



## SYSTEMATIC REVIEW

# Evaluating integrated development: are we asking the right questions? A systematic review [version 1; peer review: 3 approved with reservations]

Tessa W Ahner-McHaffie<sup>1</sup>, Greg Guest<sup>2</sup>, Tricia Petruney<sup>2</sup>, Alexandra Eterno<sup>1</sup>, Brian Dooley<sup>1</sup>

<sup>1</sup>FHI 360, 1825 Connecticut Avenue, NW; Washington, DC, USA

<sup>2</sup>FHI 360, 359 Blackwell St Suite 200; Durham, NC, USA

**V1** First published: 06 Nov 2017, 1:6  
<https://doi.org/10.12688/gatesopenres.12755.1>  
 Latest published: 29 May 2018, 1:6  
<https://doi.org/10.12688/gatesopenres.12755.2>

## Abstract

**Background:** Emerging global transformations - including a new Sustainable Development Agenda - are revealing increasingly interrelated goals and challenges, poised to be addressed by similarly integrated, multi-faceted solutions. Research to date has focused on determining the effectiveness of these approaches, yet a key question remains: are synergistic effects produced by integrating two or more sectors? We systematically reviewed impact evaluations on integrated development interventions to assess whether synergistic, amplified impacts are being measured and evaluated.

**Methods:** The International Initiative for Impact Evaluation's (3ie) Impact Evaluation Repository comprised our sampling frame (n = 4,339). Following PRISMA guidelines, we employed a three-stage screening and review process.

**Results:** We identified 601 journal articles that evaluated integrated interventions. Seventy percent used a randomized design to assess impact with regard to whether the intervention achieved its desired outcomes. Only 26 of these evaluations, however, used a full factorial design, the only design capable of statistically detecting synergistic effects produced by integrating sectors. Of those, seven showed synergistic effects.

**Conclusions:** To date, evaluations of integrated development approaches have demonstrated positive impacts in numerous contexts, but gaps remain with regard to documenting whether integrated programming produces synergistic, amplified outcomes. Research on these program models needs to extend beyond impact only, and more explicitly examine and measure the synergies and efficiencies associated with linking two or more sectors. Doing so will be critical for identifying effective integrated development strategies that will help achieve the multi-sector SDG agenda.

## Open Peer Review

Approval Status

	1	2	3
<b>version 2</b>			
(revision)			
29 May 2018	<a href="#">view</a>	<a href="#">view</a>	<a href="#">view</a>
<b>version 1</b>			
06 Nov 2017	<a href="#">view</a>	<a href="#">view</a>	<a href="#">view</a>

1. **Erica Di Ruggiero**, University of Toronto, Toronto, Canada

**Leah Shipton** , University of Toronto, Toronto, Canada

**Donald C. Cole** , University of Toronto, Toronto, Canada

2. **Reshma Trasi**, Pathfinder International, Washington, USA

3. **Thomas de Hoop** , American Institutes for Research, Washington, USA

Any reports and responses or comments on the article can be found at the end of the article.

## Keywords

integrated, development, multi-disciplinary, multi-sector, evaluation, synergy, interaction effects, SDGs

**Corresponding author:** Tessa W Ahner-McHaffie ([tahnermchaffie@gmail.com](mailto:tahnermchaffie@gmail.com))

**Author roles:** **Ahner-McHaffie TW:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Guest G:** Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Petrune T:** Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Eterno A:** Data Curation; **Dooley B:** Data Curation

**Competing interests:** No competing interests were disclosed.

**Grant information:** Bill and Melinda Gates Foundation [OPP1130126] and the FHI Foundation.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2017 Ahner-McHaffie TW *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Ahner-McHaffie TW, Guest G, Petrune T *et al.* **Evaluating integrated development: are we asking the right questions? A systematic review [version 1; peer review: 3 approved with reservations]** Gates Open Research 2017, 1:6 <https://doi.org/10.12688/gatesopenres.12755.1>

**First published:** 06 Nov 2017, 1:6 <https://doi.org/10.12688/gatesopenres.12755.1>

## Introduction

### Rationale

Twenty-first century global trends such as rapid urbanization and dramatic climate change are forcing the international community to rethink solutions to challenges that are increasingly multi-faceted and interrelated. Indeed, the Sustainable Development Goals (SDGs) – an ambitious framework of 17 goals to end extreme poverty, fight inequality and injustice, and reverse climate change over the next 15 years – emphasize the integration of previously distinct development aims. The agenda states that “[t]he goals and targets we have decided on are integrated and indivisible and balance the three crucial dimensions of sustainable development: the economic, social and environmental” (United Nations, 2015). This evolution in thinking indicates a firm shift away from narrowly isolated sectors of development toward what the authors refer to as “win-win cooperation.” A recent analysis of how each of the 169 SDG targets is related to others reveals a web of closely interrelated objectives, yet also points out that any policy and program integration founded on these underlying linkages would need to rest on evidence with regard to their means of implementation (Le Blanc, 2015).

