNICEF - Save the Children – submitted to UNICEF, funded by DFID, UNICEF, UNESCO	
Date published	December, 2011	
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of aid-funded basic education program	Strong
2. Program description	Objective: Increase number of children accessing and completing inclusive, quality and protective basic education Through a local model for the delivery of primary education	Strong
3. Evaluation objective	Objective: Assess the performance of the project, using OECD/UNICEF evaluation criteria: relevance, effectiveness, efficiency, implementation process, coverage, coherence, impact, sustainability	Strong
4. Approach	Mixed methods – focus groups and interviews, school enrollment and retention trends using a simple random sampling approach	Strong
5. Rigor		Strong
6. Target audience	Not stated – presumably UNICEF/STC/Gov't	Moderate
7. Participatory evaluation	Yes – interviews with teachers, education officials	Moderate
8. Explicit assessment of process	Yes – but analysis is weak	Weak
9. Explicit assessment of outcomes	Yes – but causal validity is weak	Weak
10. External quality measure		Weak
11. Activities evaluated	Implementation of 3 year basic education project	Strong
12. Lessons learned re: education, aid to education, evaluations		
13. Utility	Example of an evaluation that may be useful to the project coordinator due to specific findings (in bulleted list) regarding technical and operational challenges, but has very limited value in terms of providing insights into education, aid, or evaluation	

14. Additional aspects that make the study worthy of inclusion	Poorly organized, weak analysis, but methodology is moderately strong, and could be included for diversity purposes

Title	Assessment of the USAID Assistance Program to the Reform of the Benin Primary Education System	
Author/Agency		USAID
Date published		2005
Item	Description	Strong/Moderate/Weak
1. Relevance	Evaluation of aid-funded support to the Benin education system	Strong
2. Program description	Reorganization of the primary education structure (new studyprogram) NSP, computerized management of school statistics and disaggregated data, development of a planning tool for school development, a system of financial management based on budgeted reforms, and community/school-based programs (including support for parent associations)	Strong
3. Evaluation objective	Assess the impact of USAID/Benin's assistance to date – identify strengths and weaknesses and areas for potential collaboration.	Strong
4. Approach	Primarily qualitative – evaluators met with key informant and focus groups with USAID and government officials, especially those directly responsible for design and implementation, school directors, teachers, parents, and school visits (observations)	Moderate
5. Rigor	Methodology and limitations acknowledged, process of data triangulation described	Moderate
6. Target audience	Policy makers, USAID	Strong
7. Participatory evaluation	Somewhat – interviews and observations, but weak	Weak
8. Explicit assessment of process	Somewhat	Moderate
9. Explicit assessment of outcomes	Somewhat – impacts as perceptions of, attitudes, beliefs	Moderate
10. External quality measure		Moderate
11. Activities evaluated	Teacher training programs, children's knowledge and earning, and the role of parents and communities in school management.	Strong
12. Additional aspects that make the study worthy of inclusion	Good description of each intervention and challenges encountered	

Title	Program Evaluation for USAID - Guinea Basic Education Program Portfolio	
Author/Agency		USAID
Date published		May 2006
Item	Description	Strong/Moderate/Weak
1. Relevance		strong
2. Program description	Guinea basic education + community-based interventions	strong
3. Evaluation objective	The evaluation examined the efficiency of a program to deliver quality basic education to a larger percentage of Guinean children with an emphasis on girls and rural children.	strong
4. Approach		strong
5. Rigor	In addition to interviews, the team also adapted a classroom observation tool developed by EDC to observe process of change over time. Yet, to test out the tool, to avoid replicating earlier EDC studies, site visitors wrote "field notes based on their observations of teacher practices, including interaction with students, the use of active teaching methods and student assessment techniques, the availability of pedagogical materials, and gender-related practices," (p. 5). The evaluation team noted a "strong emphasis on the collection and analysis of documentation relating to program implementation," (p. 5).	strong
6. Target audience	USAID	
7. Participatory evaluation	A multinational team of six researchers from Benin, Canada, Guinea, Senegal, and the United States conducted the evaluation research.	moderate
8. Explicit assessment of process		strong
9. Explicit assessment of outcomes		strong
10. External quality measure	As in most other evaluations, no mention of how findings relate to researchers' influence, for ex, through their interactions with participants	strong/moderate
11. Activities		strong

evaluated		
12. Lessons learned re: education, aid to education, evaluations	"Decentralization of planning and decisionmaking has been met with relative success, although devolution of budgetary authority has proven more difficult to implement," (p. viii) The evaluation notes a positive impact of community participation (re: access and quality), yet, warns that this impact is fragile since it may generate a demand for education that cannot be met. Additionally, the report notes that while there has been progress regarding gender and rural/urban gaps, it is challenging to isolate the reason for these impacts because of the multiplicity of interventions by the funding agency, national government, and civil society organizations.	
13. Utility		
14. Additional aspects that make the study worthy of inclusion	Interesting research questions, including on sustainability of "strategies, models, and approaches," (p. 4) for example, on effective support to civil society groups, the impact of community participation on education, and the program's approach and impact on intersectoral issues (for example, gender, rural/urban gaps, HIV/AIDS education).	

Title	Action Communautaire pour l'education des filles: Evaluation finale (2001-2005)	
Author/Agency		USAID/ World Learning
Date published		June 2005
Item	Description	Strong/Moderate/Weak
1. Relevance	very useful given Sida's objectives	Strong
2. Program description	a four-year USAID-financed project, piloted by World Learning, to promote girls' education in rural zones and to stimulate community participation to encourage school attendance, particularly for girls' education	Strong
3. Evaluation objective	To evaluate community action for girls' education program	Strong
4. Approach	mixed methods.	
5. Rigor	Limited and unreliable statistics, this carries through to the national level	Moderate
6. Target audience	Communities, implementing agencies	Strong
7. Participatory evaluation	Highly participatory, and included multiple stakeholders, emphasis on community-based interventions.	Strong
8. Explicit assessment of process		Strong
9. Explicit assessment of outcomes		Strong
10. External quality measure		Moderate
11. Activities evaluated	Community participation to promote girls education (via NGOs).	Strong
12. Additional aspects that make the study worthy of inclusion	Utility of NGOs in community-based approaches.	

Title	An Unfinished Agenda – An evaluation of World Bank Support to Primary Education	
Author/Agency	Independent evaluation group – World Bank	
Date published	2006	
Item	Description	Strong/Moderate/ Weak
1. Relevance	Evaluation of aid funded education programs – primary education	Strong
2. Program description		
3. Evaluation objective	Objective is to assess the overall effectiveness of World Bank assistance to countries in primary education	Strong
4. Approach	Literature reviews, review of WB documents, inventory and review of WB primary education portfolio, field-based evaluations of completed primary education in 8 countries, field-based country case studies in 4 (different) countries	Strong
5. Rigor		Strong
6. Target audience	Aid policy decision makers, implementers	Strong
7. Participatory evaluation	Limited – case studies included interviews with Bank and local managers, donors, agencies, beneficiaries.	Moderate
8. Explicit assessment of process	Yes – some attention to the modalities of aid giving and monitoring and evaluation	Strong
9. Explicit assessment of outcomes	Focus is on outcomes of individual projects, rather than on the overall WB efforts in primary education	Moderate
10. External quality measure		Weak
11. Activities evaluated	Management performance, decentralization, community control and accountability, teacher incentives, M&E, research	Strong
12. Lessons learned re: education, aid to education, evaluations	Mostly descriptive analysis of evolutions of WB funding, but some critical analysis of WB policies and process of education aid	
13. Utility		
14. Additional aspects that make the study worthy of inclusion		

Title	Bangladesh Education Sector Review: Seeding fertile ground: Education that works for Bangladesh	
Author/Agency		World Bank
Date published		September 2013
Item	Description	Strong/Moderate/Weak
1. Relevance	Overall sector review of Bangladesh education sector – not an evaluation of a particular policy/program	Moderate
2. Program description	N/A	N/A
3. Evaluation objective	Support an articulated, coherent policy dialogue on education and skills development	Moderate
4. Approach	Not described, but compilation of document review, data analysis from administrative/census/HH survey data	Moderate
5. Rigor		N/A
6. Target audience	Politicians and international aid community	Strong
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	No	N/A
9. Explicit assessment of outcomes	Yes	Moderate
10. External quality measure (MMAT)		Moderate
11. Activities evaluated	Sector status – "snap shot" of educational development, trends in enrollment, equity, management	N/A
12. Lessons learned re: education, aid to education, evaluations		
13. Utility		
14. Additional aspects that make the study worthy of inclusion	Rating: C (not relevant – not an evaluation)	

Title	What Really Works to Improve Learning in Developing Countries?	
Author/Agency	World Bank (David Evans and Anna Popova)	
Date published	2015	
Item	Description	Strong/Moderate/Weak
1. Relevance	Synthesis of 6 existing systematic reviews or meta-analyses of interventions designed to improve learning	Strong
2. Program description	N/A – meta-analysis of multiple interventions in low and middle-income countries	N/A
3. Evaluation objective	Demonstrate and explain the divergent findings between the 6 existing reviews	Strong
4. Approach	Synthesis – purposive sample of existing meta-analyses and synthesis reviews, then examination of main conclusions, exclusion rules, variation in composition and categorization of studies included, and heterogeneity across results within intervention categories	Strong
5. Rigor	Strong	Strong
6. Target audience	Academics, policy makers, aid community	Strong
7. Participatory evaluation	No	N/A
8. Explicit assessment of process	No	N/A
9. Explicit assessment of outcomes	Yes	Strong
10. External quality measure		Strong
11. Activities evaluated	Many different interventions	N/A
12. Additional aspects that make the study worthy of inclusion	This study is worth including in part as it serves to highlight the challenges of identifying “what works”	

E. Evaluations selected for high-priority attention

For each evaluation selected for in-depth review, we recorded the following information: *why selected* for in-depth review; *evaluation approach/method*; *major findings*, our own analysis regarding *observations* and *lessons learned* from the evaluation about education, aid, and evaluation.

3ie - Krishnarane, S., White, H., Carpenter, E.
Quality education for all children?
September 2013

Why selected for in-depth review:

3ie has gained recognition as a leader in foreign-aid evaluation and research, mostly for the organization's work funding randomized controlled trials and systematic reviews. This is one such systematic review.

Evaluation approach/method:

3ie's meta-analysis of "what works" in education in developing countries is based on an earlier systematic review undertaken by WestEd (Petrosino et al, 2012), in which educational projects are categorized as either *demand side interventions*: reducing costs (CCTs, scholarships, and non-fee subsidies, vouchers, abolishing school fees and capitation grants), providing information to parents and students, and increasing preparedness (early childhood development, health/nutrition), and *supply side interventions*: buildings, teachers, methods and management. To be included, evaluations had to use either an experimental (randomized controlled trial) or quasi-experimental method to identify a quantitative impact on a given educational outcome (enrollment, attendance, dropout, or progression).

Major findings:

Broadly, the conclusions are the following: demand-side interventions, like CCTs, school feeding programs, and vouchers, can increase enrollment and attendance, but spending more time in school does not automatically translate into improved learning. Likewise, early childhood development programs can have an important impact on future enrollment and cognitive development, but require high quality (well trained, well supported) early childhood educators. Programs designed to provide information to parents regarding the importance of schooling have the potential to be extremely cost-effective, but thus

far there is limited evidence in favor of these programs. Supply side interventions, such as building new schools, providing learning materials (textbooks, flip charts, chalkboards), or even hiring additional teachers, have been shown to improve learning, but only when these investments are accompanied by sufficient training and support for teachers and school communities. School-based management programs have been linked to improved test scores, but it is unclear whether this is from increased parental involvement in schools or from the additional resources that often accompany school-based management programs.

The evaluators hypothesize that programs that increase enrollment may lead to new challenges in the classroom, because the newly enrolled children often come from poorer or more disadvantaged backgrounds. However, none of the studies evaluated found that these children "dragged down" the performance of already enrolled children. They also note that few of the studies reviewed were directed at "difficult to reach populations," and argue that new approaches are needed to reach these groups.

Major observations:

- **On education in poor countries:** Little attention is paid to context – differences between results of similar programs in different countries are noted, but not analyzed nor dealt with explicitly (in most cases).
- **On aid-supported education activities:** This systematic review includes almost no mention of the role of the aid sector, despite the fact that many of the projects evaluated were likely at least partially funded by aid agencies.
- **On evaluating aid-supported education activities:** The contradiction between the need to find "what works," despite acknowledging that there is no "one-size fits all" stands out. The systematic review attempts to identify "what works" by pooling the effect sizes of different evaluations, yet the authors mention in the conclusion that "the broad aggregation across all different interventions is not useful as a guide to policy," and that "for the majority of the interventions studied in this review, there is simply not enough evidence available to determine their effectiveness."

What have we learned?

This evaluation, along with many others reviewed, supports the idea that the most vulnerable populations remain un-reached by most aid-funded education projects (the authors make this claim on page 44).

This evaluation also provides another example of an effort to quantify and compare impact measures across different contexts and programs, which contrasts with the widespread acknowledgement that there is no "silver bullet solution," and that context matters – a program that works in one context will not necessarily work in another.

French Agency for international Development and the World Bank
La Cooperation Française face aux defis de l'éducation en Afrique: l'urgence d'une nouvelle dynamique
2007
Why selected for in-depth review:

This policy document/analysis explores various strategies for improved aid efficiency on a global scale and more coherence in French interventions in terms of education cooperation. The analysis argues for a renewed, dynamic approach to French aid to education. Despite very significant positive results in expanding educational access, completion rates of primary school and educational quality remain a challenge in Africa. Access to secondary schooling largely remains within the realm of students who are more privileged, despite the economic need for more secondary school graduates to facilitate development.

This evaluation was selected for in-depth review also because international development literature conducted in English tends to focus on Anglophone countries. Due to historical relationships France is a large contributor of foreign aid, in particular to its former colonies in Africa. In particular, there is a high concentration of Francophone African countries in the lowest tier of the UNDP Human Development Index (HDI). There is a need therefore to include evaluation literature across languages.

This evaluation is useful because it highlights challenges to education in Africa and the shortcomings of development agencies and multilateral agencies in addressing these challenges, as aid has not met its objectives in the educational sphere. The evaluation focuses on Sub-Saharan Africa since this is the priority zone of French aid to education, given historical linkages.

Evaluation approach/method:

This strategy evaluation describes French development objectives in terms of aid to education and the limitations of France's educational programs, as well as a need for education financing to be on par with commitments made. The evaluation indicates that in genenal, for primary education, there has been a slight lack of emphasis on educational quality, and FTI has not been entirely implemented despite good progress. Moreover, this strategy document analyses the participation of France in terms of the second and third MDGs. The evaluation also lays out some of the challenges in post-primary education and professional training, as well as future directions, as EFA has resulted in downstream pressure on the educational system specifically to secondary education.

Major findings:

Despite significant improvements in education in Sub-Saharan Africa, the evaluation indicates that progress towards education development goals remain insufficient. Regardless of increased access to primary education, primary completion rates are still mediocre and vary largely from country to country and within country, depending on household income, gender, and rural and urban areas.

Starting in 2007, the evaluation notes that France started dedicating more aid to education development, especially eventually through FTI. Yet, the evaluators find that French aid to education is not on par with commitments.

The evaluation notes that the Pole de Dakar and the PASEC (Programme d'analyse des systèmes éducatifs de la Conférence des ministres de l'éducation des pays ayant le français en partage) have not successfully examined challenges to educational quality (which, according to the evaluators should focus on: the pedagogy and training of teachers; the role of parents and communities; and the role of educational leadership wthin the school). Therefore, the evaluation proposes a "Pole Qualité" (Quality Center) providing resources to recipient countries for teacher training, as well as resources to improve the school environment. The evaluation hypothesizes that a Quality Center would advance South-South cooperation and country-specific approaches, and facilitate diffusion of best practices.

Major observations:

On education in poor countries: The evaluation, notes that the wider international community has even recognized important links between education and development. The evaluators take this linkage even further by stating that it seems education influence health more than even health interventions, for example, the low rate of HIV/AIDS among youth that have completed secondary school.

On aid–supported education activities: The document critiques French development policy as being incomplete and fragmented in the education sector, particular as the strategy does not address post-primary education.

The evaluation notes that teacher trainers and those training school management officials are oftentimes the core of the education sector, yet remain unaddressed by French development policy.

Moreover, according to the document, French strategy marginalizes the question of language of instruction, since the Ministry of Foreign Affairs tends to promote use of the French language in Francophone countries, but in other non-Francophone, contexts, local languages are favored. The evaluators argue that this contradicts the notion of aid harmonization and national alignment.

The evaluation underlines the need for continued development aid in order for Sub-Saharan Africa to achieve its educational development objectives. Despite the new dynamic in aid to education following Jomtien (1999) Dakar (2000), the MDGs (2000), Monterrey (2002), the Fast-Track Initiative (2002), and the Paris Declaration (2005) international commitments to aid to education remain weak, according to the evaluators.

The evaluation indicates that FTI has been influential in promoting partnership to achieve MDGs, especially given the use of indicators as important measurement and financial mobilization tools. Yet, there have been very minimal efforts made in terms of measuring and achieving educational quality, and the report indicates a lack of international strategy that merges the objectives of increasing access and improving results, objectives which "should" go together, according to the evaluators.

On evaluating aid-supported education activities: The report advocates that all development partners utilize their comparative advantages, and outlines perceived added value in the sector (ex: France has notable strength in teaching quality). France developed a regional analysis

center in Dakar to facilitate analysis of education initiatives and to help elaborate national public policies within the sector, and the document proposes the creation of a quality center for education within Francophone Sub-Saharan Africa.

What have we learned?

This strategy document indicates that distribution of resources between educational levels varies greatly due to inadequate and inequitable policy reflection; particularly as most public resources benefit a small minority (usually within the highest socioeconomic group). Moreover, the report indicates that teaching practices and educational management are not adapted to the current context/development objectives, posing significant challenges to educational development in Africa.

While France contributes slightly more than one quarter of its bilateral foreign aid to education (2004 figures), the evaluation notes that the majority of this amount is in the form of funding students from developing countries to study in France or in French-run institutions in recipient countries. The evaluation indicates that in 2001, France set up the Dakar education sectoral center (Pole de Dakar) as a platform of expertise alongside the UNESCO regional office in Dakar, and shortly thereafter, has acquired competency in diagnostic studies, the development of instruments, and policy documents at the request of recipient countries. The Pole de Dakar publishes reports on the evolution of EFA and more recently, developed a distance-learning program for African education professionals. The Pole de Dakar also promotes post-primary education and intersectoral approaches in its sectoral analysis, and since 2007, covers countries outside of Francophone Africa as well. Building upon the initial agreement between France and UNESCO in setting up the Pole de Dakar, the report indicates a desire that the Pole develops institutional bridges with its principal partners and formalizes a network of exchanges on the functioning of African educational systems. This may be an interesting model for other funding agencies to consider or a resource for them to learn from.

The evaluation, however, does not say anything new, and echoes the larger education development literature. The use of regional analysis centers to facilitate monitoring and evaluation, as well as to help elaborate national public policies within the sector, as well as a proposed quality center, are perhaps potential ways forward suggested by this report that should be further reflected upon in terms of

addressing continuous challenges to educational quality. Follow-up questions could address the success of the "Pole Qualité" and the result of this earlier report on AFD education sector policy.

French Agency for international Development and the World Bank

L'enseignement post-primaire en Afrique subsaharienne: Viabilite financiere des differentes options de developpement
2010

Why selected for in-depth review?

This is not an evaluation but rather a comparative analysis of post-primary education in 33 sub-Saharan African countries, led by a team of academic researchers. The evaluation provides strong contextual analysis of challenges facing sub-Saharan education systems, especially in relation to financing.

Evaluation approach/method:

Methods include document analysis and official communications and statistical analysis of institutional data on educational participation and macroeconomic conditions from the World Bank and UNESCO. Analysis consists of projections of educational supply and demand through 2020.

Major findings:

Findings highlight the similarities and differences between the different challenges facing sub-Saharan African countries as the rate of primary school completion increases. The objective of the evaluation was to compare and contrast different strategies for achieving universal secondary education by 2020, considering forecasted macroeconomic, demographic and institutional conditions. The evaluators recommend that analysts to keep in mind: (1) the "dual structure" of sub-Saharan African economies (informal and formal sector), (2) the consequences of low secondary school completion in the labor market, (3) the importance of girls' education.

By focusing on finance, the study analyzes the volume of aid needed to address the "bottleneck" in secondary education, and the level of foreign aid dependence that is "acceptable," as well as the necessary political reforms to achieve quality secondary education for all.

Major observations:

The evaluation produces a number of recommendations, namely, that governments must devote 20% of the state budget to education (23% in many countries), but achieving universal secondary education by 2020 will also require substantial foreign aid investments.

Asian development bank - independent evaluation department Uzbekistan: education sector assistance program evaluation September 2010

Why selected for in-depth review:

The evaluation is well organized and well written, has a section devoted to analyzing the performance of the aid agency itself – rather than just the implementing agency—and also enhances the diversity of our sample, in terms of both the funding agency and the aid recipient country.

Evaluation approach/method:

This sector assistance program evaluation aims to assess the performance of Asian Development Bank (ADB) support to the education in Uzbekistan from 1992 – 2009. During this time period, ADB's aid to education in Uzbekistan totaled $290.5 million (approximately 23% of total ADB aid to the country) and included a basic education textbook development, support to senior secondary education, an ICT project, and a rural education project, along with technical assistance for national educational governance capacity development (mostly monitoring and evaluation). The evaluation includes a "strategic and institutional-level assessment (top down)" as well as a "project/program level assessment (bottom-up)" assessment. Both are based on document review and consultations (interviews) with national education officials. The top down components evaluate how well the ADB has responded to the country's needs, how the ADB has contributed to overall development in the country, and the ADB's performance as a lead funding agency. The bottom-up components assess the ADB's programmatic relevance, effectiveness, efficiency, sustainability, and impact.

Major findings:

The authors conclude that ADB education programming has been "successful" in improving educational access and quality and in supporting national educational governance. The evaluation notes the following overarching lessons regarding aid to education: (1) projects

that respond to government's priorities and have the government's commitment are most successful, (2) there is a disconnect between the knowledge and capacity of aid officials and national (government) officials, especially regarding capacity to collect and analyze quantitative data, (3) projects that entail substantial change at the school level – such as "student-centered learning methods," and "learning by doing" pedagogy, require a long lead-time in order to ensure core professionals understand these new methods, (4) there is a need for more realistic time frames and "in-depth review of the terms of reference of consulting firms by national counterparts."

Major observations:

- **On education in poor countries:** How to define notions of equity or "pro-poor" growth: Is it enough to say that a project is "inclusive" if it is targeted at both male and female children, or at rural children (who are more likely to be poor), for example (as this evaluation does)?
- **On aid-supported education activities:** Some specific findings stand out – although the specific sources to which these findings are linked are not mentioned – (1) ADB officials tend to focus on the relationship with the Education Ministry, but it is "also crucial to keep in mind other ministries" – namely, the Finance Ministry, planning agency, etc. (2) the tendency for the government and the aid agency alike to focus on inputs and outputs (rather than outcomes or impacts) is bemoaned in this evaluation as in many others, (3) the authors emphasize the importance of working with the government to establish a "road map," – that is, identifying and agreeing on desired outcomes and the aid agency's role in supporting the country's progression towards these outcomes *before implementing* a project
- **On evaluating aid-supported education activities:** The link between the conclusions and the data collected and analyzed is not clear. Little attention in the analysis section is paid to issues of attribution (e.g., can national enrollment increases really be attributed to ADB funded-projects?), nor to the performance of the ADB, as opposed to the performance of the government. The primary problems identified are a lack of technical capacity at the ministerial level for monitoring and evaluation and delays in hiring international consultants – the performance of ADB itself is considered to be positive because the projects supported are relevant to the country's priorities. Claims are often made

without any discussion as to what sources of data substantiate said claims.

What have we learned?

This evaluation, like many others, emphasizes the importance of national capacity development. It is assumed that low levels of national capacity (in particular in terms of monitoring and evaluation, data collection, and analysis) pose substantial barriers to educational development. This may be the case, but almost no evaluations critically assess *how* to sustainably develop national capacity, nor the role of the aid sector in helping – or hindering—national capacity development.

The evaluation notes that "slow growth of employment opportunities continues to be a major challenge for the education sector." This seems to be representative of a common practice across evaluations: unrealistic expectations about the link between educational investments and immediate, tangible economic growth.

Belgian Development Cooperation
Thematic evaluation of Belgian development cooperation in the education sector
August 2007

Why selected for in-depth review:

This quality of the description and analysis is among the highest. This evaluation is also unique in that it encompasses all Belgian actors in aid to education, including NGOs and research councils, not just Belgian Development Cooperation.

Evaluation approach/method:

This evaluation describes the "architecture" of Belgian aid to education, including the roles of Belgian Development Cooperation, Belgian Technical Co-Operation, NGOs, research councils, and university councils. Data was drawn from policy and background documents, interviews with Belgian direct and indirect actors, and six case studies (Benin, Burundi, DR Congo, Ecuador, Tanzania and Vietnam); also based on interviews with education officials and document analysis.

Major findings:

The evaluators note that there is little evidence of coordination among the various actors involved in Belgian aid to education, except where

the national policy framework (the document that defines educational development priorities and plans of action) is clear. This lack of coordination is evident in the disconnect between Belgium's stated educational development priorities and where the money actually goes. For example, approximately 55% of Belgian education aid goes to higher education, while the stated policy priority is basic education, which receives only 7%.

The main findings and recommendations are: (1) the need to involve national and institutional partners in project design and implementation to ensure country ownership, (2) the need to update the traditional roles of development cooperation and technical assistance in order to fit the needs of SWAps (e.g., by providing for a "more flexible and better defined role for local institutions, more clearly defined roles and responsibilities", and (3) Belgian aid delivery relies (too) heavily on Belgian/European staff working in partner countries, rather than on regional or national personnel. This evaluation takes a more critical perspective, focusing on weaknesses/challenges among Belgium aid actors, rather than focusing critiques at the "low levels of capacity" among partner governments.

Major observations:

On education in poor countries: The evaluation focuses on the policy dialogue and cooperation between different Belgium actors involved in aid to education, without evaluating specific programs/projects. One observation is that much development aid goes to activities not traditionally considered in the "aid to education" debate: university cooperation (scholarships, training programs). These partnerships cover a wide range of activities with qualitatively different intentions, the authors find.

On aid-supported education activities: The evaluation focuses on the (lack of) coordination and coherence of different actors (NGOs, universities, Belgium Development Cooperation, etc.). The evaluation recommends a more "flexible and better defined management structure of interventions," with an enhanced role for local institutions. The main barriers to this seem to be political – differing incentives, strategies, priorities, and practices between different Belgium and local actors, and a tendency to rely on Belgium (or foreign) expertise, which constrains national ownership.

On evaluating aid-supported education activities: Most evaluations note the lack of quantitative indicators, and overall tendency to focus

on inputs rather than outputs. Few discuss why this is the case (besides blaming "low levels of educational planning capacity among national ministries of education"), but this evaluation identifies the following challenges: (1) educational quality is culturally defined, (2) there is no international consensus on how to measure/define educational quality, (3) education systems are "slow" to respond to inputs, (4) educationalists are ambiguous about the use of testing, and (5) education results are politically sensitive.

What have we learned?

A surprisingly limited portion of foreign aid is directed towards basic education, much goes to university partnerships and scholarships, for example (at least in Belgium). What are the implications?

Weak donor coordination is often considered to be a primary challenge facing foreign aid to education. This evaluation demonstrates the difficulties of donor coordination – even among different actors from a single country. Each organization has its own priorities and organizational framework, and there is also a lack of transparency regarding implementation and activities. This results in intra-agency overlap (in terms of activities and countries), and means that some priority areas (e.g., basic education) and countries remain underserved.

Additionally, two claims made by the authors stand out: (1) Despite commitment to gender inclusion, there is limited evidence of concrete support for gender issues in aid-supported activities: there are no gender specific indicators, and most interventions are gender neutral, (2) Foreign aid-funded technical assistance relies too heavily on foreign personnel, which is particularly inappropriate in the context of SWAPs, where external consultants lack the the legitimacy and diplomacy necessary to play a leading role in the management of SWAPs

CfBT Education Trust (Boak, E., Ndaruhuts, S.)
The impact of sector-wide approaches: where from, where now and where to?
2011

Why selected for in-depth review:

This report analyzes the sector-wide approaches to aid to education (SWAps) in terms of (1) aid effectiveness, (2) financing, (3) education outcomes, (4) fragility, and (5) planning. SWAps are promoted on the

basis that they will improve aid effectiveness by improving coordination between donors and ensuring that the aid sector is more responsive to national policies and priorities. Since the 1990s, SWAps have been used widely since the 1990s, and remain popular, despite the fact that there is limited evidence that SWAps have indeed improved aid effectiveness.

Evaluation approach/method:

The research methodology consists of a literature review on SWAps, aid effectiveness, education planning and financing, as well as interviews with aid officials, independent consultants, and non-traditional donors.

Major findings:

The report makes the following conclusions:

- The promise of harmonization has not been fulfilled, in part due to differing levels of risk aversion among donors (donors also have different interpretations of governments' "readiness" for SWAPs). Also, some donors want more visibility than others, which can create different incentives.
- Donors need to build trust and alliances among *themselves* in addition to the government
- Non-traditional actors tend to not be involved with SWAPs, which can put a strain on governments
- Strong national leadership is required to ensure that the SWAP plan is not 'stretched' to encompass all donor projects and programmes, but rather that the SWAp plan of action dictates the type of support that is most relevant
- It's often hard to foster national ownership because "*the assumption that recipient governments behave like a strong, coordinated and unified team is unwarranted*."
- "*Broader inclusion in SWAp planning processes can bestow legitimacy on non-state actors, increasing their influence unduly*"

Major observations:

- **On education in poor countries:** This evaluation makes note of one assumption often overlooked by others – the idea that aid-recipient governments behave like strong, coordinated and unified institutions.

- **On aid-supported education activities:** The importance of relationships and diplomacy between aid agencies and local institutions cannot be overstated. This evaluation has several examples: "there is a consensus among experts that much of SWAp's effectiveness lies with the personalities of key technical staff involved" (p. 19). Is this inevitable in aid activities –in particular education? What mechanisms can be put in place to be sure that all actors—both national and foreign—have the tools necessary to develop successful aid relationships?

What have we learned?

This evaluation highlights the tension between the need for donor coordination and sustainable financing on the one hand (e.g., financing that comes from multiple sources, both national and foreign, from traditional and non-traditional donors), and the need to clearly account for the impact of aid on the other (which requires specific projects being funded by specific donors).

In the short term, the costs of developing SWAps are high (in terms of developing national consensus, balancing divergent priorities and approaches between different donor agencies, and ensuring that countries exercise real ownership over SWAp activities). It remains unclear whether or not SWAPs in the long term are more cost-effective than the alternative.

Concern WorldWide (with Irish Aid and University of Sussex)
Leach, F., Slade, E. and Dunne, M.
Promising Practice in School-Related Gender-Based Violence (SRGBV) Prevention and Response Programming Globally
2014

Why selected for in-depth review:

This is an evaluation of the "best practices" in school-related gender-based violence prevention (SRGBV), an issue of growing concern, particularly in conflict and post-conflict countries. The activities and methods used are well described, analysis is in-depth and strong with actionable results.

Evaluation approach/method:

The study included a desk review of Concern's and other agencies/organizations' policies and programming in basic education and SRGBV. From this, the authors developed a set of criteria for the selection of projects: multi-level (system-wide) approach, a gender-

based approach to combatting school-related violence, delivery within formal school system, an M&E component, and that he project took place in a low or middle income country. This resulted in the selection of three agencies – ActionAid, Plan International and USAID. These agencies projects were then evaluated, via document review, supplemented by discussions over phone and skype.

Major findings:

1. Measuring data on the impact of SRGBV is challenging, and in most cases, patchy and poorly designed/executed. This makes policy makers and donors unwilling to commit to firm action.
2. There is an over-reliance on short-training and awareness activities aimed at changing attitudes, but little evidence that this works.
3. The assumption that attitude change will lead to behavior change is not supported by evidence in the case of SRGBV.
4. The most robust evidence comes from observations and interviews indicating that strategies such as girls and boys clubs, where children can safely discuss issues and seek information and advice, as well sa develop peer-mentoring relationships, may be the most effective. Sex-seggregated toilets and clean classrooms are also promising strategies, if limited.

Major observations:

Collecting data on GBV poses unique challenges, both ethical, methodological. Because GBV related findings must be qualitatively validated, it will not suffice to rely on statistics alone (e.g., increase in the number of reported cases of violence are likely ambiguous, or do not present the "whole picture").

What have we learned?

The authors recommend that SRGBV approaches "identify and work with well-established local partner organizations," which seems particularly important in the case of SRGBV, a topic that requires contextualized approaches led by well-trusted community members. A common question also emerges between the need for targeted efforts directed exclusively at addresseing SRGBV, versus (or in addition to?) incorporating gender and gender-based violence prevention in all education-related efforts.

Katherine Conn
Columbia University 2014
Identifying Effective Education Interventions in Sub-Saharan Africa: A meta-analysis of rigorous impact evaluations
(Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences - Columbia University)

Why selected for in-depth review:

Per the author, this is the "first meta-analysis in the field of educational effectiveness conducted for Sub-Saharan Africa. The focus is on meta-analysis of *what works*, and *why*. The "why" is explained through meta-analytic techniques to evaluate the relative impact of different interventions and to explain variation both within and across interventions.

Evaluation approach/method:

Conn combines 56 articles and uses a random-effects meta-analytic technique to evaluate the impact of different interventions and explain variation in effects. Interventions are categorized according to type (content): quality of instruction, interventions aimed at reducing student/community financial limitations, school or system accountability measures, student "cognitive processing abilities" (e.g., meals, health treatments) and student or teacher motivation (incentives). Within those she further classifies, leading to 12 distinct intervention areas.

Major findings:

Comparing the relative pooled effect sizes of 12 intervention areas, Conn finds that interventions in pedagogical methods have a larger effect on achievement than all other 11 intervention types included in her analysis (average effect size of 0.30 standard deviations, which is greater than all the other areas combined. Specifically, she finds that programs employing adaptive instruction and "teacher coaching" are particularly effective. Studies that provide health treatment and school meals have on average the lowest pooled effect size, although these types of treatments do have a relatively large pooled effect size on cognitive outcomes (tests of memory and attention). She also explores where the bulk of this research comes from, both in terms of academic

discipline and geographic focus – and finds that the most research is from the field of economics (62%), followed by education (23%) and public health (15%), and from six countries: Kenya, Nigeria, South Africa, Uganda, Burkina Faso, and Madagascar.

Major observations:

- **On education in poor countries:** In fact, the only intervention type that achieved a statistically significant result is "pedagogical interventions" – but this does not necessarily tell us anything we did not already know. Programs that do not attempt to improve the quality of instruction, or to improve community/home support for education are unlikely to have a strong impact on achievement. This has been shown time and time again (eliminating financial barriers, providing incentives, improving health, providing food, can improve attendance/enrollment, but the link between quantity (participation) and quality (learning) is not necessarily straight forward.
- **On aid to education:** No focus on aid – the study categorizes intervention type according to content, but does not describe or analyze according to how or who implements these various types of interventions (e.g., whether or not the approaches are government-led or implemented by NGOs; implemented at scale or as a pilot program, etc.). This seems to be a major weakness. What matters most is the quality of implementation, not just the specific content or intervention "type."
- **On evaluating aid-supported education activities:** One interesting conclusion: "topics currently under rigorous study are not necessarily representative of the major issue facing many Sub-Saharan African school systems today," such as: multi-grade or multi-shift teaching and bilingual education.

What have we learned?

This evaluation highlights the importance of pedagogical support – above and beyond input provision, but beyond that, it does not lead to actionable conclusions or guidance.

DFID (Independent Commission for Aid Impact)
DFID's Education Programmes in Three East African Countries
May 2012

Why selected for in-depth review:

This evaluation presents a detailed description and analysis of trends and mechanisms in aid to education in three countries that receive a significant proportion of aid to education (Ethiopia, Rwanda and Tanzania), focusing on what makes these programs effective,

Evaluation approach/method:

Methods include literature review of the international evidence on "what makes education effective," revision of DFID policy documents and spending patterns, interviews with DFID central staff and government officials, DFID staff, education officials, teachers, parents, and civil society experts in-country, and announced and un-announced school visits. The authors used the EQUIP2 framework (USAID) to assess whether DFID's funding to education "systematically supports the linkages between inputs and outcomes." According to this framework, the basic "building blocks of learning" are: early grade learning, pupil and teacher attendance, pupil-teacher ratios, the availability of instructional materials and the number of hours of instruction provided to students.

Major findings:

The authors conclude that DFID has largely neglected to address the basic preconditions for learning (building blocks). This is true in these three African countries, where DFID has focused on achieving universal primary education, but this is not necessarily a universal pattern in DFID aid to education. There are some positives, though – budget support has encouraged all three countries to increase education budget from around 3% to over 5% between 2000 and 2012, and has enabled a shared platform for policy dialogue. This is linked to rapid improvements in access to primary schooling and decreases in gender gaps in education.

Observations:

- **On education in poor countries**: Authors find a consistent pattern of major funding gaps –and very little time is spent in assessing gaps betw**een planned budgets and expenditures—which results in poor performance.**

- **On aid-supported education activities**: The above is partly related to the "division of labor among donors" – DFID focuses on basic education while others focus on vocational or tertiary education, for example. This results in a "limited awareness of overall financing issues."
- **On evaluations of aid-supported education activities**: Using evidence-based (academic) structures to gauge whether or not a funding agency is focused on learning is a useful evaluation strategy, results in directly actionable findings.

What have we learned?

Rising to the challenge of improving quality, not just access, requires a "deep understanding of the processes by which this can be obtained." This includes an understanding of cost structures, context, and donor/government coordination and shared responsibility. To get us there, rather than comparing and ranking different types of interventions or different countries' education systems, approaches that assess benchmarks within countries, across regions, districts and schools, and across time, are needed.

Upper Quartile and Institute of Policy Analysis – Rwanda For DFID Evaluation of Results Based Aid in Rwandan Education – 2013 Evaluation Report - Year One March 2014

Why selected for in-depth review:

Upper Quartile and the Institute of Policy Analysis- Rwanda (IPAR) completed a comprehensive evaluation of both process (focusing on the aid recipients' response to results-based aid (RBA)) and the impact of RBA on primary school completion and the number of teachers competent in using English as the medium of instruction. This evaluation was chosen for several reasons: (1) it is one of the few evaluations to include an in depth exploration of the way Results Based Aid (RBA) affects institutional capacity, (2) it is one of few evaluations to include an impact evaluation, process evaluation, and value for money assessment, with clearly defined methodology and findings for each.

Evaluation approach/method:

The evaluation adopts what the authors refer to as a realist approach, with the goal being to "explore what works, for whom, in what

circumstances, and why." In order to do so, the authors conduct (1) an impact evaluation, using three different econometric models to explore trends in school completion (defined as sitting for school exams), and 2) a process evaluation, which focuses in particular on how the RBA approach has been perceived by the Rwandan government. Limitations and constraints, as well as research ethics, are thoroughly and explicitly explained.

Major findings:

The evaluation devotes a significant amount of time to explaining the context in depth – not an overview of broad national education trends, but a careful analysis of how the political economy of education relates to the implementation of RBA. In terms of school completion (as measured by the number of students sitting for the exam), the impact assessment estimates that the implementation of RBA has NOT significantly increased the number exam sitters. The overall time trend finds an increase in exam sitters, but this change is not attributable to RBA (coefficients on RBA in the econometric models are insignificant, and in 2012, negative (but insignificant)).

The process evaluation finds that the RBA agreement is "highly relevant" in the Rwandan context, but the focus on indicators (quantitative, with readily available data) has detracted from the government's capacity to focus on quality. Despite this recognition, the evaluation finds that the awareness of and government ownership of RBA as a funding mechanism is high.

Major observations:

On education in poor countries: A positive factor on enrollment has been the progressive introduction of free education – BUT "increases in enrollment will only have a positive impact on completion if repetition is reduced, and quality is increased. Poor teacher motivation and low proficiency in English (the medium of instruction) are integral explanatory factors of school quality (per the evaluation's econometric modeling). Teachers' gender is another factor affecting completion (at the primary level - female teachers have a greater positive effect on completion than male teachers, especially on female students, at the secondary level – male teachers have a positive effect on completion on both female and male learners).

What have we learned?

This evaluation takes seriously many aspects that (most) evaluations ignore or address only superficially: attribution, participation/ownership, process and partnership, the role of aid to education in promoting/detracting from national capacity in implementation and measurement, and multiple impacts – both direct and indirect, qualitative and quantitative, and purpose/audience of the evaluation.

DFID, with the Institute of Education at the University of London
Kingdon, G., Little, A, Aslam, M., Rawal, S., Moe, T. Patrinos, H. Beteille, T., Banaerji, R., Parton, B., and S. Sharma
A rigorous review of the political economy of education systems in developing countries
April 2014

Why selected for in-depth review:

This is not an evaluation of aid to education, but rather a multi-disciplinary literature review designed to explore decision-making processes related to education policy and implementation in low-income countries. In other words, this research should inform aid to education. The intent is to "put the theory of political economy to use in evaluating the research on education systems in developing countries" (p. 7).

Evaluation approach/method:
The review examines the "interests, incentives, strategies, contexts and exercise of power of key stakeholders in the formulation and implementation of educational decisions," focusing on decisions related to (1) schooling access and (2) improving school quality (p. 7). Methods follow a systematic review – with inclusion/exclusion criteria based on conceptual framing, openness and transparency, appropriateness and rigor, validity, reliability, and cogency (for both quantitative and qualitative studies).

Major findings:
Roles and responsibilities: teacher unions exert great influence, due to their political bargaining power, which can be good and bad for education access and quality, depending on the context. Parents, conversely, have no collective voice and therefor very limited power –

even in countries where accountability measures have sought to involve parents in education decision-making.

Rent-seeking and patronage politics: "…are rife in public education sectors in developing countries" (p. 2). The authors site quantitative studies from India, Mexico and the US showing that teacher union membership is associated with significantly reduced student achievement.

Decision-making: Research suggests that the theoretical benefits of decentralization are rarely met, especially in rural areas, where "local elites close up space for wider community representation and participation in schools." However, some institutional factors that can improve the performance of decentralization reforms are: centralized examinations, teacher autonomy over teaching methods, "scrutiny of students' achievement," and teacher incentive structures and competition from private schools" (p.2)

Implementation: Most research focuses on gaps between policy and practice, arguing that poor local capacity and corruption are the causes of poor delivery. The authors here argue that missing from these analyses is the role of "political will" – political will to implement reforms, OR political will to advocate and pass legislation related to school inputs in order to facilitate leakages.

Driving forces: Likewise, the role of political will needs to be "pitched at multiple levels" – national and local political will- which can be either mutually reinforcing or neutralizing, or even undermining. Regime type and openness also influence education spending – with democracy and openness being associated with increases in public spending on education, decreases in private education funding. However, increased spending does not necessarily lead to improved outcomes.

The authors conclude that the literature in the political economy of education is under-developed, particularly in Africa and South-east Asia, where most countries "remain virtually untouched by research on the ways in which political-economy forces affect their education-sector decisions, processes and outcomes (p. 46).

What have we learned?

How to incorporate these political economic analyses into the design, implementation and evaluation of aid-funded education projects? In particular those related to political will and driving forces of educational change: teachers unions, state and local officials, parents, all with potentially conflicting interests.