Thus, decisions about when and how to most effectively implement the integrated, multi-disciplinary SDG agenda need to be driven by evidence, rather than by assumptions about the amplified results of ‘doing more together’. Importantly, a large volume of research carried out to assess various types of integrated programs suggests that in many cases these approaches are successful in achieving positive impacts. But are program evaluators addressing a critical question: is  $1 + 1 > 2$ ? In other words, are these interventions generating amplified impacts that go beyond the sum of single sector interventions? To effectively advocate for integrated, multi-disciplinary approaches to development, it behooves us to understand under which circumstances integrating two or more development sectors enhances impacts in amplified or synergistic ways. In this paper we present the results of a systematic review designed to identify if and how synergies and interaction effects between sectors in integrated interventions are being measured.

### Objectives

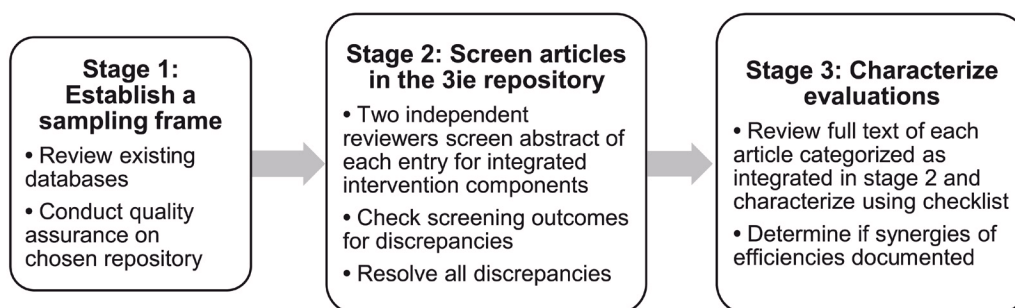
The objectives of this systematic review were threefold: to summarize how integrated development programs are being evaluated; to assess whether these evaluations seek to statistically measure synergistic effects or efficiencies associated with integrated interventions; and, for those that do, to document if synergies are detected. Secondary objectives include documenting other characteristics associated with study design, such as: inclusion of cost analyses or qualitative analyses, types of sectors involved, and regions where the intervention took place.

### Methods

Our review consisted of a three-stage process (Figure 1), based on the PRISMA guidelines (Moher *et al.*, 2009), as well as recommendations from Waddington *et al.* (2012). We first established a sampling frame from which to screen and review published articles. Given our objectives, we required a large, relatively exhaustive development database, or combination of databases, that included a broad range of evaluations not limited to any particular development sector. The first stage in our process, therefore, entailed establishing a sampling frame from which to identify and review the highest number of relevant publications. Once the sampling frame was established, we screened each article’s abstract and identified evaluations of development programs that we defined as “integrated development.” We then reviewed the full text of each article in the integrated development subset and documented characteristics essential to the objectives of our review.

#### Stage 1: Establish a sampling frame

The universe of human development literature and evidence is extensive. Published articles from this broad field are scattered throughout numerous databases, span a multitude of sectors, and use a diverse range of keywords and terminology. We explored many of these specialized databases and possible combinations of databases from which we could specifically identify the most evaluations of programs that integrated two or more traditional development sectors (described below). Fortunately, we found a



**Figure 1.** Review process for systematic review on integrated, multi-sector programs.

high-quality database that was not specific to any particular sector that would serve well as a single sampling frame for our review.

The International Initiative for Impact Evaluation (3ie) Impact Evaluation Repository is an index of all published impact evaluations of development interventions. To be included in the repository, an impact evaluation must be published (as a journal article, book chapter, report, or working paper), take place in a developing country, examine the effectiveness of a specific development intervention, and use a specifically defined experimental or quasi-experimental estimation strategy. The 3ie review process has no restriction on publication date; however, the systematic review upon which it is based was completed in July, 2016.

During the creation of the repository, 3ie systematically searched more than 45 databases, search engines, journal collections, and websites with an aim to identify all published development impact evaluations (Figure 2) (Mishra & Cameron, 2014; Jorge Miranda, 2017, personal communication). At the time of our analysis (September 8, 2016), 3ie had reviewed more than 140,000 potential studies, rendering an index of 4,339 eligible studies (Mishra & Cameron, 2014; Jorge Miranda, 2017, personal communication). The repository, including a full description of its inclusion criteria and review methodology, is available [here](#).

As part of our due diligence, we sought to confirm that 3ie's repository was a thorough and sufficient sole-sampling source. A library information science specialist audited the methodology 3ie staff used to create the repository. Her objective was to assess whether the searches used were both sensitive (i.e., broad) and specific (i.e., focused) enough to ensure that the vast majority of relevant and eligible impact evaluations were included in the final repository. She reviewed the databases that were used and how they were searched with regard to subject scope, time frame limits, and geographic coverage. She found that some lesser-known and regional databases were excluded from the 3ie repository; however, she determined that these smaller databases would not likely have added a notable number of new references. She concluded that the overall methodology design was strong and its implementation consistent. We therefore feel confident that using the 3ie repository as our sampling frame provided us with a sufficient index of development evaluations.