DFID, with the University of Sussex Centre for International Education

Westbrook, J., Durrani, N., Brown, R., Orr, D., Pryor, J., Boddy, J. and Salvi, F.

Pedagogy, curriculum, teaching practices and teacher education in developing countries

December 2013

Why selected for in-depth review:
While this is not an evaluation of aid-funded education activities, the research question is directly applicable to our work: "which pedagogic practices, in which contexts and under what conditions, most effectively support all students to learn at primary and secondary levels in developing countries?"

Evaluation approach/method:
The systematic review comprised two stages: a "mapping" exercise of the 489 studies that met the initial inclusion criteria, and then studies that met the relevance and methods clarity secondary inclusion criteria were reviewed in-depth (54 studies). An advisory group of education officials, teacher educators, researchers, NGOs, foundations and other development partners also provided input.

Major findings:
The primary finding is that "communicative strategies" contribute to interactive pedagogic practices, which are in turn more likely to have a positive impact on student learning outcomes. The review identifies three specific strategies that promote interactive pedagogy (1) feedback, attention and inclusion, (2) safe learning environments, and (3) pedagogy practices that draw on students' backgrounds and experiences. In turn, these strategies form the basis of six "effective teaching practices: (1) flexible use of whole-class, group and pair work, (2) frequent and relevant use of learning materials beyond the textbook, (3) open and closed [student] questioning, (4) demonstration and explanation – drawing on sound pedagogical content knowledge, (5) use of local languages and code switching, and (6) planning and varying lesson sequences" (p. 2). Effective teachers use these strategies communicatively – actively paying attention to students' learning processes and evolving/modifying classroom practices based on student learning. The review also identifies ways that teacher education can support these practices: "(1) teacher peer support, (2) alignment of teacher professional development with

teachers' needs, (3) support from head teachers, and (4) alignment of assessment with curriculum" (p. 3).

What have we learned?
From this review it can be extrapolated that aid-funded education should support education systems to develop communicative strategies – through strategies such as those outlined above. The review provides a very useful framework to do so – but of course these are not silver-bullet solutions.

DFID, with GRADE, 3ie, EPPI, IOE
Guerrero, G., Leon, J., Zapata, M., Sugimaru, C. and S. Cueto
What works to improve teacher attendance in developing countries? A systematic review
October 2012

Why selected for in-depth review:
Teacher absenteeism ranges from 3 percent to 27 percent (national average) in developing countries (per this report). This report assesses the research on the effectiveness of interventions aimed at increasing teacher attendance in developing countries.

Evaluation approach/method:
The study is a systematic review of 9 studies that meet the following inclusion criteria: (1) must assess impact of programs on teacher attendance/absenteeism, (2) study location must be a developing country, (3) must be carried out with teachers in primary or secondary schools, (4) must use a quantitative experimental or quasi-experimental design, and (5) must be published from 1990-2010, inclusive.

Major findings:
Findings suggest that programs that combine monitoring systems with rewards, or that involve the community in students' education and provides incentives for students are the most effective (on teacher attendance), but there is no evidence of an effect on student achievement.

What have we learned?

This study is useful in providing a framework through which to think about addressing one of the many challenges of education systems in low-income countries, but the finding of no impact on achievement provides further evidence of the fact that getting teachers in the classroom is but one of the important steps needed to improve learning in developing countries

DFID
Nag, S., Chiat, S., Torgerson, C. and Snowling, M. ***Literacy, Foundation Learning and Assessment in Developing Countries***
2014

Why selected for in-depth review:
This study reviews the evidence on foundational learning and literacy in order to identify key components of interventions that are appropriate to specific cultural and linguistic contexts. The author's approach and methodology are closely aligned with the realist methodology we employ in our synthesis, focusing on what works, why, and in what contexts.

Evaluation approach/method:
The review is informed by research from 1990 to January 2013, with exclusion criteria focusing on methodological quality and cultural sensitivity.

Major findings:

1. Literacy development depends on oral language skills – as does numeracy skill development. Thus, oral (spoken) language proficiency is a critical component of early learning.
2. Both child-level and school-level factors influence education attainment, but it is hard to distinguish the relative impact of the two sources.
3. Some predictors of literacy are different in different languages and writing systems– for example, phoneme recognition in Bahasa Indonesia, morphological knowledge for Turkish. This means that good quality assessments require psycholinguistic measures of the skills that are most relevant for the language of literacy – simple translations/adaptations will not suffice.

4. Measuring literacy needs to include comprehension, not just fluency.
5. Rote learning, particularly for early math teaching, is too common (this particular finding is consistent with many other evaluations and research)
6. There is "moderate evidence" of the efficacy of preschool enrichment programs on foundational learning
7. Another key message is that there are many examples of local research of practices that have been found acceptable by local communities that should be given greater weight

Major observations:
Learning and teaching in multi-lingual contexts is an issue that is not particular to developing countries, but certainly very prevalent and under-studied. This document provides a useful review of the evidence.

What have we learned?
One of the standout lessons of this paper is the need for supporting oral language skill development – where children do not have the oral language skills necessary for achieving literacy, "an intervention targeting these skills is vital."

DFID, with the University of Birmingham, IOE, and ODI
Ashley, L.D., Mclouglin, C., Aslam, M., Engel, J. Wales, J., Rawal, S., Batley, R., Kingdon, G., Nicolai, S., Rose, P.
The role and impact of private schools in developing countries
April 2014

Why selected for in-depth review:
This systematic review explores whether private schools can improve education for children in developing countries. This is a very relevant topic for anyone interested in "what works" to improve education in developing countries, given that low-fee private schools are often heralded as a promising solution to inadequate state-run education systems in developing countries. The selection and synthesis methods are transparent and rigorous.

Evaluation approach/method:
The authors selected 59 studies based on the following inclusion criteria: (1) publication date (2008 and onwards for the), (2) relevance – focus on empirical studies assessing the role and impact of private schools in low-income countries, (3) geography (DFID priority countries), (4) English (written in English), (5) high quality empirical research – qualitative and quantitative.

Major findings:
Findings are presented according to the strength of evidence. The authors find strong evidence that teaching is better in private schools, moderate evidence that private school pupils achieve better than public school pupils, moderate evidence that the cost of delivery is lower in private schools, moderate evidence that the "perceived" quality of schools is better than public schools, and moderate evidence that state intervention in private education is constrained or ineffective. The evidence is weak or inconclusive in regards to whether or not private schools are equally accessed by boys and girls, whether private schools reach the poor, whether private schools are accountable to students/families, whether private schools are sustainable.

Major observations: On education in poor countries: private schools look very different in different countries – and serve very different populations country to country. The authors are clear that private schools cannot be lumped together into a single category, but this fact inevitably leads to inconclusive results.

What have we learned?
The assumption that the poor access low-cost private schools more so than the wealthy is not substantiated by the studies reviewed in this synthesis. There is no evidence that low-cost private schools improve educational quality or equity in developing countries, and some evidence that low-cost private schools expand existing gaps between boys and girls, and between rich and poor, although in both cases the evidence is context-specific.

Upper Quartile and Institute of Policy Analysis – Rwanda
For DFID
Evaluation of Results Based Aid in Rwandan Education – Year

Two
May 2015

Why selected for in-depth review:
This evaluation is the sequel (year 2) to the previous entry. Both are included in order to explore the continuation of the process designed to be flexible and evolving, considering carefully the context and utility of the evaluation's findings. Like the year one evaluation, the year 2 evaluation takes limitations, attribution and context very seriously. To illustrate: "The programme cannot be isolated or kept constant – the evaluation approach views change as a continuous process. The evaluation must seek to understand how observed changes in completion and teachers' proficiency in English come about in a dynamic system" (p. 7). Likewise, the evaluation design is explicitly constructed in accordance with the evaluation audiences, interests, and needs in order to ensure the evaluation's utility: "In line with the realist approach, the evaluation methods are flexible and evolving to meet the needs of the study and the client group" (p. 9).

Evaluation approach/method:
The evaluation purpose is to "determine any contribution of the RBA pilot to additional learners completing key stages in primary and secondary education and additional teachers becoming competent in the use of English as the medium of instruction." The "year two" evaluation builds on the findings from the previous year ("year one") evaluation (a final evaluation – "year three" – is planned as well, and the objective is to allow for sequential implementation of qualitative and quantitative findings). The year two evaluation complements the process and impact evaluations with a more extensive Value for Money evaluation.

Major findings:
Impact findings: Completion (as measured by the number of exam sitters in 2013 compared to the number of exam sitters in 2012) rose for S3 but not for P6 and S6, where completion was actually found to decrease. However, there was no apparent effect on "GoR actions or messages," and qualitative data (from interviews) suggest that the focus on completion has negatively affected quality.
Process related findings: The evaluation finds that the RBA finance mechanism is not well known outside of the highest levels of the Government. This calls in to question the theory of change of the RBA mechanism, given that RBM is based on the assumption that

institutions will change given the right incentives (e.g., pay for performance). Moreover the English language component of the RBA modality was found to be only weakly enforced; completion was prioritized over English language proficiency. The evaluators note that the English language component was included at the behest of the GoR, and was against the initial wishes of DFID.
VfM: If the additional completion rates are attributable to RBA, then the RBA model represents very good VfM. However, although RBA has served to reinforce GoR efforts, the evidence does not suggest that completion rates would have been different in the absence of RBA. IN other words, the evidence indicates that DFID's investment in Rwandan education is cost effective, but it is not clear with RBM is more effective than alternative aid modalities.

Major observations:
On education in poor countries: "The existence of sufficient management controls and accountability mechanisms to ensure communication, compliance and action on policy priorities set by the central government will facilitate success" (p. iv)

On aid-supported education activities: The relationship between school completion and quality (learning) calls into question the use of completion as the primary outcome linked to RBA payments.

The evaluation also calls into question the use of results based management as means of establishing incentives to improve quality: "Alignment of RBA with pre-existing government priorities may remove/reduce the potential incentive for additional action to achieve results" (p. iv) – these are findings that were initially hypothesized in the year one evaluation, and substantiated in the year two evaluation.

On evaluating aid-supported education activities: This evaluation directly builds off of the previous year's evaluation (for example, from page 7: "In keeping with findings from year one of the evaluation, the research in year two highlighted the wide range of factors that seen to affect completion…").

What have we learned?
This evaluation stands out for its attention to context, history, and utility, as well as its careful analysis of the aid relationship, aid impact, and process.
The authors endeavored to explain the evaluation's findings in light of context and history: "Through this process the evaluation found that Rwanda's imihigo system has been used by GOR to mainstream

messages and incentivize action to promote completion" (p. 47). "…Rwanda was a results-oriented country prior to RBA, and there is no evidence that GOR's approach has been altered…It is unclear, from the evidence available, how RBA may or may not function in a country that was less results oriented to start with" (p.50).
This evaluation seems much more useful than the majority of the evaluations we have reviewed. Is this the case? How has this evaluation been utilized by its intended audience—the Rwandan government and DFID?

European Commission (EC)
Thematic global evaluation of European Commission support to the education sector in partnering countries (including basic and secondary education)
2010

Why selected for in-depth review:
The evaluation itself is comprehensive in terms of scope and description of methods and activities. Also, the EC approach to educational development is unique – including a strong focus on country ownership and the provision of general budget support to aid-recipient countries (e.g., funds that are un-earmarked, which countries use to cover overall public sector financial needs, including teachers' salaries).

Evaluation approach/method:
The evaluation consists of four main components: inception phase (to establish an inventory of EC support to education and define the scope of the evaluation, desk phase (survey of EU delegates, interviews with EU officials at head quarters, and document analysis, field phase ("information gaps were filled and hypotheses were tested") and a synthesis phase. Overall, more than 6,000 documents were reviewed, 200 interviews, 3 video focus groups, and 6 field-based case studies were conducted. The main limitations are lack of attribution (causal impact), access to and availability /quality of qualitative and quantitative primary data.

Major findings:
The evaluators find that EC support is highly relevant for aid-recipient countries' national priorities and policies, in part because of the general budget support (unearmarked funds), which allows them to

meet budget shortfalls and increase basic education access. This is especially the case in conflict/emergency or post-conflict/emergency countries, where EU budget support has helped countries meet the shortfall in school provisions in certain regions. One of the EC's main contributions are projects designed to improve girls' enrollment through a variety of measures, including training female teachers, and focusing on vulnerable populations (e.g., students with disabilities, and pro-poor school investments). However, EC support to improving quality "needs further focus" – EC support "has not so far enhanced basic literacy and numeracy skills." The evaluators refer to a "quality crisis" – confirmed by data on learning achievements and school-leaving exam results in aid-recipient countries.

Major observations:
The evaluation describes the "delays in aid-disbursement" having to do with non-compliance of indicators and "weak capacity of national staff." This again demonstrates the problems encountered when indicators and monitoring strategies, accountability mechanisms are based on aid agencies' experience and capacity, rather than national capacity/interest. Efforts to improve local monitoring/evaluation/planning capacity are called for, but not evaluated. Neither are efforts to foster country ownership of these systems at all stages.

What have we learned?
This evaluation largely confirms what many other evaluations have found – that aid funding can help aid recipient countries meet immediate budget needs and increase accesss, and can at times support the development of gender-sensitive and pro-poor education policy frameworks. Results on decentralization and school-based management are mixed, and in many cases, only a very limited portion of the education budget is fully managed at the decentralized level.

GIZ
Évaluation ex-post 2012 – Rapport de synthèse: Promotion de l'éducation de base, Tchad
July 2013

Why selected for in-depth review:
This GIZ evaluation was very well-done and highly participatory, and addressed the role of parent associations and potential networks. The

evaluation also described educational interventions at great depth, unlike many other evaluations that we have seen that focus solely on strategy. The project operated autonomously outside of the Chadian Education Administration. Executed by GIZ, other organizations were implicated, including KfW and the World Bank, within the framework of the national education program.

The stated goal was to evaluate innovative approaches ameliorating basic education access and quality, in particular for girls in three regions within Chad, integrated within the national sector policy.

Evaluation approach/method:
Concerning reinforcing capacities, the project took into account three levels of intervention: (1) development of human resources, (2) organizational development, and (3) society. The evaluation utilized qualitative methods, which included 36 individual interviews and 19 group interviews in the capital, N'Djamena, and in the region of Mayo-Kebbi. On top of this, 220 students and 33 teachers were also surveyed through a standardized questionnaire in one region. The evaluation team consisted of an international expert alongside a national expert. Yet, in terms of results dissemination the target audience was the funding agencies and the Ministry of Education.

Major findings:
The evaluation indicates that the project was primarily active in the areas of support to community schools. The evaluators note that in this context, innovative approaches were developed and implemented, different textbooks were developed and numerous trainings for parents' associations of members of students and teachers were conducted. Social interventions were successfully implemented through the formation of networks between parent associations. New pedagogical approaches were put into place with success, yet at the end of the project they were not continued as had been planned. The findings indicate that parent associations were eventually included in the national sectoral policy as a result of the project, and therefore strengthened at the institutional level.

Major observations:
On education in poor countries: The evaluators observe that primary schools receive little funding relative to the overall state budget dedicated to education. As a potential solution, the evaluators endorse the project that was evaluated, and advocate for more reinforcement

of parent associations. The various interventions of this project such as "promotion of parent associations", "teaching in maternal languages", "promotion of girls education" and "HIV/AIDS education in schools" were successfully completed, according to the evaluation, which also notes a need for follow-up on the long-term sustainability of parent associations.

The evaluation indicates that girls especially benefited from the project. Interviewees viewed the GTZ project positively, indicating it led to better access to education as well as improved academic results for girls. The evaluation indicates that the goal is to scale the project to the national policy level. Though girls benefited most from increased access to basic education, further work is needed. Finally, the evaluation indicates that the school dropout rate and educational quality need to be improved.

On aid-supported education activities: The evaluation recommends that projects aiming for scalable impact ensure that measures of capacity building are already taken into consideration before the project begins.

On evaluating aid-supported education activities: The evaluators suggest that evaluations should be planned and budgeted for from the very beginning of a project, in order to properly address the various approaches introduced during project implementation. The authors conclude that "proven success" is the only way to achieve buy-in from other actors of concepts generated by a project.

What have we learned? In terms of instruction in maternal languages, the evaluation suggests that teachers and teacher trainers must be on board with new approaches before they are implemented.

The evaluation provides evidence in favor of discussion events and the widespread diffusion of information (with the assistance of public media), and indicates that they are necessary as a first step towards effective implementation of new approaches. Additionally, the evaluators indicate "a local approach and coordination with targeted groups and intermediaries are preconditions for successful interventions during project execution," (p. 8).

This evaluation is unique because of its wide inclusion of local stakeholders, which achieved positive results. KfW's promotion of parent associations was very successful. While the goal of promoting a diversity of pedagogical approaches was successful, evaluators noted that the degree of institutional development and cooperation at the national level were insufficient. The evaluation indicates several recommendations for follow-up...it would be curious to see how often such recommendations are acted upon.

Inter-American Development Bank - Office of Evaluation and Oversight
Thematic Evaluation: Review of IDB Support to Secondary Education: Improving Access, Quality and Institutions, 1995-2012
October 2013

Why selected for in-depth review: The evaluation is well organized and well written, includes a well-defined methodology and limitations, and also enhances the diversity of our sample, in terms of both the funding agency and the aid recipient country.

Evaluation approach/method:

This evaluation aims to determine the "extent to which the Bank supported equitable access to secondary education, improvements in secondary education quality, and reforms in institutions to improve management capacity." The data are mostly qualitative, drawn from document analysis and case studies in Argentina, Paraná Brazil, Dominican Republic, Paraguay, Peru, Trinidad and Tobago, and Uruguay, but quantitative analyses of program costs, outputs, outcomes and test results are also included.

Major findings: The evaluators note that there is limited evidence that any IDB supported programs have achieved the desired impact in terms of access, quality, efficiency, and institutional strengthening, largely because individual program evaluations focus on inputs and outputs, rather than outcomes/impact. Thus, one of the evaluators' recommendations is to improve M&E capabilities, which, it is argued, requires strong institutions. Along those lines, the evaluators also recommend emphasizing "innovation and knowledge development" to strengthen the "repository of evidence-based interventions." This could include vocational education and cost-effective uses of technology (despite the fact that the evaluation notes that there is limited evidence supporting the role of ICT in improving educational outcomes).

Major observations: On education in poor countries: Changing curricula to reflect global and local societal changes: IDB evaluation notes that "keeping up with the pace of change in society is difficult for Ministries of Education." This evaluation (like many others) laments the persistent use of rote teaching methodologies.

On evaluating aid-supported education activities: Indicators used in most evaluations are tied to specific activity inputs/outputs, which limits the sustainability or replication of M&E practices developed under aid-financed projects. This is one of the few evaluations that encourages participation in regional and international assessments as a primary way to improve national institutional capacity

What have we learned?
This evaluation, along with others reviewed, supports the idea that the most vulnerable populations remain un-reached by most aid-funded education projects. This evaluation exhibits the difficulty of evaluating programs designed to strengthen institutions according to quantifiable output/outcome measure. The authors recommend "paying more attention at the design phase to the statement of outcome indicators, and to the type and quality of data to be collected." These reforms in particular need to be flexible and well aligned with local political and economic contexts, meaning that the typical input-output framework for project monitoring may not work.
There is a tension between the need for sustainable educational strategies, based on national ownership and priorities, and multi-lateral banks' preference for policy-based and performance-based loans, which put pressure on national governments to implement bank-promoted reforms.

Irish Aid
Zambia Country Strategy Paper: Evaluation 2007-2010
2012

Why selected for in-depth review:

This evaluation was in part selected to provide sample diversity, as Irish Aid is, to a certain extent, a non-traditional donor (or at least one of the lesser-studied donors). The evaluation also takes a somewhat unique approach, focusing on development strategy, development processes, development results and development management. The methods are transparent and well described.

Evaluation approach/methods:
This evaluation employs contribution analysis (way of assessing the extent to which it is plausible that observed changes can be attributed to Irish Aid programming) as its primary methodology, with data

collected from desk research, telephone interviews and fieldwork in Zambia.

Major findings:
Overall, Irish Aid's support to education was "relevant to country priorities," and "built on areas of added value that Irish Aid identified" – especially in terms of gender inclusion and support to civil society education. The evaluation authors still claim, however, that despite the effort to improve aid effectiveness in the education sector, progress has been disappointing. With regards to development management, the evaluation team finds that program efficiency and effectiveness were limited due to challenges in logistical and managerial structures between staff and senior management.

Major observations:
On aid-supported education activities: The authors note the importance of the "personal approach" and flexibility employed by Irish Aid officials – informants in Zambia mentioned that Irish Aid was "supportive and approachable" – and "understanding of the problems that they faced." For this reason, Irish Aid is perceived to be one of the primary development partners committed to aid effectiveness, and allows Irish Aid to provide support "in a way that other donors are often unable to do." Similarly, the authors note the high degree of "long-term institutional memory" – because staff stay on for long(er) periods of time and are well known for their professionalism and technical competence.

What have we learned?
This evaluation focuses on the role of development agencies in influencing national priorities – in this case, gender equality in education –more so than influencing outcomes, *per se*. That this is a positive result of aid-support to education is unquestioned by the evaluators. Is it, though?
This evaluation also provides a strong description of the challenges of donor coordination and the challenges of developing and maintaining a country strategy in the context of multiple donors and competing interests for distinct types of funding modalities and accountability systems.

Mathematica Policy Research, Inc.
Levy, D., Sloan, M., Linden, L., and Kazianga, H.
Impact Evaluation of Burkina Faso's BRIGHT Program: Final Report
June 2009

Why selected for in-depth review?
This evaluation follows rigorous quantitative methodology – impact assessment, with detailed description of design, sample selection, causal analysis, limitations, etc. using a quasi-experimental: regression discontinuity, a methodology considered to be "as good as an RCT" if the assumptions are met. The evaluation assesses the impact of a comprehensive program (including school meals and take-home rations, provision of school kits and textbooks, community mobilization campaign, literacy programming (adult literacy and training for girls), and local partner capacity building) on enrollment and test scores. Given the program's focus on gender (girls' education), the evaluation also looks at how these effects differed for boys and girls.

Methods/approach:
The authors use regression continuity design to compare communities that participated in BRIGHT to a similar group of comparison group. Assignment is based on an eligibility score – scores above participated, below did not. Data was collected at baseline and end line two years after implementation – including a household survey, school-based survey and school administrative data.

Major findings:
The BRIGHT program increased enrollment and test scores, and the impact is larger than those observed in most evaluations – the authors argue that this mostly is a result of constructing schools in areas that had no school before the program was implemented.
Major observations:
On education in poor countries: Information from parents (household survey) suggests that school construction was a crucial feature of the program – enabling children to travel shorter distances to schools. The program was implemented in communities that did not have a school before the program was implemented, however, which is important to note in considering possible policy lessons for other contexts.

What have we learned?
Two big questions remain unanswered: whether the effects will be sustainable, and whether the approach is cost-effective and or "scalable" – given that it was implemented in a relatively small scale pilot program led by NGOs. The evaluation is very useful in demonstrating a quantifiable impact assessment, which is particularly valuable for accountability purposes, but less so for knowing any more about "what works" for education, other than the fact that if you construct schools in places with no schools, attendance will increase (although, test scores did too in this case).

Netherlands Ministry of Foreign Affairs
The two-pronged approach: Evaluation of Netherlands support to primary education in Bangladesh
August 2011

Why selected for in-depth review:
Interesting conceptual approach: two-pronged approach. Focusing initially on BRAC, Netherlands education sector support diversified towards a 'two pronged' approach. The approach of the study is unique in that it is a case study evaluating support to formal AND non-formal basic education. Additionally, Bangladesh is one of the largest recipients of aid for education, and Dutch support was provided through two distinct channels – for non-formal primary education through BRAC (a major NGO in Bangladesh), and for formal primary education, through the national government. The research methodology is quite solid, as the evaluation used a mixed-method approach, including both qualitative (interviews, focus group discussions, school visits and classroom observations, document analysis, etc.) and quantitative research methods.

Evaluation approach/method:The country evaluation permits comparison of the two unique channels of Dutch aid in reaching the MDGs and EFA objectives. The evaluation used mixed-methods, and included a literature review, quantitative modeling, interviews with key players in the education sector in Dhaka (including a working group) and a qualitative field study in two districts among local education officials, different types of primary schools, and teacher training institutes. The evaluators note that "no primary quantitative data collection was conducted for the impact evaluation," (p. 31). Qualitative research was comprised by semi-structured interviews with various education stakeholders (including government officials

involved in education, the Dutch Embassy, and the donor community in Dhaka, which all participated in an education working group within a local reference group; semi-structured interviews were also conducted with staff members of various NGOs, the ILO, and research institutes, (see p. 31). The evaluators triangulated interview data with findings acquired from other evaluation tools. Secondly, a comprehensive review was undertaken of the literature on education development in Bangladesh. Thirdly, qualitative research was undertaken at the school level.

Major findings:
The evaluation finds that both Government and NGOs such as BRAC have made targeted efforts to increase girls' enrolment in school (the groad spectrum of efforts include awareness raising campaigns to (secondary) school stipends). The evaluators indicate that combined efforts resulted in virtual gender parity in primary education in Bangladesh. The study also indicates that attendance of poor boys is increasingly surfacing as an issue that merits more attention.
The evaluators indicate that delayed enrolment leads to increased opportunity costs and dropout rates. Additionally, the report points out that substantial age differences in the classroom affect teaching and learning. Against the background of a standard national curriculum and standard textbooks, the evaluators suggest that these differences require close attention. School attendance is below the national average at madrash schools, according to the data, and the evaluators warrant that this merits more study given increased enrolment of students at madrash schools.

Major observations:
The evaluation points to various reasons for the non-enrolment and poor attendance of boys from poor families, ranging from higher prevalence of male child labour to a lack of interest in education among boys. While beyond the scope of this evaluation, this topic needs further analysis as this phenomenon occurs not only in Bangladesh but also in other countries (e.g. in Pakistan).
Equity of access was also addressed through Netherlands support for 'hard to reach' children in very remote areas and in slums in Dhaka.
The evaluation was hampered by a lack of consistent, comprehensive and up-to-date data on various key indicators such as dropout and school completion rates. Moreover, the history of student assessment is limited in Bangladesh.

What have we learned?
In particular: The BRAC experience shows that, at a time when the Bengali government is shaping its own education system, it is possible to provide non-formal education through NGOs that is less costly, takes less time than formal education, and yields good results in terms of learning outcomes. Sustainability of this external support for non-formal education initiatives remains, nevertheless, a key concern.
In terms of the utility and limits of methods, the evaluators indicate that the regression analysis confirms earlier findings as regard improvements in learning but remains somewhat inconclusive with regard to the determinants of these improvements.

RTI
Heyward, M., Cannon, R., and Sarjono Implementing School-Based Management in Indonesia September 2011

Why selected for in-depth review? The study provides a comprehensive assessment of the project context relevance, take-up, and implementation, and explore outcomes on school management and governance as well as the extent to which the content of the project was taken up and replicated beyond the scope of the project, an element that is ignored in most evaluations but no doubt essential to long(er) term sustainability. Methods – data and limitations – are thoroughly described and justified. Additionally, school-based management is a popular strategy among aid agencies and donors.

Approach/methods
To impact evaluation employs a non-experimental quantitative design, comparing baseline and end line data from target schools. To account for the lack of counterfactual, the authors use a mixed-method approach to track changes over time and triangulate findings from routine project monitoring and performance indicators, two qualitative field studies carried out in 98 schools, three studies of school funding, and participant observation in eight school clusters in two provinces.

Major findings:
Findings suggest a positive impact of the school-based management program on planning, community participation, and transparency, and the intervention was adopted by other (non-target) schools, thereby expanding the impact.

Major observations
On aid-supported education countries: The authors attribute the program's success in large part to the fact that it was "firmly and explicitly based on government policy" (p. 10). The implementation design and implementation was flexible – and was seen as a partnership with local governments and project implementation team – with shared responsibility for achieving objectives (this evidence is from interviews and observations).
On evaluations of aid-supported education countries: The mix of qualitative and quantitative methods is effective in this case – qualitative methods are used to corroborate quantitative findings and explore how communities have participated in and taken-up the various program components, and to explore factors that likely contributed, or explained, the program's success:
What have we learned?This evaluation provides additional support and specific examples for the importance of country (government) ownership at all stages of program design, implementation and evaluation. Despite the overall success of the school-based management program, the impact was quite heterogeneous between regions, however – and the most significant element in the project's success, per the authors' interviews and observations, "was the level of commitment of the district or province and the capacity of the implementation team to leverage and build that commitment" (p. 11).

Sida
Swedish support to the education sector in Mozambique 2004

Why selected for in-depth review:
The evaluation examines Swedish aid to the education sector in Mozambique since the country's independence in 1975 and through a turbulent and fragile period, providing possible lessons learnt in terms of aid to education and evaluation in contexts of fragility. Additionally, this country study adds to the diversity of our compliation of evaluations by including a Lusophone countru.

Evaluation approach/method:
Documentary anlaysis included "country analyses, country strategies and country plans for development cooperation, project- and programme documents, project reviews, evaluations and auditors reports, research reports and other relevant studies from Sida, Sweden, Mozambique and international organisations,' (p. 6).

Major findings:
This document highlights the changes in Sida's support to education in Mozambique. These changes in strategy were due to various events (providing a historical backdrop), including civil war and destabilization in Mozambique, international politics during the Cold War, neoliberalism and structural adjustment, shifts in international development approaches, new emphases on environmental challenges, alongside HIV/AIDs and refugee flows, and then the MDGs.

What have we learned?
This study highlights some important consequences of the transition to sector-wide support for education in Mozambique (donor coordination), namely:
The evaluators indicate that Sida withdrew too rapidly close contacts with government officials, in order to concentrate support and give greater ownership to Mozambique "*To transfer responsibility and not wish to follow the process closely, entails that one neither ought, nor can, have complete control over all events, nor over the mistakes or errors that might arise. This relationship must in some way be part of the agreement itself. At the same time, all the parties involved must share knowledge of suitable methods of work and the demands that are made, so that all the parties will be able to respond to what is expected of the final results. There must also be regulations prepared that come into force to correct eventual mistakes and errors. Not all of this capacity was in place when the changed methodology of collaboration was introduced" (p. 39).* The authors also focus on the "development of competence and greater ownership among recipient governments" (p. 61) – which is deemed to be essential, but only possible with "decentralization of authority, good communication and greater 'freedom of movement' wihin the development cooperation."

Sida
Evaluation and Monitoring of Poverty Reduction Strategies – 2005- Budgeting for
Education: Bolivia, Honduras and Nicaragua
2005

Why selected for in-depth review:
This evaluation is relevant because it was commissioned by Sida, and is a cross-country comparative cost-effectiveness analysis that includes discussion of monitoring and evaluation. The goal of the program was to align the poverty reduction strategy to achieve education-related MDGs through the mechanism of output-oriented budgets.

Evaluation approach/method:
The evaluation aims to measure successes within the education sector in Bolivia, Honduras and Nicaragua, as well as to present the results of a cost-effectiveness analysis conducted in each country. The evaluation approach is essentially a needs assessment ("human, physical, and financial resources") to estimate "cost of achieving MDGs", (p. 5); measures current education sector achievements and conducts a cost-effectiveness analysis. Using a simulation model as well as case studies, the evaluation includes a stock-taking of local actors through field visits. "The case studies on cost-effectiveness analysis and result-oriented budgeting presented in this report build on the methods and framework developed by Gertler and Van Der Gaag (1988), Gertler and Glewwe (1990) and applied, among others by Bedi and Marshall (1999), Bedi et al. (2004) and Vos and Ponce (2004)…household survey data and appropriate econometric methods were used to estimate the empirical model and to identify the effect of school costs and of schooling inputson enrolment," (p.11). Schooling inputs included availability of books, qualification of teachers, and school infrastructure.
The evaluation asserts that it presents a unique theoretical model, as reaching the MDGs requires policy that takes into account "human, physical, and financial resources…in its design and implentation" (p. 5). Yet, the evaluation points out many data limitations and weaknesses of the regression models, in particular, indicates that the limitations of the simulation confirm the need to look at demand-side variables. Other challenges and limitations presented by the regression models include (1) the absence of a reliable database makes budget simulations virtually impossible, (2) many variables are municipal level and not at the school levels, affecting estimations and reducing sample

variation, (3) "high degree of aggregation in the simulation model," masking rural/urban, geographic, and gender differences, etc, and (4) several other limitations.

Major findings

On education in poor countries: While net primary enrolment has significantly increased, challenges remain in terms of educational quality.

On aid to education: Evaluators and policymakers need to also "look at demand-side variables - in particular the reduction of poverty - to reach the goal of universal primary education," (p. 5-6).

On evaluating aid-supported educational activities: Cost-effectiveness analysis illustrates that reaching the MDG of 100% net primary enrolment in Bolivia, Honduras and Nicaragua, is impossible "using only one or more of the education policy instruments considered in the enrolment models estimated for these countries," (p. 5-6).

Major observations:

On evaluating aid-supported educational activities: "The determinants of access to schooling are context-specific as shown by the three cases; hence, these exercises have to be conducted in-depth for each country and cannot be generalized across countries, as for instance the MDGs have done to some extent. (p. 29).

While "costing exercises are usually based on one type of methodology," evaluators note that there are tradeoffs and choices between different approaches to achieving the MDGs and " a lot of qualitative judgment is involved in determining what a 'good' policy might be," especially when considering various alternatives, (p. 8).

What have we learned?

"It is important to avoid a technocratic approach to result-oriented budgeting, as budgeting should be the outcome of a negotiation process which not only considers the (expected) impact of policies and budgetary implications, but also takes due account of political economy and institutional factors. Institutional weaknesses, lack of coordination between institutions (within the central government and between central and local governments) and political pressures to alter agreed budgets are likely to hamper a move towards ROB, so ways should be found to strengthen institutions and improve coordination between them, as well as to reduce political pressures," (p. 30).

Sida
Sida's contribution 2006: Progress in educational development
2007
Why selected for in-depth review:
This report is an "analytical results oriented" review of Sida's support to education globally, in the context of the agency's intention to increase its efforts to improve learning for "poor boys, girls, women and men by providing equitable access and better quality education." This is one evaluation conducted in-house, prior to Sida's outsourcing of evaluations to InDevelop.

Evaluation approach/method:
Major findings:
The following patterns regarding education outcomes in the countries receiving Sida support to education stand out:
Early childhood education is seldom prioritized, there is very low coverage
Access to primary education has increased, but marginal groups remain excluded
Only a small percentage of youth participate in secondary education
Adult education is limited – literacy rates are improving, but slowly
The low quality of teaching and learning is a persistent problem
The gender gap in primary education is decreasing, but education systems still have a long way to go to achieve gender parity is far from reality

Major observations:
"The shift from project support to sector and budget support puts technical issues regarding aid modalities at the forefront" (p. 5). This is because donor coordination and harmonization are costly (particularly in terms of time), but necessary. The authors argue that Sida will need to maintain its focus on educational issues – focusing on conditions that support effective teaching and learning – through increased/improved monitoring and assessment.

What have we learned? The country-level analyses within focus on the overall growth and challenges in the education sector of each countries, rather than the role of Sida specifically.

Sida - Wort, M., Sumra, S., Schaik, P. Mbasha, E.
Swedish Support in the Education Sector in Zanzibar, 2002 – 2007
2007

Why selected for in-depth review:
This is one of the few (English language) evaluations of Swedish aid to education produced by Sida. The evaluation is also well written, well organized, with the explicit objective of guiding used to future Swedish aid policy.

Evaluation approach/method:
The methods are similar to many of the other evaluations reviewed – a combination of document analysis, interviews with policy officials at the Ministry of Education and Sida, and group interviews with educators, community members and education officials around the country.

Major findings:
The authors find that overall Sida support has been successful, in that the "outputs are evident;" that is, schools have been constructed and classrooms have been refurbished. However, in terms of lessons learned, the authors argue that going forward, education aid should be based on a "well thought out framework and methodologies." The lack of a Sida document outlining objectives resulted in a tendency to focus on inputs rather than outcomes. In particular, the authors highlight the need for a more coordinated "school mapping system" in order to reach the most marginalized communities. Another key finding is the overall lack of national ownership, or even understanding, of the education sector SWAp. Efforts to support government capacity development have mostly consisted of Sida-led consultancy groups (report writing, diagnostics), and several education officials have been sent to trainings and PhD programs abroad. At the community level, the authors find that parents and community leaders are actively involved in school construction processes, but once the schools are completed, the sense is that "the community responsibility is no longer, and is passed on to the Ministry of Education." The authors make the case that communities should be more involved in school management and decision-making processes.
This evaluation stands out for its focus on "progress as processes," which the authors note leads to a focus on systematic issues set in the context of constraints at the national, district and school levels.

Major observations:
On education in poor countries: Community ownership of school construction is easier to achieve than community ownership of school management/decision-making processes. Also – on the limits of community ownership: the evaluation describes a program in which communities are responsible for providing toilets at schools, but the evaluators find that there is a lack of suitable and separate toilets for girls. Local institutions/communities don't always know best.
On aid-supported education activities: On the limits of EMIS: EMIS relies heavily on technical skills, this is challenging in situations w/low levels of programming skills/math skills. An Integrated Information Management System - linked to EMIS and to the school mapping system - might be better. This would be a decentralized system that would be the domain of "all Directorates" - not just the Planning Directorate. South Africa and Namibia are in the early stages of developing an IIMS
On evaluating aid-supported education activities: Findings often are not explained, not well interpreted. For example: "the roles of school committees will need to be broadened and capacity strengthened to ensure their participation in managing schools is done in a more meaningful way."

What have we learned?
Some evaluations call for increased flexibility (among aid agencies), while others (such as this one) bemoan the lack of clearly defined desired outcomes. The authors attribute the focus on inputs rather than what should be achieved (outcomes/impact) in part to an absence of overall country-wide "Sida-backed objectives." What is the correct balance between aid agency flexibility and clearly established priorities? *Is* there a correct balance?
This evaluation, together with others (IDB 2013, World Bank 2006), demonstrate the limits of aid-funded efforts to improve national educational governance capacity by sending education officials abroad for trainings/scholarships. This is one of the primary strategies utilized by aid agencies, but there is limited evidence that these trainings lead to lasting improvements in education planning, policy-making, or finance.

Sida
Policy Guidance and Results-Based Management of Sida's

Educational Support
2008

Why selected for in-depth review:
As Sida has commissioned our study, it is key to understand Sida's previous evaluation of educational support. The justification for the evaluation as well as the conclusions drawn are useful for our synthesis, and particularly the discussion of results-based management of educational support. The evaluation also considers organisational conditions (such as the division of labor, absorbative capacity of different units, management) influencing the extent to which evaluation findings are accessible, and the extent to which findings are used to inform policy, for instance, by looking at information flows and relevance for users. Yet, though the lessons learned are relevant, the very weak response rate for the questionnaire involved in this study indicates a strong limitation within this study.

Evaluation approach/method:
The evaluation examines strengths and weaknesses of the entire management process in education, in order to respond more closely to the commitments of the Paris Declaration (which became a strong benchmark, internally and externally, of Sweden's development cooperation). The evaluation uses documentary analysis, surveys, and interviews to assess steering instruments, as well as results information acquired from monitoring and evaluation and evaluation instruments. The methodology and approach included an attempted survey sending out questionnaire by e-mail. In the event, the response to questionnaires (only 7 returns – 20% response) was limited and the questionnaire findings were restricted to a collation and analysis of informed comments from the respondents…extensive consultations were undertaken in Stockholm and in selected case countries," (p. 19).

Major findings:
The evaluation found continued challenges to educational quality, despite success in expansion of access, insufficient measures of quality. There has been some successful support in capacity development of monitoring and evaluation, yet the evaluators note that there are limited links between information on results and the change in the design and implementation of programs. For instance, evaluators observe a limited use of results from pilot studies, and advocate that Sida take a more holistic approach, overall, to the education sector. Overall, the evaluators note that "country results information flows

are uneven, insufficiently strategic," (p. 11) yet there is a reduction in information needs when there is good harmonization and alignment... "participation in country sector donor working groups is becoming a critical source of results information," (p. 11) reducing the individual burden on information gathering for funding agencies. At the same time, the evaluators note that results information flows need continued improvement.

Major observations:
On education in poor countries: more country-specific education guidance is needed, according to the evaluators.
On aid-supported education activities: With increased harmonization of aid (via Paris Declaration), the evaluators not a need for a more sector-wide and outcome-oriented focus on guidance instruments. There is also a lack in guidance in how to link education with broader development goals, such as human rights, poverty reduction, etc, a lack in guidance on implementation of education innovations (oftentimes small-scale approaches through CSOs), and on how to conduct policy dialogue and on SWAps in general, as well as a need for guidance on transitioning from emergency situations to development contexts, and on transitioning from individual projects to more harmonized projects with other partners or projects.
On evaluating aid-supported education activities: "Basic preconditions for results-based management are lacking in the educational sector. An overall conclusion is that management in the education sector is based on blueprint formats rather than a systematic use of policy instruments or information on results," Stefan Molund, Acting Director, Sida Dept for Evaluation, p. iii.
The evaluators note the importance of understanding all "sub-sectors of education and their inter-relationship," (p. 11), yet that this broader understanding of education cooperation may be difficult to realize due to organizational structures with different mandates within the funding agency (p. 11).

What have we learned?
The evaluators note a need for more guidance on how to ensure education projects are tailored to specific country contexts, and this is perhaps due to uneven results information flows regarding various countries, and limitations in the internal reflection process within funding agencies to actually use evaluation results. Additionally, they indicate that additional guidance gaps include how to incorporate a more sector-wide approach, particularly for secondary and higher

education, while also considering the role of civil society and private sector.

Sida (conducted by InDevelop, for Sida)
Swedish Development Cooperation in Transition? Lessons and Reflections from 71 Sida Decentralised Evaluations (April 2011-April 2013)
2013
Christoplos, I., Liljelund Hedqvist, A., Rotham, J.

Why selected for in-depth review:
While not an evaluation of aid-funded education activities, this study is of relevance given the focus on Sida and its overall objective to "identify lessons of relevance to strengthen management for results" and contribute to "evidence-based policy," thus closely aligned with our synthesis review.
Evaluation approach/methods:
The synthesis followed a format designed to extract the main findings from each of the 71 evaluation reports, and then develop findings and recommendations related to aid management. Evaluations were selected to be representative of different sectors and different portfolios, but the sample is not necessarily representative of Sida's aid portfolio as a whole.

Major findings:
The review finds four main "success factors" of effective programs: (1) committed and engaged organizations and individuals, (2) capacity within the partner organization, (3) developing programming based on a thorough political and economic assessment prior to implementation, and (4) ownership and political will. In terms of management, the review finds that most interventions assess activities and outputs, rather than monitoring, and that there is an overall weak culture of monitoring and evaluation.

Major observations:
- **On aid-supported education activities**: The reviewers conclude that most efforts to support capacity development among partner organizations (aid recipient governments) fail to demonstrate results.
- **On evaluations of aid activities**: In addition to a focus on activities and outputs, rather than outcomes and impact, low budget for evaluations is a major limitation, also the fact that "evaluations are not

intended to generate general lessons" – and very limited time frame allowed for country, regional and thematic reviews of evaluated programs.

What have we learned?
In efforts to improve aid accountability and transparency, there is a risk that partners (e.g., aid recipients) come to see their responsibility as being limited to producing outputs. This can in turn limit the expectations about what a program should achieve.
Also interesting is the observation that "evaluations almost invariably include recommendations relating to the need to improve results-based management systems among Sida and partner organizations," and little evidence of institutional learning or effective "evaluative relationships." The exception to this rule is where there is a "constructive dialogue" between evaluators, Sida, and program partners" (p. 26).

Sida (conducted by InDevelop, for Sida)
Lessons and Reflections from 84 Sida Decentralised Evaluations 2013 – a Synthesis Review
2014

Why selected for in-depth review:
This review builds on the previous entry (*Swedish Development Cooperation in Transition? Lessons and reflections from 71 Sida Decentralized Evaluations)* by providing deeper analyses into several areas of focus: the use of theories of change, the focus on poverty, and the efficiency of Sida supported projects and programs.

Evaluation approach/method:
The synthesis draws from a purposive sample of 84 evaluation reports from 2013, covering all of Sida's country categories and thematic sectors (though the sample is not necessarily representative of Sida's overall portfolio). The synthesis followed a 28-item tool designed to record and collate qualitative and quantitative data from the reports.

Major findings:
The primary success factors for development results are: (1) a coherent and unified Swedish approach to development, (2) selection of strategic partners with the "right" approach and capacity, and (3) strong and committed leadership. Overall, the report finds that "Sida has yet to overcome institutional hurdles and develop sufficient

mechanisms to learn from experience in general and evaluations in particular" (p. 9).

Major observations:
- On capacity development: How to measure and assess capacity development? It is often listed as one of the main outcomes of a project, yet it is simultaneously stressed that low capacity is a major limitation in achieving desirable long(er)-term outcomes/impact.
- On theories of change: Many programs fail to critically assess the assumptions guiding the theory of change and to identify direct and indirect beneficiaries. Sustainability is also missing in most design strategy and evaluations.
What have we learned?
This review provides a useful analysis of the limited utility of evaluations as commonly conceived – measuring inputs/outputs, as well as more examples of the importance of ownership and participation (of aid recipients, agencies), and the need to ground program design, implementation and evaluation in strong assessments of the relevant political economy.

UNICEF (American Institutes for Research, conducted for UNICEF)
Child Friendly Schools Programming, Global Evaluation Report

2009

Why selected for in-depth review:
This study explores the extent to which the child friendly school model has been taken up and operationalized among participating countries, an important step along the theory of change connecting the intervention to student outcomes (learning) that many evaluations ignore. The methods are comprehensive and thoroughly described. The study also enhances the geographic diversity of our synthesis.

Evaluation approach/method:
Methodology consists of a desk review of child-friendly school (CFS) documents from all regions, and primary data collection in Guyana, Nicaragua, Nigeria, Philippines, South Africa and Thailand (interviews with teachers, school leaders, parents, and students). In country, schools, students, teachers and families were selected randomly, and qualitative data was combined with quantitative data (hierarchical linear modeling) to explore patterns.

Major findings:
Broadly, the evaluation finds that CFS implementation across contexts successfully adheres to the three key principals of CFS models: inclusiveness, child-centeredness and democratic participation. All actors interviewed, except for in one school, appear to have internalized and actively taken up the concept of CFS. Ministries, as well, have embraced the concept of CFS. The main perceived challenge is a lack of resources to support CSF – from instructional materials to trained teachers.

Observations:

- **On education in poor countries**:

- A major challenge in many developing countries, including the six included in this study, is expanding inclusive education to include students with disabilities.
- Also, although teachers interviewed have clearly internalized the importance of community and parental involvement in schools, the teachers.
- Interesting to note that teachers are somewhat more positive than students in their assessment of school climate

What have we learned?
This evaluation is one example that not all effective programs are resource intensive – the positive outcomes observed in this study, changes in teachers', families,' and communities' attitudes, were achieved mostly through funds channeled through the Ministry of Education to support community/school led initiatives based on the CSF model.

UNICEF (David Clarke, conducted for Unicef)
Independent Evaluation of UNICEF Education Programme Improving Access to Quality Basic Education in Myanmar (2006-2010)
2010

Why selected for in-depth review:
The evaluation provides a comprehensive analysis of the context and the framework for program implementation, focusing on perceptions and attitudes of key informants. The methods and approach are not as

rigorous, but the evaluation also serves to enhance the geographic diversity of our sample.

Evaluation approach/methods:
Methods include a document review, interviews and focus group discussions with key informants, and a rapid situation analysis of the education sector, including field visits to observe and conduct interviews/focus groups with local-level actors (teachers, students, members of parent teacher association).

Major findings:
The evaluation notes that the three main outputs were achieved (increased access to and quality of ECD program, increased equitable access to primary education through the Child Friendly Schools programme, and improved access to learning about "life skills"), but major challenges are "lack of MoE ownership," failure to target the most disadvantaged children, a need for better data to inform monitoring and evaluation, and overall resource shortages (teachers, supplies, etc.). The program is found to have "limited impact on the quality of teaching and learning."

Major observations:

- **On aid-supported education activities**: Like many evaluations, this one considers the program relevant because it was aligned with government policy and "developed models of good practice that can be taken to scale." What does that mean? Many different things across different contexts and evaluations – these words have become jargon.

What have we learned?
The lack of positive findings is in part blamed on the failure to develop strong monitoring and evaluation structures, to follow an education sector plan, and lack of exit strategy (e.g., no efforts were made to ensure that the policy is sustainable). These conclusions are repeated over and over again across evaluations. When will we learn? And to what end can these alarming consistencies be explained by the evaluation process itself?

UNICEF (Anna Haas, independent consultant)
Evaluation of UNICEF's role as Lead Partner in the Education Sector in Sierra Leone

2012

Why selected for in-depth review:
This evaluation assesses the role and mechanisms of aid providers in contributing to policy dialogue and educational provision. Specifically, the evaluation assesses UNICEF's role as the leader of the aid-supported Education Sector Plan in Sierra Leone. The evaluation is also one of the only *formative* evaluations encountered, meaning the objective is to "learn from past experience and provide guidance on how UNICEF can best fulfill its role as Lead Partner in the years to come" (p. 6).

Evaluation approach/method:
The evaluator conducted interviews with 22 actors, reviewed relevant policy documents, and observed the 2012 annual Education Sector Review process in order to review the relevance, effectiveness, risks and benefits of UNICEF's role as Lead Partner.

Major findings:
Overall, the evaluation finds that UNICEF was effective as the Lead Partner – successfully fostering stronger coordination in the education sector. Key achievements include establishing, chairing, and managing the Educational Development Group and UNICEF's consistent insistence on government involvement and control of the education sector. Major areas for improvement are: improving the clarity of roles and responsibilities among donors, better integration of work of all stakeholders, and more regular planning and monitoring.

Observations:

- **On aid-supported education activities**: Those interviewed note a change from the "let's do it for them" to a marked "insistence" on working with and through the government – to give the government control.

What have we learned?
Evaluations again and again mention that monitoring and evaluation capacity (at the government – or aid recipient level) needs to be improved. There are a few analyses of the trade offs between investing time and energy in improving monitoring and evaluation capacity, versus providing and improving education delivery, nor of different strategies for capacity development. Capacity development does not lend itself to impact evaluation in the traditional sense – you can't

randomize which government official or government ministry you provide training/capacity development for – but surely other strategies would provide valuable feedback for development agencies' efforts to improve capacity.

UNICEF:
Democratic Republic of Congo: Evaluation du Programme Education de base 2008-2012
2012

Why selected for in-depth review:
The goal of this evaluation was to evaluate the activities, process, and results of the UNICEF Basic Education Program for the DRC from 2008-2012. This study was selected due to the quality of its methodological approach, and to provide a case study of education in emergencies and progress towards the MDGs after a long period of political instability and conflict. Some areas of the DRC are classified as post-conflict, while others still experience conflict, therefore DRC provides and interesting case of education and fragility. The main potential beneficiary groups that were affected by program interventions included displaced, returning, and host populations in emergency and transition zones.

As stated in the evaluation, one of the added values of the UNICEF Basic Education Program in the DRC is its adaptation to the realities and priorities of the social, economic, political, and security situation in the DRC. Despite its great utility and importance, education in fragile states oftentimes very under-funded due to hesitancy on the part of donors who typically prioritize other sectors, or other sub-sectors within education.

The project includes a new partnership framework, centered on improving the quality of life of Congolese women and children. Goals include ensuring quality formal and informal education to children (in a "secure, healthy, and integrated environment"), while also focusing on gender equality. More specifically, the Child-Friendly School approach ensures that children are educated. The Child-Friendly School approach is based on partnership between schools and communities.

The projects evaluated were the Integrated Development Young Child project, Quality of Primary Education project (included infrastructure), Adolescent Participation and Development (life skills

development for adolescents), Education in Emergencies (to address specific regions of the DRC vulnerable to emergencies), Education in the Transition zones (unique to Eastern DRC—transitioning from conflict to development).

Evaluation approach/method:
As the program was managed jointly by the Ministry of Foreign Affairs and International Cooperation and UNICEF, various central and decentralized structures involved in M&E included an interministerial committee and provincial coordinating committees. The evaluation was conducted by a team comprised of two international experts, two local experts, and forty-eight field surveyors.

The study examines the planning context, interventions, results, and impact of the basic education program, focusing on implementation gaps, constraints, weaknesses, and achievements as well as sustainability. The evaluation aimed at providing UNICEF, the Congolese Government, and various technical partners with recommendations for future programming.

The methodology incorporated a participatory approach, implicating actors at all levels. Using a mixed-methods approach, the evaluation integrated information from diverse sources and triangulated the information with a real-world lens. The evaluation process included (1) documentary research at the central level, and the consultation of diverse partners within the education sector, and (2) field visits with the education administration, which consisted of semi-directed questionnaires in five provinces, including Katanga (where a case study was realized). Semi-directed interviews took place with various managers and program partners. Data was analyzed using triangulation, which extrapolated information related to the most pertinent issues, according to the evaluators, using the specific expertise of the consultants, available documentation, and information provided by respondents. Within the UNICEF evaluation system, an external independent company reviews and rates all evaluation reports.

Major findings:
The planned strategic outcomes were achieved, or nearly achieved, for the first three years, according to the evaluators, with sufficient resources and distribution of resources.

The evaluators particularly note that activities involving local populations advance the following objectives: learning, the school environment, hygiene, and the security and well-being of children. Remaining challenges, according to the evaluators included an "underutilization of available funds, insufficient monitoring and evaluation, and unfinished school buildings," (p. 10). Surveys by local beneficiaries confirmed the program's success.

The study investigated the involvement of beneficiary communities and other local actors operating on the ground, to see if they were willing to participate in the development and promotion of the educational projects and if so, if this participation had observable results. Though community involvement worked very well for early childhood education and in involving communities in rapid assessments for IDPs in emergency situations, the the evaluators indicated that the extent of local ownership remains very uneven and local actors feel frustrated by their perceived limited involvement in terms of UNICEF interventions.

The evaluation found that the impact of knowledge and skill transfer was "reduced by the lack of motivation or capability among some teachers and trainers, though on the whole demonstrate a high sense of professional duty," (p. 12). NGOs and provincial officials indicated that they have not witnessed a lack of coordination to the extent that it would negatively affect interventions, however the evaluation instrument was unable to elaborate on conclusions on this point. Numerous actors, on the other hand, indicated a lack of collaboration between various stakeholders, therefore there is some hesitancy on behalf of the evaluation team to draw conclusions on the extent of coordination.

Major observations:

On education in poor countries: The program's strong emphasis on gender and equity, including reducing barriers to education for vulnerable children outside the school system by providing remedial programs, was also assisted by taking into account psychological care of children in emergency zones. However, the evaluation notes that children in rural areas still experience challenges to access education and more needs to be done to facilitate their access as well as the access of children with disabilities.

The evaluators note that wareness activities have had a positive impact on the political level as well as facilitated cooperation between partners. Future directions for funding agencies would be to develop and capitalize off the increased involvement of grassroots organizations. Even when actors at all levels are consulted, the program will not be sustainable without increased expertise and training for technical directors.

On aid-supported education activities: Funding for activities was greatly curtailed due to political crisis, posing challenges to the feasibility of the program. While on the ground, program objectives of certain funding mechanisms, as well as difficulties with disbursement, delayed implementation.

The evaluators note that while most funds were well-managed, shortcomings could be addressed by reducing delays in the delivery of assistance (for example, distributing school kits or building and restoring schools, particularly in remote areas). Additionally, the evaluation notes that involving local communities and local resources has been largely insufficient, though more local involvement would lead to economies of scale. The evaluators suggest that UNICEF should ensure that local instruments have an integrated approach to structural challenges to programming, through the "Child Friendly School Approach," which can eventually be self-sustaining.

In terms of coordination mechanisms, the evaluators note the importance of defining roles and responsibilities to facilitate program alignment. For instance, as an example, they note that the Government might invite each partner to indicate where they would like to intervene.

The evaluation notes that measures should be taken to assure that school is free, and that all employees of the school system have a motivating career with decent salaries. Additional findings call for investment in educational quality as well as a diverse pedagogical offering; and in particular in road and school infrastructure in flood-prone areas.

On evaluating aid-supported education activities: The program is vast geographically and thematically, and therefore monitoring and evaluation and tracking the disbursement of funds, was challenging, according to the evaluation. UNICEF acknowledges that "it cannot do it all" but should be able to mobilize and guide targeted local populations to take charge of educational activities, to achieve the greatest impact. Finally, the evaluation notes the need for more effort

towards quantitative data collection, taking into account different indicators. The evaluation calls for capacity-building of M&E at the regional level and suggests this can be achieved by providing training in M&E, as well as by better coordination between UNICEF field staff and institutional actors, and by improving information circulation at all levels to provide a global snapshot. The evaluators point out that though staff often have a heavy workload with new programs, UNICEF can help alleviate this workload by setting up a documentation service or at the very least, an internal electronic platform to help centralize information, and a cost accounting system.

What have we learned?: Themes: Education and fragility, sustainability

According to the evaluation, an intervention hypothesis establishes how collective problem solving can be mitigated or resolved by the program / project implementation. The evaluation also provides a case example of education and fragility. Additionally, the evaluators indicate that participatory approaches and the involvement of local actors facilitate sustainability. Yet, the evaluators highlight that if there is rhetoric of community based support but the community does not feel involved, this will lead to frustration. Additionally, the evaluation indicates that even if the conditions of an integrated approach to sector financing across national, regional, and local actors are met, and all voices are heard, there needs to be adequate training for those working in operations. The evaluators suggest that a focus on the broader impact of stakeholders at all levels within the sector, so that each individual stakeholder can contribute systematically (affecting educational quality and the school environment), and by tracking the progress made at that level, may be useful in assigning roles and to assure implementation.

This evaluation was also useful in adding dimensions of humanitarian assistance and education in emergencies to our synthesis. Humanitarian assistance and development assistance have been typically viewed as operating in separate spheres however as education is a long-term process, it is important to have an integrated approach to programming in contexts of fragility.

UNICEF, JIMAT Development Consultants, Ifakara Health Institute for UNICEF & Government of Tanzania

Evaluation of Government of Tanzania and UNICEF Interventions in 7 Learning Districts
July 2013

Why selected for in-depth review:
This evaluation utilized a unique mixed-methods approach: document analysis, group interviews, and a quantitative difference-in-difference impact estimate.

Evaluation approach/method:
The evaluation of UNICEF's country program in Tanzania is comprehensive, covering sector wide programming in health, child protection, water and sanitation, and education. Methods include household survey and interview data analyzed through a difference-in-difference approach comparing targeted "learning districts" (LDs) with non-learning districts (NLDs), along with an "ethnographic approach aimed at exploring key child survival, growth and development practices."

Major findings:
The evaluation draws a number of conclusions, none of which are clearly linked to data and analyses. The conclusions related to education include: (1) pass rates increased, particularly for girls. This is likely due to advocacy and affirmative action policies designed to achieve gender parity in upper secondary school. (2) Supporting schools through public-private partnerships should be supported and scaled up, in particular through initiatives that grant schools autonomy in fund raising. (3) The Whole School program was well liked by community members and would be easily scaled up. One notable achievement of this initiative is that it encouraged district education officials to develop routine supervision practices. Another key finding highlights the limits of "cascade trainings," in which foreign "experts" train high-level education officials, who in turn train the district officials, who in turn train community leaders, who in turn train teachers and parents. The authors demonstrate that this type of training results in a "funnel shaped" resource allocation, with more resources being concentrated in institution-based residential courses, and very few resources reaching the grassroots. As a result, the quality of the training at the community level is often quite low.

Major observations:

- **On education in poor countries:** Affirmative action to improve gender parity? The evaluation finds increased pass rates among girls, which the authors argue is the result of UNICEF –funded and government coordinated campaigns to improve girls' access to secondary education. There is no discussion of improved learning among girls, however. What are the trade-offs?
- **On aid-supported education activities:** "Too broad a coverage of thematic areas or of districts with a standard package of supply-led interventions spreads UNICEF human and financial resources too thinly and is not supported by evidence of impact, suggesting the importance of taking up more demand-driven, well researched and customized activities that target fewer (than the current seven) districts for testing innovations so as to guarantee the depth required to sufficiently reach the grassroots"
- **On evaluating aid-supported education activities:** Again, this evaluation, like many others, does not clearly link findings to data sources/analysis (e.g., where does the claim that public-private partnerships should be pursued come from?) However, unlike many other evaluations, this one does describe potential limitations, mostly having to do with attribution. Another observation: many evaluations, including this one, evaluate programs based on their relevance to country priorities and policies. As one might guess, almost all are deemed "highly relevant," given their focus on improving educational quality. What purpose does this evaluation criteria (as used) serve? Why not focus more on whether or not the projects are aligned with policy, rather than priorities?

USAID/ World Learning
Action Communautaire pour l'education des filles: Evaluation finale (2001-2005)
June 2005

Why selected for in-depth review:
This evaluation was selected because of its emphasis on participatory approaches and inclusion of multiple stakeholders. The evaluation aimed at informing communities and implementing agencies. Additionally, the evaluation applies a mixed-methods approach, and was a joint evaluation between an NGO and a bilateral donor. The description of the methodology was very clear and specific to each

case study. Given Sida's interest in girls' education, the objective of the program was highly relevant, and there is a focus on marginalized rural areas.

Evaluation approach/method:
The underlying hypothesis of the project was that the best way to advance girls' education is by stimulating community participation via organizations. The evaluation used mixed-methods to conclude that objectives were achieved using both methods.

Major findings:
The evaluation finds that the project, rooted in participatory approaches, good communication, and social mobilization to achieve buy-in for girls' education in rural areas, succeeded in achieving high participation, and in mobilizing communities. World Learning supported and trained three local NGOs to put the project into place across ninety-one communities within five departments.

Major observations:
On education in poor countries: The evaluators note that the short duration of the program complicated finalizing all steps for community leadership of initiatives. Girls' education received much more support due to the immersion and inclusion of local communities by NGO staff.
On aid-supported education activities: The study indicates that capacity to stimulate buy-in and local/social mobilization included communication with local stakeholders at all phases of the project, in terms of implementation, analysis, discussion of obstacles and strategies, up until the financing of micro projects. The evaluators note that social communication was very good and useful in stimulating buy-in on the local level.
On evaluating aid-supported education activities: The main difficulty confronted by this evaluation, according to the authors, is the lack of available or reliable statistics, which is pervasive at the national level.

What have we learned?
The evalutators conclude that community-based approaches, and capacity building of local NGOs, are a potentially useful tool to accompany decentralization in terms of knowledge and financial transfer.

USAID
Assessment of the USAID Assistance Program to the Reform of the Benin Primary Education System
August 2005

Why selected for in-depth review:
This evaluation, which focuses on pedagogical reforms and assistance with institutional planning at the primary and secondary level, was selected because of its mixed-methods and community-based approaches. The evaluators signal the impact of USAID/Benin's activities on the final beneficiaries: children, parents, teachers, directors, and communities at large… for example, the impact of teacher training programs, children's acquisition of knowledge and competencies, and the increased role of parents and communities in school management. Finally, it is interesting to compare this evaluation with the AFD/DANIDA/MCPD evaluation on Benin, and Benin is a good case study for observing the dynamics of decentralization. Activities evaluated were holistic, including: computerized management of school statistics and disaggregated data, development of a planning tool for school development, a system of financial management based on budgeted reforms, and community/school-based programs (including support for parent associations).

Evaluation approach/method:
The evaluation assesses USAID's assistance to primary education in Benin to date, determining strengths and weaknesses in implementation and identifying areas for future collaboration, as well as past, present, and potential future constraints. The evaluation utilizes a mixed-methods approach to monitor the reorganization of the primary education structure (known as the "New Study Program").
More specifically, several activities that were assessed were:

- Sustainability of current USAID-funded technical assistance and training to support school district operations and in-service teacher training
- Implementation and performance of the "New Study Program" in private schools, comparing findings with public schools
- Current teacher training model in Benin
- Impact and relevance of the USAID/Benin education program

The methodological approach was rigorous and thoroughly described, particularly in terms of qualitative data (primary method used, though mixed methods): evaluators met with key informant and focus groups with USAID and government officials, especially those directly responsible for design and implementation, school directors, teachers, parents, and school visits (observations). Of note, there were also focus group interviews with main beneficiaries. The evaluators note that though children were not interviewed for this evaluation, they were "informally addressed and observed when team members visited schools," (p. 4). The evaluators "synthesized their observations and findings to identify points of commonality and difference across various geographic contexts (e.g. north/south, rural/urban, small/large cities, etc), to facilitate triangulation and the robustness of findings as well as allowing for generalization yet accounting for differences in context," (p. 4). However, at the same time, the evaluation period was too short for the evaluators to consult with a large number of actors. Data limitations were also addressed in this study.

Major findings:

In terms of reforms to education, the evaluation notes that while USAID/Benin's impacts at the national level are significant, they are even greater at the local level. The evaluators state that local associations, particularly parents' associations, are better prepared and have established networks at the provincial and national levels; they also point out that in certain communities a variety of organizations are now involved in the debates on education.

Yet, the evaluators note that despite a good extent of success, some communities are still not actively promoting education initiatives. The study notes that limitations are persistent, particularly given decentralization, as the integration of local government (communes) into the national education system is complex. "The vision of a centralized school system clashes with one of the school as a responsibility of local government, and the development of local schools is tightly intertwined with local sociocultural and economic realities," (p. 2).

The findings note a lack of competency among school inspectors, affecting education sector management and hindering decentralization. Moreover, the evaluators note that "new financial procedures have placed additional burdens on school inspectors, since they have not been trained in financial management," (p. 2). The evaluators note that at the school level, there are too few teachers, crumbing infrastructure, and poor working conditions.

Major observations:

On education in poor countries: The study suggests that schools should be able to manage their own budgets, to effectively respond to these various problems. Additionally, the evaluators remark that support to community organizations helps mobilize debate on education issues.

On aid-supported education activities: The context of decentralization particularly necessitates a framework for teacher training centers in terms of their organization, functions, administration and the curriculum design. Additionally, the evaluators suggest that funding agencies should "revitalize teacher networks, reinforce the capacity of school inspectors through close collaboration, provide practical training to teachers in the execution and use of results of student assessment, and ensure that curricula and guides are clear and adapted to the level of students in teachers (both in terms of language and volume)," (p. 3). Funding agencies should also "institutionalize a coherent and systematic communications program on education in general including the objective of any reforms, reinforce and facilitate collaboration between local authorities and the education system, work with women's groups to further promote girls' education and develop mentoring and tutoring programs for regions where girls' participation in education is weak, as well as plan and implement strategies for the education of other disadvantaged groups," (p. 3). Moreover, the evaluators suggest that the funding agencies "should expand awareness on the prevention and treatment of HIV/AIDS," (p. 3).

On evaluating aid-supported education activities: The evaluators indicate that evaluations can be improved upon, and implementation can be improved upon, if evaluations involve those directly responsible for the design and implementation of the strategy (technical directors, etc) as well as interviewing the main beneficiaries of education interventions. The evaluators suggest that even children can be informally addressed and observed while in school. The evaluators also promote synthesizing observations and findings across the evaluation team to identify points of commonality and difference across various geographic contexts, to facilitate triangulation.

The evaluators indicate that the main difficulty of this evaluation was the lack of reliable and available educational statistics, particularly at the national level. Therefore, it was difficult the evaluators to measure the achievement of several of the project objectives. The project was however able to put a database into place to do the necessary analyses --> but did this compensate for the overall lack of reliable statistics?

What have we learned?
Though there is a widespread problem in acquiring reliable and available educational statistics in many developing country contexts, the evaluators note that qualitative data also remains essential for a pilot project based on a new approach or project, particularly as triangulation can address data limitations in statistical analysis. Yet, triangulation and qualitative data take time and oftentimes the evaluation team is faced with time constraints, and unable to consult with a wide variety of actors. Finally, the evaluation tells us that community events and close interaction with civil society help disseminate information about educational interventions and evaluations.
"There are many constraints to quality…. these can generally be summarized in terms of **continuing weaknesses in the ministry's institutional capacity and in the involvement of communities in school affairs…**school organization and management therefore suffers from a lack of coordination and monitoring of instructions from the top…politicization of the educational administration constitutes another brake on quality, meaning human resources are not being used optimally," (p. 2).

USAID*Program Evaluation for USAID - Guinea Basic Education Program Portfolio*
May 2006

Why selected for in-depth review:
This evaluation was selected because it examines community-based interventions to increase enrollment and co-management of schools, within the context of a fragile state and with a special emphasis on girls and rural children. The evaluation approach was also unique in that instead of replicating the approach of earlier evaluations, site visitors were instructed to write field notes "based on their observations of teacher practices, including interaction with students, the use of active teaching methods and student assessment techniques, the availability of pedagogical materials, and gender-related practices," (p. 5). Additionally, the evaluation team also placed a "strong emphasis on the collection and analysis of documentation relating to program implementation," (Eval: USAID/Guinea, 2006: 5).

Evaluation approach/method:
A multinational team of six researchers from Benin, Canada, Guinea, Senegal, and the United States conducted the evaluation research. The

evaluation sought to determine "the principal capacity-building activities and their effects on policy, sectoral strategic planning, management, and decision-making in education," the contribution of USAID, and USAID-funded programs, to educational quality at the primary level, the contribution of USAID to supporting civil society organisations, the program's approach to intersectoral issues (gender, rural/urban, HIV/AIDS education), and the "sustainability of strategies, models and approaches in all of these activities," (p. 5).

The evaluators note that "in addition to interviews, the team also adapted a classroom observation tool originally developed by EDC for tracking change over time….rather than replicate earlier studies… visitors wrote field notes based on their observations of teacher practices, including interaction with students, the use of active teaching methods and student assessment techniques, the availability of pedagogical materials, and gender-related practices," (p. 5).

Noting the importance of documentary analsysis for any evaluation, but especially in a complex context such as Guinea, the evaluators addressed a large information gap: the lack of any external evaluation within the past ten years (possibly given the political context). Therefore, the evaluators had to collect and analyse hundreds of documents, deemed "essential to developing a deeper understanding of the various activities funded by USAID and the context in which program implementation occurred," (p. 5), resulting in a data collection matrix and interview guides for various stakeholder groups (ministry decentralized structures, school principals, teachers, students, implementing partners in community-based education, local NGOs, and for civil society groups including parent associations, coordinating bodies, alliances for girls' education, and rural development committees), favoring open-ended questions. For instance, " evaluation team members often asked respondents to identify areas in which methods and strategies introduced by projects were most useful to them, how these methods were applied, and with what results," (p. 5).

Major findings:

The evaluators found a positive impact of activities on community participation, and significant in terms of promotion of greater transparency and governance. The evaluators indicate "democratic principles are taking root in the practices of parent associations and are generating a ripple effect in the political life of the communities," (p. xiii).

The evaluators found that gender and the rural/urban equity gaps are ameliorating, yet note that "it is difficult to isolate specific impacts because of the multiplicity of interventions on the part of the government, other technical and financial partners, and members of Guinea's civil society," (p. xiii).

Major observations:

On education in poor countries: "Decentralization of planning and decisionmaking has been met with relative success, although devolution of budgetary authority has proven more difficult to implement," (p. viii). The evaluators note that one step in the right direction that has been achieved is the development of a reliable management information system, which should help better ensure an equal distribution of resources. The evaluators note "evidence of a shift from centrally-driven decisionmaking to the more broadly participative process that is now an integral part of the Ministry's practice," (p. viii).

The evaluators note that community involvement has increased the demand for education and, to some extent, the quality of schooling…yet outcomes are fragile, as demand generated by eduation promotion activities cannot always be met, and there is a lack of effective coordination at higher levels to help grow the impact of grassroots organizations.

On aid-supported education activities:

On evaluating aid-supported education activities: as in most other evaluations, no mention of how findings relate to researchers' influence, for ex, through their interactions with participants*On evaluating aid-supported education activities:*

The evaluators note that rather than solely relying on evaluation templates, a good practice to track change over time includes writing field notes based on observations, and developing classroom observation tools that are flexible and adjusted to context.

What have we learned?:

While important in all contexts, the collection and analysis of documentation relating to program implementation is of particular importance in post-crisis situations where there has been likely a deficit in documentations and external evaluation of education programs.

World Bank - Independent Evaluation Group

From Schooling Access to Learning Outcomes: An Unfinished Agenda – An evaluation of World Bank Support to Primary Education
2006

Why selected for in-depth review:
The World Bank is one of the most prominent actors in global educational development, and this evaluation in particular is well organized, well written, and provides a comprehensive and critical overview of World Bank support to education from 1960 – 2005.

Evaluation approach/method:
Literature reviews, review of WB documents, inventory and review of WB primary education portfolio, field-based evaluations of completed primary education in 8 countries, field-based country case studies in 4 countries. Case studies included interviews with Bank and local managers, donors, agencies, beneficiaries.

Major findings:
The evaluation tracks the evolution of lending to education from 1960 to 2005 and finds that the number of education investments managed by sectors other than the education sector has increased, due to "a proliferation of projects with relatively small primary education components."
Enrollment growth over the last twenty years has been expansive, and can partially be attributed to Bank support for infrastructural development, although in many cases the elimination of school fees was the driving force. The authors note that much of the expansion in access has come through projects managed by Bank units from other sectors – through social funds and public works projects. However, one risk of these programs is that "their focus on quantitative growth can overshadow improvements in educational quality and outcomes."
Regarding conditional cash transfers (CCTs): Bank experience with CCTs suggests that CCTs can be effective in increasing enrollment (although not necessarily in improving learning), but require strong targeting mechanisms, monitoring requirements, and administration structures.
Many World Bank projects include an emphasis on improving educational equity, but equity is typically framed in terms of access only. About half of the evaluations included in the World Bank review dealt with "equity of treatment" – eliminating biases against disadvantaged children in the classroom, or equity in learning outcomes.

Relatively few projects have assessed learning improvements over time, but among those that have, learning has improved for disadvantaged and in some cases the gaps between advantaged and disadvantaged students has reduced significantly.
The three Bank-supported countries that have seen the strongest improvements in learning (Ghana, India, and Uruguay) all have explicit national education policies and strong national commitments to educational development. They also reveal what the authors refer to as "sequencing of learning outcomes support:" (1) provision of basic inputs (e.g., school construction), (2) teacher support, and (2) pedagogical renewal to targeted programs for the most disadvantaged.
Projects designed to provide support for local school governance (such as school-based management) have in general been more effective than support for central management—largely because efforts to improve central management have "not been sufficiently founded in institutional-political analysis." However, community management has been linked to improved facilities and staffing, but not improved instructional quality or learning.
The evaluation notes that recent projects have given more attention to evaluating outcomes (rather than inputs/outputs). However, the following challenges persist: (1) systems for monitoring, student assessment, and research are rarely used in decision-making, and (2) "lingering problems with data quality" in countries where EMIS has been developed.

Major observations:
On education in poor countries: Decentralization has been a popular development strategy, but there is evidence that decentralization can have adverse effects on equity in access and quality. The issue of teacher recruitment and performance is often overlooked – partly because the experiences that do exist have been unsuccessful: e.g. contract teachers, financial incentives to bring teachers to rural and underserved areas are unsustainable and often not successful.
On aid-supported education activities: Many evaluations emphasize the need to improve education planning, policymaking, and financing at the central level. However, this World Bank evaluation notes that aid-support to these activities have (in general) failed to meet the targeted outcomes. Is aid money better spent on specific projects—pedagogical or infrastructural—rather than management and policy projects?
On evaluating aid-supported education activities: The Bank's

analytic work in education has not focused on learning outcomes or equity – despite Bank commitment to these aspects.

What have we learned?

Drop out and grade repetition remain consistent challenges long after countries expand enrollment in basic education.

Divergent classifications lead to different conclusions and policy recommendations: For example, school-based management programs are categorized as "institutional strengthening" in the Inter-American Development Bank's review, while financial support for school committees is classified as an "innovative form of infrastructure development" in the World Bank's review. These discrepancies have been noted by other scholars – see Evans and Popova 2015, for one example. What are the implications?

In regards to M&E – it seems that establishing the appropriate M&E systems is only one step, it is also important to ensure that the national political culture encourages the use of such systems in policy making. This is more likely if local institutions guide the decisions regarding how and what to monitor and evaluate

World Bank (David Evans and Anna Popova)
What Really Works to Improve Learning in Developing Countries?
February 2015

Why selected for in-depth review:

This working paper was selected because it is a synthesis of six existing systematic reviews/meta-analyses of education interventions to improve learning in low and middle-income countries. It is useful to our synthesis as it finds that systematic reviews sometimes reach starkly different conclusions, driven by differences in the samples of research as a result of inclusion/exclusion on the basis of methods, by each review.

Evaluation approach/method:

The study examines six reviews and explains divergent findings across the systematic reviews/meta-analyss: Conn (2014), Glewwe et al. (2014), Kremer, Brannen, & Glennerster (2013), Krishnaratne, White, & Carpenter (2013), McEwan (2014), and Murnane & Ganimian (2014). The target audience consists of evaluators, academics, and funding agencies, and the overall aid community.

The synthesis approach utilized in this paper takes a purposive sample of existing meta-analyses and synthesis reviews, then examines the main conclusions, exclusion rules, variation in composition and categorization of all the reviews. It then examines the extent of heterogeneity across results within intervention categories as well as differences across categories.

Major findings:
"In the past two years alone, at least six systematic reviews or meta-analyses have examined the interventions that improve learning outcomes in low- and middle-income countries. However, these reviews have sometimes reached starkly different conclusions: reviews, in turn, recommend information technology, interventions that provide information about school quality, or even basic infrastructure (such as desks) to achieve the greatest improvements in student learning. This paper demonstrates that these divergent conclusions are largely driven by differences in the samples of research incorporated by each review," (p. 1)

Major observations:
On evaluating aid-supported education activities: The evaluators find that much of the divergence in conclusions is driven by strikingly different compositions of studies across the reviews:
Of the 227 studies that look at learning outcomes, only three are included in all six systematic reviews, whereas almost three-quarters (159) are included in only one of the reviews. While some of these compositional differences are driven by explicit exclusion rules (e.g., some reviews include only randomized trials and one focuses only on evidence from Sub-Saharan Africa), many are not. This divergence does not mean that reviews are incorrect in characterizing what works well: The main conclusions of each review are supported by evidence from papers that attempt to explicitly establish a counterfactual. Indeed, the strongest positive results in each review are driven by randomized controlled trials. However, each review incorporates different evidence, leading to different ultimate conclusions," (p. 3)

The least systematic form of analysis, the narrative review, can incorporate the largest number of studies but requires non-scientific tallying and weighting across studies, and is the most susceptible to influence by authors' prior beliefs. The most systematic form of analysis, the meta-analysis, may limit the included studies because of stringent requirements on the data reported in order compute strictly

comparable effect sizes, and it may fail to illuminate the mechanisms behind the most effective interventions. Each method has flaws that keep it from being both systematic and exhaustive," (p. 3).

What have we learned?

Systematic reviews may not be exhaustive, as each meta-analysis or review includes different evidence, and additionally, may restrict their sample to only specific methods, leading to very different conclusions at times.

F. Case studies

CASE STUDY 1

Agence Francaise de Developpement (AFD), Denmark Development Cooperation (DANIDA), and Benin Ministry of Development, Economis Analysis and Forecasting (MCPD) Joint Evaluation

Evaluation a mi-parcours du Plan decennal de developpement du secteur de l'education du Benin (PDDSE 2006-2015)[4]

February 2012

Why selected for in-depth review:

This evaluation was selected because it scored highly on our review criteria, particularly in terms of relevance to our synthesis, its objectives (in particular, the management, leadership, and facilitation of sectorial dialogue, and sector financing). Additionally, it is a joint evaluation, between two funding agencies but also with the recipient ministry of development. Though the evaluation is very descriptive, and methods are not discussed at great length, the evaluation totaled 228 pages in length and offered insight into the challenges of policy implementation in a decentralizaed context, as well as challenges to data collection and capacity-building in terms of monitoring and evaluation. As it was mostly conducted in French, the evaluation provides geographic and linguistic diversity, particularly as Francophone West Africa is disproportionally represented on the lowest tier of the UNDP Human Development Index.

[4] All quotations are my French translation, here and elsewhere in this report.

The AFD/DANIDA/Benin Ministry of Development Joint Evaluation (MCPD) provides a lengthy discussion of the process of decentralization of the education sector in Benin, which is relevant to our synthesis, as the topic is frequently raised within the evaluations we have thus far reviewed. This evaluation, however, provides the most thorough discussion of the dynamics that we have seen thus far. The evaluation also was initially commissioned at the request of the recipient country, as Benin wishes to develop more capacity in terms of monitoring and evaluation.

Why selected for case study:

This evaluation traces the process of decentralization and implementation of reforms, and notes significant process-related challenges. It is also a joint evaluation between two funding agencies and an aid recipient, and was initiated at the request of the aid recipient in recognition of the need for external consultation due to challenges at fully realizing decentralization, and due to a desire, on behalf of Benin, to develop its own capacity in monitoring and evaluation.

Overview of education sector in Benin and justification for evaluation: A new educational system was put into place in Benin in 1990, yet has not achieved the expected progress in sector management or results and is covered up by more and more new reforms (Attanasso, 2010). Yet, this implementation of these reforms has a cyclical effect, as the sector is increasingly confounded by "fragmented management structures," "bottlenecks" in data collection and information management, and challenges in human resources management (Attanasso, 2010). As the Beninese educational system is faced with numerous governance problems, the ministry requested external evaluation by the Agence Française de développement and DANIDA (Denmark's development cooperation agency).

Despite the key role of aid to education in Benin, information on interventions provided by external funding agencies is "poorly distributed" (Interviews, August and September 2015). Therefore, joint missions by funding agencies assist in supplying sectoral ministries the information on aid to education that is "not revealed to them by the central ministry of budget and finance" (Interviews, August and September 2015). However, "aside from press releases and

a few newspapers that address the results of these external missions," the results found by evaluations conducted by external funding agencies are usually subject to recipient government approval before dissemination to end users and remain exclusively within the government (Interviews, August and September 2015). Since reports by various donors are either non-harmonised in their findings or not distributed, there is limited information concerning conditions for aid distribution (Interviews, August and September 2015). Funding agencies experience difficulty in procuring contact information for beneficiaries as well as civil society representatives with an interest or involvement in aid-funded educational activities (Interviews, August and September 2015). Moreover, there are few documents on the compensation of individuals implicated in projects funded by aid agencies (Interviews, August and September 2015). However, there are very active civil society organizations in West Africa, and our synthesis overall illustrates the importance of strengthening these organizations, and in particular, fostering the role they might have in promoting aid to education and contributing to and disseminating results.

Evaluation approach/method:

This mid-term evaluation took place just before the last phase of Benin's ten-year education sector development plan, and was initiated by the governments of Denmark, France and Benin (represented through the '*Observatoire du Changement Social*'). The evaluation assessed the extent to which the objectives and assumptions of the education plan remained relevant, the results achieved over the past five years, and provided lessons learned. The evaluation covered all levels of the education sector in Benin: preschool, primary, secondary, technical education and vocational training, higher education and scientific research, as well as literacy and adult education. The evaluation was based on a documentation review, data collection in Benin (including interviews with key actors at central, deconcentrated ('*départementales*'), decentralized ('*communes*') and institutional levels,"and analysis.

Major findings, observations and recommendations of the evaluation:

Early Childhood Education: Evaluators note that EFA has generated support and demand for early childhood education…yet, the evaluation questions whether the Government will be able to meet

commitments for this growing demand without resorting to financial contributions from parents.

Primary Education: In addition to progress in net enrollment due to EFA, the evaluation indicates that inequalities between girls and boys in school attendance have continued to decline. However, the evaluation notes limited progress on integrating some of the most vulnerable children: those with special needs and those that are out-of-school. The evaluators suggest that the most effective strategies for increasing enrollment are cost reduction for families and continued advocacy and awareness activities at the national level. The evaluators observe that strategies that are more focused on local needs are likely to be most successful in reaching excluded and vulnerable children, particularly as regional and rural/urban differences remain. More work remains in terms of reducing repetition rates, as the evaluation mentions some resistance from teachers.

Secondary education: The report illustrates the downstream pressure of EFA on secondary education, and gaps in girls' access to education still persist despite measures to promote girls' secondary education.

Educational Quality: Despite efforts to introduce new programs, the majority of students are not performing at grade level. In tandem with increased teacher training interventions, the report indicates the roll-out of a new competency-based approach for teaching methods, yet highlights the need for sustained training and pedagogical and material support. Decentralization has not led to the strengthening of school management, as noted by evaluators.

Decentralization: According to the report, decentralization has been more successful in the health sector than the education sector in Benin. In education, the evaluators point out that decision-making remains highly centralized with limited devolution of responsibilities. The evaluation overall reveals a great deal about the dynamics of decentralization.

In Benin, the evaluators note the creation of a new management structure at the national level, to accompany decentralization in the education sector, consisting of an oversight committee, a steering committee, a coordination committee, and a technical secretariat (to coordinate action plans, reports, and reviews). However, the evaluation notes that the management structure has not been operational because of high inactivity across various committees, either because there were too many members in a particular committee, because roles were not defined, or because the committee

stated it lacked sufficient financial resources for meetings and events aimed at coordination. While sector dialogue between the government funding agencies has improved, the government has not been a leader in the decentralization dialogue nor in coordination with donors regarding decentralization, as evaluators note. The evaluators suggest future steps would consist of involving donors in reflection on the integration of policy decisions and their strategic management. The evaluators note that the coordination of sub-programs was ineffective because of poor communication between ministries, whom, evaluators observed, do not meet often enough to truly coordinate on educational policy and reforms.

The evaluation observes that consulting with the private sector and civil society organizations (though their objectives aligned with the education sector plan) has achieved mixed results. Though several NGOs conducted pilot projects involving participatory approaches, the evaluators indicated that they were not included in the education plan though they should have been included. Therefore, the participation of civil society organisations has not really been adequately measured since several organizations that should have been involved in the sector plan were not included. Overall, in Benin, the evaluation illustrates that deconcentration (structures and human resources), and decentralization (government services and management) process has made no significant progress within the education sector.

Evaluation and Measurement: The evaluators find that "current information systems are not capable of informing decisionmakers, particularly as the use of indicators is greatly limited by the weakness of databases in the education sector," (AFD/DANIDA/MCPD, 2012: 48). The evaluators note that even wen information does exist, it is often unreliable…"the collection and analysis of data remain highly centralized and the production of annual statistics is typically not without considerable delay…and that the use of key indicators is not applied across ministries, and therefore performance reports lack a solid informational basis and credibility," (AFD/DANIDA/MCPD, 2012: 48).

In this report, the evaluation team noted the assistance of an evaluation management committee, as well as a local reference group comprised of a diversity of stakeholders (ministerial representatives, trade union representatives, and relative civil society groups) that

facilitated access to information sources and assisted the report enormously through sharing their insights.

Financing: The MDGs have resulted in pressure on the government to meet financing objectives (pressure to increase spending due to free enrollment in preschool, primary, and higher education; the transfer of community teachers to state employee status, and the lack of strategic management leading to more equibable resource distribution), as we have seen across evaluations. At the same time, the evaluation indicates that in this particular case, the MDGs may have weakened the presence of the secondary education ministry, due to their strong emphasis on primary education, perhaps resulting in downstream effects. The DANIDA interview revealed that ministers are unlikely to reduce the higher education budget because higher education benefits mostly the elite. Therefore, increased funding for early childhood and primary education has largely come from reducing the secondary education budget (Interviews, September 2015). In terms of budget performance, the evaluation indicates that efficiency was low across all ministries, due to cumbersome bureaucracy, insufficient knowledge of procedures by some managers, and delays.

Sustainability: The evaluation indicates that EFA was not part of the initial education plan, and therefore was not included in the financial simulations of the Plan, resulting in huge financial ramifications. Since primary teachers who were previously paid by the community are now state employees, EFA coupled with decentralization may be dangerously financially overwhelming for the state. Subsequently, the evaluators indicate the importance of evaluating the financial consequences of EFA in terms of sector-wide financial sustainability.

Equity: Despite expansion of access and improvements on girls' education at the primary level, challenges to educational equity remain, according to the evaluators. Given the high allocation of the budget to higher education, the evaluators note that the most privileged members of society tend to benefit from this particular budget allocation (Eval: AFD/DANIDA/MCPD, 2012: 51). This may be the source of "passive" resistance to decentralization at the ministerial level in the education sector, since decentralization has been successful in other sectors such as water, health, and sanitation (Interviews, August and September 2015).

Supplementary references in this Annex Case Study:

Attanesso, M.O. (2010). "Bénin: Prestation Efficace des Services Publics de l'Education: Une etude d'AfriMAP et de l'Open Society Initiative for West Africa (OSIWA)." Dakar: OSIWA.

Supplementary information from interviews:

The interviews below provided significant insight into the mechanisms described above, and in particular, enabled us to understand more fully the challenges of decentralization of the education sector in Benin, as well as evaluation preferences and evaluation use, and institutional learning among funders and recipients.

I. Interview with two AFD staff members[5]

1. Through what processes do organizations determine what to measure, how to measure, and how to use evaluation findings?
Evaluation is cultural. There has been a new evaluation push at the ministry level within France that will eventually reach the AFD. Additionally, the added-value of France is its economic approach to education, therefore there is an emphasis on economic impact of development aid activities, and the organization's priorities in the sector are education, training and employment.

The AFD has limited use of experimental designs. They tested a few experimental designs and randomized control trials, but found the results limited and lacking, as well as the design highly costly. One example is a microfinance study in Cambodia. The limited use of experimental and quasi-experimental methods is relegated to measuring the impact of scholarships and conditional cash transfers. They have also conducted regular quasi-experimental studies on the impact of school feedings.

The AFD in general is skeptical of RCTs because they are expensive and very difficult to run, the timeframe for measuring educational outcomes is long and that is the unique challenge in education as opposed to other sectors of development. Finally, when the AFD ran RCTs the agency found that there were inconclusive results and poor explanations for the results. In particular, the measurement of

[5] My French translation, here and elsewhere in this report. Original interview in French.

education programs is complicated and perhaps ill-suited to RCTs because education interventions usually take place over a long period of time. Yet, there are also examples of randomized evaluations for bilingual education/teaching methods. Quasi-experimental designs have been undertaken for teacher training and observation, as well as examining new technologies for teacher training.

The AFD evaluation department indicated that a measure they might like to further explore are evidence maps, as used by 3ie: http://www.3ieimpact.org/evaluation/evidence-gap-maps/, since they look at a multitude of factors in development. They noted that evidence maps have not really been used in Francophone countries.

2. Which sorts of evaluations are most useful for different constituencies involved in aid to education, and why?

At times, joint evaluations are conducted with recipient country ministries, keeping in mind public policy within the country and national sovereignty. In this case study, Benin was selected in particular because it is active in terms of evaluation policy and in terms of education. Some countries are more active than others in evaluation policy and they are more interested in joint evaluations, in terms of building up country capacity.

In the Benin case study, the evaluation took place halfway during the program cycle, which was a key point given Benin is in the process of decentralization, therefore this mid-point evaluation was crucial in terms of the dynamics of this process. The country is still not quite organized enough to fully carry out decentralization.

The ministries in charge of education are not inclined to significantly transfer competencies to the commune level, though this has occurred successfully in other sectors such as water and health. When services are decentralized, there are limited resources to accompany their management, and this is particularly the case in educational quality equity, and delivery (AFD/DANIDA/MCPD).

Oftentimes, the education minister has far less power to implement changes than the finance minister. However, the health minister is also typically weak, so this raises the question: is there something unique about the education sector that makes it more difficult to decentralize? In Benin, overall there has been very limited transfer of competencies from the national level to the local level, and the

dialogue of management and decision-making transfer has gone poorly.

3. What is AFD's evaluation strategy? Have there been recent changes in this strategy, in terms of the usage of experimental and quasi-experimental methods, and impact evaluations?

Given historical and linguistic ties, French development assistance to education is concentrated among fifteen countries in Francophone Africa. Evaluations are classified as strategic or programmatic. Strategic evaluations view the AFD and recipient country ministries as the end user and are generally not participatory.

Programmatic evaluations increasingly view the integration of diverse actors in the evaluations as vital, especially the role of civil society. Yet, usually evaluations are only disseminated at the ministerial or organizational level (to NGOs, which are also viewed as actors in the system), and at most to the school headmaster if on the microlevel. Statistical evidence and data overall remains the constant challenge, especially in terms of guiding reallocation of funding. When NGOs are included they are viewed as actors within the educational system. Typically, there is also a quality control group.

Decentralized evaluations reflect a desire to control the money spent. The agency (AFD) verifies that the effect of aid is generally positive to avoid the pitfalls associated with poorly organized aid programs. The primary objective here is to help improve the overall situation of the beneficiaries.

Randomized experiments make us grasp the contrary— aid as an economic strategy, it must be the most effective, efficient, resulting in objective and quantifiable improvements (reduction in the prevalence of disease, increased tuition rates . This method also refers to aid "experiment" for testing economic theories (psychological effects, externalities, etc.).

4. What evidence is there of evaluation-induced learning or change?

Evaluations are referenced and possibly consulted for future projects/allocations but there is no systematic review of evaluations. The AFD is in the middle of developing a monitoring and evaluation system, and then hopefully this will lead to more evaluation-induced learning or change. At present, monitoring and evaluation are not really part of the project cycle, like they might be for the World Bank, for example, where there is typically a completion report review before the next project. Yet, the evaluation culture in France is undergoing a

transformation, as there is a new evaluation initiative across the ministries, which will eventually find its way to the AFD.

Another main obstacle to evaluation-induced change is that as those piloting the evaluation are internal, but know that eventually they will be moved around within the organization, it is easier to not be critical when evaluating programs since it might cause problems with colleagues once they are in the same department again. Therefore, the internal evaluators (a limited number) cannot find themselves later on in the same departments that they have evaluated. An alternative model to address this challenge exists at the European Commission, where evaluators are selected for a period of three years and are very autonomous.

5. At the AFD, are evaluations typically done internally or externally?
Evaluations are done externally, as the internal evaluation team at the AFD is very small, under-staffed, and under-resourced. Yet, there is a quality control group. It is likely that the department will expand eventually. The evaluations are launched internally but undertaken externally. The problem with external evaluations is that there is pressure to be less critical of the AFD within the evaluation. The external evaluators work very quickly but at the same time, the implementing agency understands the program much better. If the external evaluators state something negative, it is also up to them to explain and investigate in depth.

Before 2006, the AFD did not systematically evaluate its interventions, and the reauthorization of funding was based on the aspired impact of the program in the annual report. However, the pressure of public opinion as well as the need to internally improve aid efficiency, led the AFD to start to follow the global evaluation movement. The AFD's programs are systematically evaluated but since the evaluation department only receives a mediocre part of their overall budget, the AFD has a preference for decentralized evaluations. Though decentralized evaluations are approximative and less precise, the AFD maintains that they do reveal any major problems and give a good idea of the program impact. As decentralized evaluations solicit the feedback of beneficiaries, this evaluation method allows the AFD to understand how aid is absorbed, and to take into account the opinions of those most affected by aid, to have human contact.

RCTs are a luxurious tool, and trendy, and though they may give scientific legitimacy, RCTs are not always appropriate to the diversity of actions led by the AFD, therefore a clear vision of other methods, such as metaevaluations, is key. The problem with RCTs is that they do not permit a global vision and the proper human contact as in a decentralized evaluation.

6. Example of exemplary evaluation:
DFID 2010 (3 country: Rwanda, Ethiopia, and Tanzania) the European Commission 2006 meta-synthesis.

7. Example of an evaluation that was not useful:
The RCTs (the example was not in education, but in microfinance). When the AFD ran an RCT found that there were no explanations for the results. The difficulty with RCTs is that the conditions are difficult to reenact.

8. Would evaluations be more useful if more funding was allocated to them?
Evaluations need political will to valorize their use. One main challenge also to evaluation is weak data availability. Some countries are developing more of an evaluation culture, like Benin which makes them easier to evaluate.

9. Can evaluations, at times, impede education program implementation? If so, have you had this situation?
The added difficulty with RCTs is that the protocol for the RCT and the evaluation take place at the same time as the program, which is very complicated to run simultaneously.

10. In the absence of financial or time constraints, how would you evaluate an aid-funded education project in Benin?
In a world without time or monetary constraints, participatory approaches would merit more attention (at the organizational level).

11. In your opinion, what should be the purpose of evaluations? How does that compare with the way in which evaluations are conducted and used?
At present, the direct application of evaluations is within project instructions. The AFD is currently developing a monitoring and evaluation system. The World Bank is a good example of how

evaluations are integrated in the project cycle, there is a completion report review before the next project. Oftentimes, data availability and collection are weak, and evaluations are easier to conduct if the country has already evaluated its public assistance.

12. In the report describing the decentralization process in Benin, it was noted that the decentralization of the educational system had more challenges as compared to the health sector. Is there something unique about aid to education that is different as compared with aid in other sectors?
Decentralization depends largely on the quality of the transfer, an efficient dialogue, and the management of the transfer. In Benin, this dialogue started to not go well. There has been very little transfer in the decentralization of education in Benin.

13. What is the role of civil society and other actors in evaluation, and more specifically, in education?
NGOs are consulted in programmatic evaluations.

14. How important are contextual considerations?
The AFD appears to place a strong value on context.

II. Interview with a former DANIDA staff member

1. The decentralization process in Benin, in the education sector, has not gone well. Is this due to a lack of dialogue? What are some of the institutional dynamics at play? DANIDA played an important role in technical transfer and accompaniment during the transition process - what are some lessons learnt for external aid agencies and their role in decentralization?
If it is part of the government's strategy to decentralize, donors can have an important role. In Benin, there was a lot of money spent on academic/tertiary education which was benefiting most well-off people in country and not often leading to qualifications needed in Benin. Additionally, there was not a lot of incentive to change this. In fact, decentralization in Benin was more complicated than elsewhere DANIDA has worked recently. When there is resistance to decentralization, the only thing donors can do is to try to influence ministries by providing evidence from other countries and to use the MDGs as an argument to move towards the target. Donors can then help assist in developing the decentralization strategy.

2. Why has the education sector had more difficulties in decentralizing in Benin, as opposed to other sectors like health and water?

The education sector in Benin was run by four agencies, and cooperation was not as good as it could have been. The government was transitioning from one education strategy to another, and tried to create a coordination unit but it never really functioned. At DANIDA there were not very many evaluations conducted overall, since the evaluation department was located in the ministry, and only the ministry has authorization to do evaluations. There are only eight to ten conducted per year. Oftentimes embassies within country will conduct their own evaluations, at a smaller scale. Benin needed someone from the outside to do the things that were already well-known; therefore the Government needed external evaluation consultants.

In terms of results communicated to beneficiaries: students, teachers, and locals were not really consulted, though results are always communicated to people within ministry, government, and other institutions involved. Additionally, there was a good mix of academia, donors, civil society organisations (international and local NGOs) in the dissemination workshop.

3. *What are some of the complexities and challenges in conducting joint evaluations?*

Joint evaluations are more work because there are more people who need to agree on the focus of the evaluation. Logistical challenges existed, as with European partners, it is easier to hold meetings in Europe – and joint meetings over Skype are sometimes complicated. There was civil society involvement in policy dialogue, through the evaluation. Joint evaluations allow for better results and access to more information, and also help lesson the demand on developing countries in terms of monitoring and evaluation.

4. What are some of the complexities/challenges in capacity building in terms of monitoring and evaluation in developing country contexts (for instance, Benin has sought external support to build up its evaluation and monitoring capacity)?

A general problem in monitoring in evaluation is that most people are working on implementation and have limited time/priority for M&E. Accountability tends to take over instead of learning. However, we need to use monitoring and evaluation for learning and not just

accountability processed. At the same time, M&E is not well-resourced, and developing indicators are not really in line with African contexts. There is a process of setting targets and trying to work out a strategy, targets were asked for by the government, but this is unrealistic. Instead of "firefighting" and looking at what the current situation is, etc., we need to think about how can we address the situation and create a strategy. There also needs to be much more training of people on the job.

5. Through what processes do organizations determine what to measure, how to measure, and how to use evaluation findings?

DANIDA conducts only eight to ten evaluations per year; other agencies conduct far more. There is of course the political perspective, what sorts of topics are prioritized, for example. Other considerations include requests from embassies, like the Benin embassy in this case study. Evaluations are also an opportunity to voice concerns – whether this is the funding agency, or the minister. In terms of evaluating aid to education in developing countries, the process of evaluation is largely crafted by the donors who are supporting education, alongside the ministries. They decide what and how to monitor, and what the milestones should be.

On the other hand, given my work with NGOs, monitoring and evaluation is quite different for NGOs, which tend to work in a vacuum. NGOs for the most part do not directly work with the government; their job is to hold the government accountable. Bilateral donors work more with government.

6. Which sorts of evaluations are most useful for different constituencies involved in aid to education, and why?

DANIDA conducts very few quantitative evaluations, especially for country-wide programs. Quantitative evaluations where you have a control group can be very effective in showing results, and they can be useful for example to measure the impact of the introduction of school canteens or the types of school canteens. Out of twenty evaluations in three years, only two used RCT-type methods, and DANIDA was pleased with the evaluation. However, RCTs risk a lot of spillover effects. Contamination (for instance, people coming in from other villages) is easy in RCTs. Additionally, ethical considerations are a strong concern. While DANIDA does not do a lot of RCTs (or evaluations as a whole), the staff had a training session in quantitative evaluation methods by an external instructor.

To do quantitative evaluation, evaluators have to be able to compare, and if the funding agency is withholding a strategy that covers the whole country then an RCT does not make sense. Additionally, RCTs are good at showing whether there is an effect, but not good at showing why or why not. Qualitative methods cannot do comparison in same way but they are much richer and more detailed in drawing out dynamics, why something worked and did not work, what were the challenges, who were the beneficiaries, etc. In qualitative methods, the impact part is where they are maybe less strong, oftentimes observers cannot see all the impacts right away sometimes, impacts take years to become observable. Quantitative evaluation, without qualitative evaluation, can easily jump to the wrong conclusions. At 3ie we had a meeting on the useful application of quantitative studies and this was a huge battleground.

7. What are some of the frustrations you've encountered in terms of evaluation?

Real-time evaluations are becoming more popular, especially in terms of humanitarian assistance (where they make sense). Yet, evaluations are always prescripted in terms of learning, as they are usually conducted towards the end of the program. In such a cycle, information and knowledge can be fed into next phase, but would have been useful had it been applied earlier. Recommendations include ways of using M&E systems as sort of more ongoing learning and evaluation.

DANIDA always made a management response that tended to respond, which maybe shows that some info generated by evaluations was not new. Evaluations are so retrospective. It is a big exercise to do an evaluation. There is always new info generated, sometimes there is not, but they just need an external person to say it and to make some recommendations for the donor community.

III. Interview with aid recipient (former Beninese government official)[6]

The process of decentralization has not gone well, at least for the following reasons. First, French culture always tends towards an

[6] My French translation, here and elsewhere in this report. Original interview in French.

excessive centralization of power, involving control, and in reality, inefficiency. The global trend is towards decentralization, but the habit of centralization of power is stubborn and difficult to break. Consequently, we have difficulty to undo the tradition of centralization. Second, oftentimes those in power wish to maintain personal advantages where there is a direct daily influence in management by the ministry, or, at least, by the regional government.

The huge gap between commitments made and daily practice are due to the lack of willpower to implement centralization and the desire to please financial partners who condition aid upon the stated commitments. The institutional organization blocks the process, for instance, teachers continue to be influenced by the ministry in schools, whereas regional directors only make slight, minor adjustments to the school system. Another example of challenges to implementation at the institutional level is that the financial management of the educational system is not always decentralized. Oftentimes, the role of funding agencies does not go beyond the summit, and is confined to interacting at the ministry level.

To properly implement decentralization, at least three conditions are necessary: first, political will; second, a real, sincere, desire for change; and third, ending corruption in all forms. Aid may be more useful if dispersed at the local (commune) level, rather than through the ministry. Management is difficult at the school level because there is no administration. Perhaps if funding agencies work directly with civil society, the ministry will be frustrated and block everything, and therefore risk a return to the major inconveniences of a centralized system.

Interview Responses

In person: AFD Headquarters, Paris, France

Two staff members from the AFD (2015.07.24)

Skype interview:

Former DANIDA staff member (2015.09.17)

E-mail Questionnaire and Open-Ended Responses:

Former Beninese Government Official (aid recipient)

CASE STUDY 2

Norad (conducted by Cambridge Education Ltd and METCON Consultants)

Joint Evaluation of Nepal's Education for All 2004-2009 Sector Programme.

March 2009

Why selected for in-depth review:

The quality of this evaluation is among the highest. The data collection and analysis processes are thoroughly described, the limitations are explicitly addressed, and the qualitative and quantitative data are well integrated. Approach is comprehensive and analyses are directly linked to data sources.

Why selected for case study:

An evaluation of a suite of programs funded by multiple donor agencies; broadly linked together via Education for All, sets this case study apart from the two other case study evaluations of specific education projects. Contacts at Norad permitted (remote) direct discussions with an aid official who played a leading role in producing the evaluation. Direct discussions with Nepali education officials, teachers, families, and students involved in the evaluation were not possible for a number of reasons (mainly, unreliable Internet connection in Katmandu and elsewhere in Nepal, ongoing challenges caused by the April 2015 earthquake, and political turmoil). Instead, 6 respondents from Nepal (current or former officials from the Ministry of Education) filled out an open-ended survey about their participation in and perceptions of evaluations of aid-funded education activities.

Evaluation approach/method:

This document is a sector wide evaluation of the EFA program in Nepal. The primary objective of the evaluation is to "assess the effectiveness and efficiency of the EFA programme." The evaluation methods include document analysis, descriptive analysis of national and district administrative data, and interviews with students, teachers,

and education officials. Where possible, evaluators took care to ensure that interviews included an equal male/female representation and to individually interview respondents who remained silent in the group interviews.

Major findings of the evaluation:

The findings section focuses on Nepal's overall progress towards expanding education participation, improving achievement, and strengthening institutional capacity. Overall, EFA is considered a success; enrollment has increased and gender and caste/ethnic enrollment disparities have decreased, although quality remains a challenge—in particular among schools serving the poorest and most marginalized communities.

A brief note in the introduction reminds readers that these findings are not causal, that is, these improvements are not necessarily due to EFA programming.

Besides a brief discussion of the status of donor cooperation and the use of external performance audits, there is little mention of the aid community's role in EFA.

Major recommendations:

Policy: develop a cost-sharing mechanisms and seek to better understand what educational costs are borne by families, develop a more complete policy on language use in classrooms, aligned with Nepal's multilingual context, develop improved policy for inclusive education – including non formal and alternative education programmes

Access, equity and quality: simplify scholarship systems while keeping basic education free, target funding to disadvantaged schools through School Improvement Plans, strengthen in-service teacher training, further integrate child-friendliness, gender sensitivity in all aspects of schooling, improve national assessment capacity, develop standards for early childhood education, non-formal education and adult literacy programs, improve the capacity of school management committees and parent teacher associations (involving all members, not just

chairperson), ensure equitable distribution of teachers between schools/districts, strengthen monitoring and evaluation capacity within the ministry – at district and national levels – and in particular in the use of "qualitative information to illuminate observations from quantitative analyses"

Finance, planning and audit: ensure the Government of Nepal maintains its commitment to 20% of national budget to education, with at least 60% going towards EFA goals, incorporate evaluations from the outset of programs – evaluating processes as well as outcomes, and including baseline studies in all EFA programming

Major observations:

On education in poor countries: Broadly, the findings from this evaluation echo many other evaluations – quantity (access) has improved, but quality has not, or it is hard to say whether or not quality has improved because there is no data on student learning. Some specific findings emerge regarding equity and inclusion, however, such as the need to attract more female teachers and teachers from disadvantaged groups, socio-cultural barriers to schooling among disabled children an marginalized castes/ethnic groups,

On evaluating aid-supported education activities: The evaluators dedicate one page to assessing community-managed schools. A few advantages and disadvantages are identified, but not analyzed, and the section concludes that the composition and leadership of the school management committee are the main factors in determining the success of this strategy. A brief review of the recent literature in Nepal suggests that decentralization and school-based management are central components of Nepal's educational development strategy—and both have been met with much criticism, mainly regarding the tendency of these reforms to exacerbate existing inequities and to perpetuate the chronic underfunding of public education (Carney, Bista, & Agergaard, 2007). What explains the apolitical analysis found in this evaluation (as in most) of decentralization and school-based management?

Supplementary information from interview and survey responses:

Interview and survey responses provide a more thorough understanding of the evaluation process, findings, and subsequent use of the evaluation. Several observations stand out:

- **Mismatched timelines**: The evaluation was requested within three months before decisions about the subsequent round of EFA funding had to be made. Not surprisingly, this meant that the evaluation was not finished until *after* the second phase had already started, severely limiting the utility of this particular evaluation. Per interview and survey respondents, this is a problem commonly experienced in evaluations of aid-funded education programs.
- **Mismatched objectives:** "*Evaluate the country's development! Don't evaluate us!*" One challenge to evaluating aid-funded activities has to do with divergent ideas regarding what (and whom) should be evaluated. The notion that evaluations should focus on evaluating the efficiency, relevance, and efficacy of the aid agency itself is at odds with the idea that evaluations should focus on evaluating the overall "state of development" in aid-recipient countries, or aid-recipient government's progress towards established national development goals. The idea of evaluating the role of the aid agency itself in supporting/detracting from a project's success or sustainability is not readily accepted—and is at times resisted–by aid agency staff in aid-recipient countries.
- Despite widespread notion of the importance of evaluations in promoting evidence-informed policies/programs, **few respondents could provide concrete examples of evaluations' use in practice.** Why? Several explanations stand out:
    - Evaluations by themselves are not sufficient. A culture of evaluation use must be institutionalized; there must be a political commitment to using evaluations. Several respondents mentioned that there is no "institutional structure or mechanism exists to ensure findings are used."
    - Evaluations are not always useful:
        - … because they are generally carried out by outsiders; with the government playing a role as "information provider," or "drafting the

terms of reference," or "providing comments once the evaluation has been drafted" (survey respondents, all Nepali education officials). This could limit the extent to which the government is inclined to make use of findings, especially "if the context and culture have not been duly considered while designing the evaluation criteria, the findings may deviate with that of the intended purpose of the program project" (survey respondent, Nepali education official).

        - … because they provide only very general or theoretical findings, based on averages or abstract statements, which lead to recommendations that are extremely difficult (or take too many resources) to implement.

- **Several respondents respondent did provide some examples of policies that have been created directly in response to evaluation findings**, these are:
    - The incorporation of early grade reading strategies
    - Revisions to teacher training programs
    - Continuous student assessment system
    - Respondents note that funding agencies use evaluations to determine which programs to support and national and district education officials use evaluations to improve "gaps in program implementation." However, one respondent noted that, "teachers tried to use the findings of the evaluation but in some cases had reservations regarding the findings." Of course, evaluations funded by aid agencies are most often conducted in order to determine future funding decisions, rather than provide concrete recommendations or implementation guidelines to teachers.

Interview and Survey Responses

Skype interview:

Former Norad Evaluation Specialist (2015.10.16)

Survey responses:

Former Secretary of the Government of Nepal (Planning Division, Department of Education)

Director, Human Resource Development Division (National Centre for Educational Development, Government of Nepal)

Government Official; responsible for planning and coordination of the School Sector Reform Program (Department of Education, Ministry of Education, Government of Nepal)

Government official; responsible for administration of basic education, higher education, and technical education (Ministry of Education, Government of Nepal).

CASE STUDY 3

Sida (conducted by Indevelop: Bernt Andersson; Edephonce Ngemera Nfuka; Suleman Sumra; Paula Uimonen; Adam Pain).

Evaluation of Implementation of ICT in Teachers' CollegesProject in Tanzania. Final Report

May 2014. Sida Decentralised Evaluation 2014:26

Why selected for in-depth review?
Systematic evaluation of a significant project (size; ICT prominence in education ministry's overall strategy) in Tanzania, supported by the Swedish International Development Cooperation Agency. Attentive to objectives and methodology. Interviews with several sets of participants in the funded activities.

Why selected for case study?
A clear example of a specified set of activities with direct funding agency support. The evaluation documents those activities carefully and employs a broad approach, including site visits and participant interviews. Visits to Stockholm and Dar es Salaam permitted direct discussions with funding agency staff, evaluators, and people in Tanzania involved in or directly familiar with the funded activities.

Activities evaluated
The project, funded by Sida (USD 3,733,000) and implemented by the

Ministry of Education and Vocational Training from 2005-2008, provided computers and related equipment, training, and internet connections to 34 government Teachers' Colleges. The primary objective was to improve teacher education, specifically to enable all new teachers to be computer-literate and to be able to use information and communications technology in their teaching.

Evaluation approach/method

Commissioned by Sida and implemented by a team assembled by InDevelop, the evaluation reviewed documents and government education reports and statistics, surveyed tutors in 12 Teachers' Colleges, and undertook interviews in Dar es Salaam and at 13 other sites. The report is largely descriptive, with analysis developed through the presentation and interpretation of the findings.

Major findings:

Most of the basic objectives were achieved: computers were delivered and installed in Teachers' Colleges; most of the tutors were trained; internet access was provided. Some tutors were not trained, and the intended ratio of functional computers per student-teachers remained below the intended target. Project management was generally efficient. The project fit well within Tanzania's national aspirations for ICT in education. Some of the teachers with newly developed ICT competence were assigned to schools with no computers. Notwithstanding the project input, the technical challenges—maintaining, repairing, and replacing computers; assuring sufficient internet access, increasing the hardware to reach more students—remain substantial, requiring substantial additional foreign assistance. Because neither the initial project nor the evaluation followed the teachers to their assigned schools, there is no evidence that the project had a significant impact on teaching and learning at secondary or primary level. Since the project had no explicit gender component, the evaluation did not address either gender inequality in the use of ICT in Teachers' Colleges or the impact of the project on gender inequality.

Major recommendations:

The Ministry of Education and Vocational Training should assign higher priority to developing and extending the role of ICT in education at multiple levels. Significant financial and human resources will be required.

Equipment provided by the project has reached its intended lifespan and must be replaced.

MOEVT should shift internet access from satellite to fiber cable to expand access and reduce its cost.

The teaching on ICT should be re-focused from the technical and theoretical dimensions of ICT systems to the use of ICT in teaching.

Sida should continue and extend its support for ICT in education.

Increased use of ICT offers a strategy for addressing the severe shortage of teachers in mathematics and science. ICT can enable experienced teachers to teach in distant schools. Lessons can be recorded and distributed on DVDs.

Major observations:

On the ICT support project. For the most part, the project addressed the first level issues (hardware provided, training sessions; access to computers and internet), with little explicit attention to teaching and learning through the use of ICT. Surprisingly for a Swedish initiative, the project did not address inequality, especially gender inequality. The project did not explicitly address sequels and sustainability, which are always important and perhaps even more important where the activities depend on equipment that is expensive and has a limited functional lifespan. While the project fit within Tanzania's national ICT education policies and plans, it apparently included no explicit coordination with institutions and organizations in Tanzania other than MOEVT or with other funding and technical assistance agencies then providing or planning to provide ICT support to Tanzania.

On the evaluation. In many respects the evaluation was systematic, thorough, and thoughtfully implemented. Site visits to 13 Teachers' Colleges and surveys of program participants went well beyond the more common review of documents and counts of equipment and participants. Yet, the evaluation understood its task relatively narrowly and focused its primary attention on the first level issues the project addressed. The evaluation noted but did not explore *why* equipment maintenance and replacement remain a major obstacle. The evaluation did not recognize a tension between its report that the project was consistent with national ICT and education policy, had MOEVT commitment, and was managed efficiently on the one hand and on the other MOEVT's inability to resolve either the technical problems or the continued use of ICT in teacher education. Some of the evaluation's recommendations are more wish lists than reasoned

analyses of what is needed and what is possible (for example, vast increase in the number of computers and computer-equipped classrooms, all with reliable broadband internet access). Other recommendations reflect unfamiliarity with relevant education research or uncritical advocacy of strategies that have proved ineffective in other settings (for example, substituting DVDs for science and mathematics teachers).

On evaluation aid-supported education activities. Though extensive and likely costly, the evaluation did not address the larger education issues—teaching, learning, education as an integrated system. Even gender inequality, long a Swedish concern, received no attention in the evaluation. Nor did the evaluation address the structural and institutional context: an external funding agency provided hardware and training, and then moved on, with at best limited attention to national ownership and integration into sustainable education development. Especially problematic in that regard is that there are now several decades of experience with externally-provided computers and other hardware and a substantial evaluation and research literature that apparently informed neither the evaluation nor its recommendations. Like most others, this evaluation did not address the aid relationship and its consequences for education improvement.

Supplementary information from interviews in Sweden and Tanzania

To understand more fully the development, implementation, and sequels of this project, Samoff undertook interviews in Sweden and Tanzania. The major concerns were to explore the content and context of the project and especially to learn more about the receipt and use of the evaluation. For whom was the evaluation useful? From those interviews, along with extended document review, several observations stand out.

- Even the most dedicated and sensitive aid agencies operate on a cycle that is much shorter than education innovation and reform require. While there is some continuity across projects, for the most part the aid agency develop and support and activity, monitor its implementation and evaluate it, and then move on. Regularly, the staff involved in developing the project have assumed other responsibilities by the time the project reaches fruition. That is compounded when education

ministry staff also move to new posts before project completion.

- Aid agencies and evaluators are inclined to assess consistency with national policies and plans by comparing documents. Are the objectives in the project document consistent with statements in a national policy or strategy document? Everyone understands, however, that documents are formal statements and may not reflect policies or priorities in practice. A national document, for example, may affirm that education is free and compulsory. Policy and practice, however, may effectively exclude some learners because their families cannot meet the costs, or because they live in remote areas or regularly migrate, or because schools cannot accommodate their learning disabilities, or for other reasons. Consistency with national policy-in-practice is too important to assessed entirely or largely through comparisons of documents. In this case, evaluations were insufficiently attentive to relevant policies-in-practice—measured by allocations and specific actions. Ironically, initial Swedish skepticism about the importance of ICT in education in Tanzania was a more accurate assessment of the contemporary situation than the vision and expectations embedded in the project documents.

- Both aid agency staff and their evaluators have little time for—in practice, that means assign low priority to—reviewing relevant previous experiences and research. For this project, that is especially striking, since there are three decades of evaluations of and research on projects designed to deliver computer technology and training to teachers and schools in Africa. Notwithstanding the regularly reiterated commitment to learning from experience, most often the aid process has little room for that learning.

- The oft-repeated concerns in aid funding, for example national ownership and sustainability, are generally not explicit concerns of evaluators and thus generally not systematically measured or assessed.

- Where there is generation of new knowledge and learning from experience, most often that occurs among those involved in the education activities, not in the aid agency or its evaluations. One example stands out here. While the

evaluators did not address gender inequality, Tanzanian educators highlighted the failure of this project to address or remedy that.

- Currently, nearly all aid-funded activities require evaluations. Most often, those evaluators, even where they involve significant local (Tanzanian) participation, as this one did, are organized and presented in ways that may meet funding agency needs but that do not serve well those directly involved in the aid-funded activities. When asked, local educators are clear that they see the evaluations as an aid agency process. In this case, while the evaluation was submitted to the education ministry for comment and then in final form, relevant senior ministry staff were unfamiliar with it and were skeptical that they could locate a copy.
- In both Sweden and Tanzania, hardly anyone knew about the evaluation, its findings, and its recommendations. No one could accurately identify a policy, or program, or allocation, or education activity that was informed, influenced, or shaped by the evaluation.
- That suggests the importance of re-focusing the evaluation process. Evaluations that are primarily intended to meet the funding agency's need to monitor the project (were the specified activities undertaken? were the funds spent as intended? was the target population reached) can be far less elaborate and less costly. Evaluations intended to assist those responsible for the education innovation or reform will need to involve them directly, from conception through implementation and analysis. It may be advantageous to shift the balance from summative to formative evaluations.

Interviews and discussions

Sweden
Stellan Arvidsson Hyving (2015.06.11)
Education, Swedish International Development Cooperation Agency
Evaluation Synthesis Reference Group

Mats Borgenvall (2015.06.09)
Evaluations, Ministry of Foreign Affairs, Sweden

Hallgerd Dyrssen (2015.06.07)
Former Head, Public Administration and Management Division, Swedish International Development Cooperation Agency

Paula Engwall (2015.09.11)
Principal International Secretary/Head of International Unit, Teachers Union

Kim Forss (2015.06.10)
Evaluator
Chair, Evaluation Synthesis Reference Group

Sarah Gharbi (2015.06.09)
Evaluator, Indevelop

Ulrika Hertel (2015.06.10)
Senior Programme Specialist, Swedish International Development Cooperation Agency

Emma Holmberg (2015.06.11)
International Department, Save the Children (Rädda Barnen)

Birgitta Jansson (2015.06.10)
Senior Policy Specialist, Afghanistan Unit, Swedish International Development Cooperation Agency

Agneta Lind (2015.06.07, 10)
Former Head, Education, Swedish International Development Cooperation Agency

Susanne Mattsson (2015.06.08)
Unit for Monitoring and Evaluation, Swedish International Development Cooperation Agency

Christine McNab (2015.06.07)
Former head, development cooperation, Embassy of Sweden, Dar es Salaam

Bertil Oskarsson (2015.06.09)
Education, Indevelop

Jessica Rothman (2015.06.09)
Project Manager/Advisor, Indevelop

Magnus Saemundsson (2015.06.10)
Senior Education Specialist, Cambodia coordinator, Swedish International Development Cooperation Agency

Eva Tobisson (2015.06.09)
Evaluations, Ministry of Foreign Affairs, Sweden

Tanzania
Dr. Elia Kibga (2015.09.08)
Director of Research Information and Publications Department, Ministry of Education and Vocational Training, Tanzania Institute of Education

Sara Kironde (2015.09.09)
Department of Teacher Education, Ministry of Education and Vocational Training

Helen A. Lihawa (2015.09.09)
Acting Director, Department of Teacher Education, Ministry of Education and Vocational Training

Samwel Makunde (brief telephone discussion, 2015.09.08)
Assistant Director, Department of Teacher Education, Ministry of Education and Vocational Training

Stella Mayenje (2015.09.07)
Education and Global Partnership for Education, Embassy of Sweden, Dar es Salaam

Omar Mzee (2015.09.06)
Managing Director, Studiacademy
Formerly, Education, Embassy of Sweden, Dar es Salaam

Helena Reuterswärd (2015.09.07)
Education Adviser, Embassy of Sweden, Dar es Salaam

Joseph Rugumyamheto (2015.09.08)
Former director of Tanzania's civil service

Dr. Frank Tilya (2015.09.06)
University of Dodoma

Pius Wanzala (2015.09.07)
Civil Society Organizations and Education, Embassy of Sweden, Dar es Salaam

G. Terms of reference

Terms of reference for proposal: Synthesis evaluation of education aid

The Expert group for aid studies (EBA) is a government committee with the mandate to evaluate and analyze Swedish international development assistance. EBA commissions studies and arrange seminars on issues and thematic areas of relevance for Swedish development aid.

EBA has decided to commission a synthesis evaluation of aid to the education sector. Hereby we invite researchers and evaluators to submit proposals for such an evaluation.

Synthesis evaluation of aid to the education sector

Education is a human right. The millennium development goals set out to ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling. The follow-up of the MDGs show encouraging results. More children than ever before are attending primary school. However, in a 2011 follow-up of MDG progress, 57 million children of primary school age where still out of school.

Education is a priority in Swedish development assistance. In the recently launched Swedish aid platform, education is one of the six sub-targets. However over time education has been given less priority in Swedish development assistance, despite the great remaining needs in low-income countries. Education has not been a focus area for Swedish evaluations and analyses of development assistance. Other donor countries and organizations have, however, done more and lessons and guidance for Swedish education aid may therefore be drawn from evaluations carried out by others.

The synthesis evaluation is expected to compile and analyze findings and conclusions from high quality evaluations and syntheses of development aid to primary and secondary education in various contexts. Two overall questions should guide the synthesis:

1. What type of programs and/or aid modalities for primary and secondary education have proven to be effective or not effective? Where, when, how and why have these programs been effective?

2. Are there other conclusions, experiences and best practices are described in these evaluations?

The main objective of the synthesis evaluation is to provide grounded and elaborated responses to the questions above and highlight potential lessons for Swedish development assistance in the education sector. The conclusions from the synthesis should be linked to the Swedish portfolio of development assistance for education.
By education we are primarily referring to primary and secondary school, not higher learning (college and university) nor job training, internships, apprenticeship etc. The study should be limited to educational programs financed by development aid.
A secondary objective of the study is to contribute to developing a model for EBA´s synthesis evaluations.

Evaluation implementation and methods

A detailed analytical framework for the evaluation should be attached to the proposal. It is up to the evaluator to choose study design and methods for the synthesis, but the choices should be justified.

At an early stage of the evaluation a database, or comprehensive list, covering as many evaluations as possible should be developed. From this list, a selection of evaluations to be included in the synthesis should be made covering a specific time period. The proposal should suggest criteria for selection and describe them in the proposal. Quality should be an important selection criteria and as many high quality evaluations as possible should be covered by the synthesis. Selection, limitations and the consequences thereof should be described as thoroughly as possible in the proposal.
The evaluator should differentiate between types of aid to the education sector. Swedish aid to the education sector is distributed through various channels. Bilateral support (about 650 million SEK for 2013), support through multilateral organizations and through NGOs. It is important that the proposal manages to capture all these types of aid to the education sector.

The synthesis evaluation should be focused on evaluations and not synthesize research more generally. Both evaluations financed by Sweden and evaluations conducted by other donors or by recipients should be included in the study. If there are robust evaluations of education aid financed by Sweden, these should be accounted for

separately in the evaluation. The conclusions of the synthesis should be related to the Swedish aid portfolio and Swedish policies for education aid.

For the second objective, the evaluator should review the selected evaluations with regard to how they have been designed and the methods applied. Based on this review, the evaluation team should propose a model for how EBA could conduct synthesis evaluations in the future. The foreseen model should be suitable for evaluations with a fairly limited budget and short- medium-term time plan.

The conclusions should be presented in a report written in English.

Administration, budget and timetable

The project proposal should be no longer than 15 pages (excluding annexes, CVs etc) and should in addition to the proposal and the team presentation include a budget and detailed preliminary timetable. The maximum cost is 500.000 SEK (approximately 65.000 USD). The timetable should include details regarding time used for each member of the evaluation team.

The budget should accommodate 2-4 reference group meetings with the reference group the EBA attaches to the study (in dialogue with the other). If the team resides outside Sweden the meetings could be conducted via video link. The following preliminary time plan should be considered:

Deadline expression of interest	8 December 2014
Selection of 3-5 authors who are invited to submit a full proposal	18 December 2014
Deadline for proposal	25 January 2015
Evaluation of proposals	25-30 January 2015
Proposal selected and decided by the EBA	10 February 2015
Contract signed	20 February 2015
Delivery of inception report	31 March 2015
Reference group meetings	March – December 2015

Draft final report	30 October 2015
Final report delivered	15 December 2015

Questions can be answered by Jesper Sundewall (jesper.sundewall@gov.se, +46 70 245 2889)

The proposal should be sent to jesper.sundewall@gov.se

The following criteria will be used in the screening of proposals

1. Quality of proposal in terms of plan for implementation, evaluation design and methods (weight 60%)

2. Experiences and qualifications of team members in the areas of education, evaluation and development assistance. (weight 30%)

3. Cost (weight 10%)

About the Expert Group for Aid Studies (EBA)

The EBA is tasked with commissioning, compiling, conducting and communicating analyses, studies and evaluations of Swedish development assistance, in particular its execution, results and efficiency. EBA strives to use existing research and knowledge on international development assistance and contribute to such knowledge being put to effective use in development policy. EBA's studies focus mainly on overall issues in Swedish development assistance.

The EBA works with "dual independence". This means that the EBA independently defines what issues to explore and which studies to commission. The content and the conclusion of each report is however the responsibility of each author.
The expert group consists of: Lars Heikensten, chairperson, Kim Forss, Maria Gustavson, Torgny Holmgren, Eva Lithman, Anna Nilsdotter, Hans Rosling, Julia Schalk, Jakob Svensson and Johanna Stålö.

8. Previous EBA-reports

2016:02 *What Education Policies and Programmes Affect Learning and Time in School in Developing Countries?* Amy Damon, Paul Glewwe, Suzanne Wisniewski, Bixuan Sun

2016:01 *Support to regional cooperation and integration in Africa - what works and why?* Fredrik Söderbaum and Therese Brolin

2015:09 *In search of double dividens from climate change intervention evidence from forest conservation and household energy transition.* Gunnar Köhlin, Subhrendu K Pattanyak, Erin Sills, Eskil Mattson, Madelene Ostwald, Ariana Salas, Daniel Ternald

2015:08 *Business and Human Rights in development cooperation – has Sweden incorporated the UN guiding principles?* Rasmus Kløcker Larsen and Sandra Atler

2015:07, *Making development work: the quality of government approach,* Bo Rothstein and Marcus Tannenberg

2015:06, *Now open for business: joint development initiatives between the private and public sectors in development cooperation,* Sara Johansson de Silva, Ari Kokko and Hanna Norberg

2015:05, *Has Sweden injected realism into public financial management reforms in partner countries?* Matt Andrews

2015:04, *Youth, entrepreneurship and development*, Kjetil Bjorvatn

2015:03, *Concentration difficulties? An analysis of Swedish aid proliferation*, Rune Jansen Hagen

2015:02, *Utvärdering av svenskt bistånd – en kartläggning*, Expertgruppen för biståndsanalys

2015:01, *Rethinking Civil Society and Support for Democracy*, Richard Youngs

2014:05, *Svenskt statligt internationellt bistånd i Sverige: en översikt*, Expertgruppen för biståndsanalys

2014:04, *The African Development Bank: ready to face the challenges of a changing Africa?* Christopher Humphrey

2014:03, *International party assistance – what do we know about the effects?* Lars Svåsand

2014:02, *Sweden´s development assistance for health – policy options to support the global health 2035 goals*, Gavin Yamey, Helen Saxenian, Robert Hecht, Jesper Sundewall and Dean Jamison

2014:01, *Randomized controlled trials: strengths, weaknesses and policy relevance*, Anders Olofsgård.