## Stage 2: Screen articles in the repository for integrated development approaches

The purpose of this stage of the review was to identify all of the publications in the 3ie repository that evaluated integrated programs. No terms to denote integrated development have universal agreement. The concept of integrated or multi-sector development in published papers is described by many different terms (e.g., cross-sector, linked, combined, blended). Moreover,

authors rarely self-identify their interventions in this way within an article, let alone an abstract. We could not, therefore, rely on key search terms to identify evaluations of programs that were integrated in nature. Instead, we manually reviewed the abstracts of every study in the 3ie repository (as of September 8, 2016) against our organization's working definition of integrated development:

*"Integrated development approaches intentionally link the design and delivery of programs across more than one core sector."*

Note that our definition of integrated development encompasses studies that would be classified as "multi-sector" or "multi-disciplinary" by others. More precisely, our definition focuses on the integrated nature of the intervention itself and excludes programs that:

- Only integrate different subsectors of a core sector (e.g., health programs that link family planning and HIV/AIDS); or
- Measure outcomes in multiple sectors but do not include multi-sector intervention components (e.g., education programs that measure both education and nutrition outcomes but only deliver education services).

There are no universal or definitive lists of development 'sectors'. Global bodies and implementing organizations characterize thematic areas in fluid ways, at times bundling some fields (e.g., health and nutrition) and at others ensuring they are distinctly separate. For this review we used the following core sector categories and illustrative interventions. We used these sector categories to classify interventions as well as outcome measures. These categories were used as general guiding parameters rather than strictly exclusive definitions:

- Agriculture and food security (e.g., farming, food supply chains, famine prevention);
- Economic development (e.g., income, livelihood, cash transfers, microfinance);
- Education (e.g., early education, primary/secondary/tertiary school);
- Environment (e.g., environmental/land management, conservation, climate change);
- Governance (e.g., peace building, conflict management, election monitoring, democracy);
- Health (e.g., HIV, tuberculosis, maternal and child health, sexual and reproductive health, non-communicable disease, malaria, immunization/vaccine);

- Nutrition (e.g., micronutrients, food fortification, malnutrition, feeding programs, diet diversification); and
- Water, sanitation, and hygiene (e.g., water quality, management, supply).

All of the interventions in the studies reviewed fell within these sector categories. We added an “other” category to describe outcomes measured, to capture more amorphous, non-sector specific measures, such as “child labor.” For our review, cross-cutting topics such as gender, youth, civil society, and technology were considered aspects of, and relevant to, the interventions and outcomes in each sector, but not sectors in and of themselves. During the review we had also initially included ‘humanitarian’ as a sector. With further discussion and analysis, it was clear that this sampling frame was not inclusive of the humanitarian sector, nor is humanitarian work represented at the same level of the conventional development sectors included above. Therefore, the final analysis was completed without the humanitarian sector category in either intervention sector or outcome sector measured.

To enhance reliability, two individuals independently reviewed all of the abstracts and identified the sectors represented in the interventions being evaluated. If more than one sector was identified, the study was categorized as “yes” for integrated development; all other studies were marked as “no”. These two reviewers met at predetermined intervals to compare their results, and had an average of 89% agreement.

All discrepancies in coding were resolved at each comparison point after the team discussed the interpretation of the integrated development definition (resulting in 100% agreement). In the few cases in which the two reviewers could not agree on a study’s categorization, a third party reviewed the abstract and made the final decision. Any study that members of the review team both categorized as integrated development was included in the second round of review. For cases in which the abstract alone did not contain enough information to make a determination, studies were advanced to the next round of review so that a final determination could be made during review of the full text.

Importantly, although the repository includes impact evaluations published in any form, our inclusion criteria for this review required a study to be published in a scientific journal. Therefore, only those publications moved on to Stage 3.

### Stage 3: Characterize evaluations of interventions identified as integrated

Full text articles of the subset of studies on integrated programs that were published in scientific journals were reviewed by two individuals. Each article was compared against a checklist, to ascertain the study’s scope and methodology (our checklist is presented with the corresponding results in the next section).

In particular, we noted the number of control, single-sector treatment, and integrated sector treatment arms in each evaluation. We further identified those evaluations which employed either a partial factorial or full factorial experimental design. For the purposes of our review, partial factorial designs included at least one single-sector arm (but not all single-sector arms), at least one integrated arm, and at least one control (no intervention) arm. Full factorial designs included all possible single-sector arms, at least one integrated arm, and at least one control arm. Factorial designs are exceptionally rigorous and permit evaluators to determine the effects of multiple interventions on an outcome. Since they include all possible combinations of intervention arms, full factorial designs are able to reveal differential effects of single-sector and multi-sector interventions and measure potential synergistic effects associated with integrated approaches.

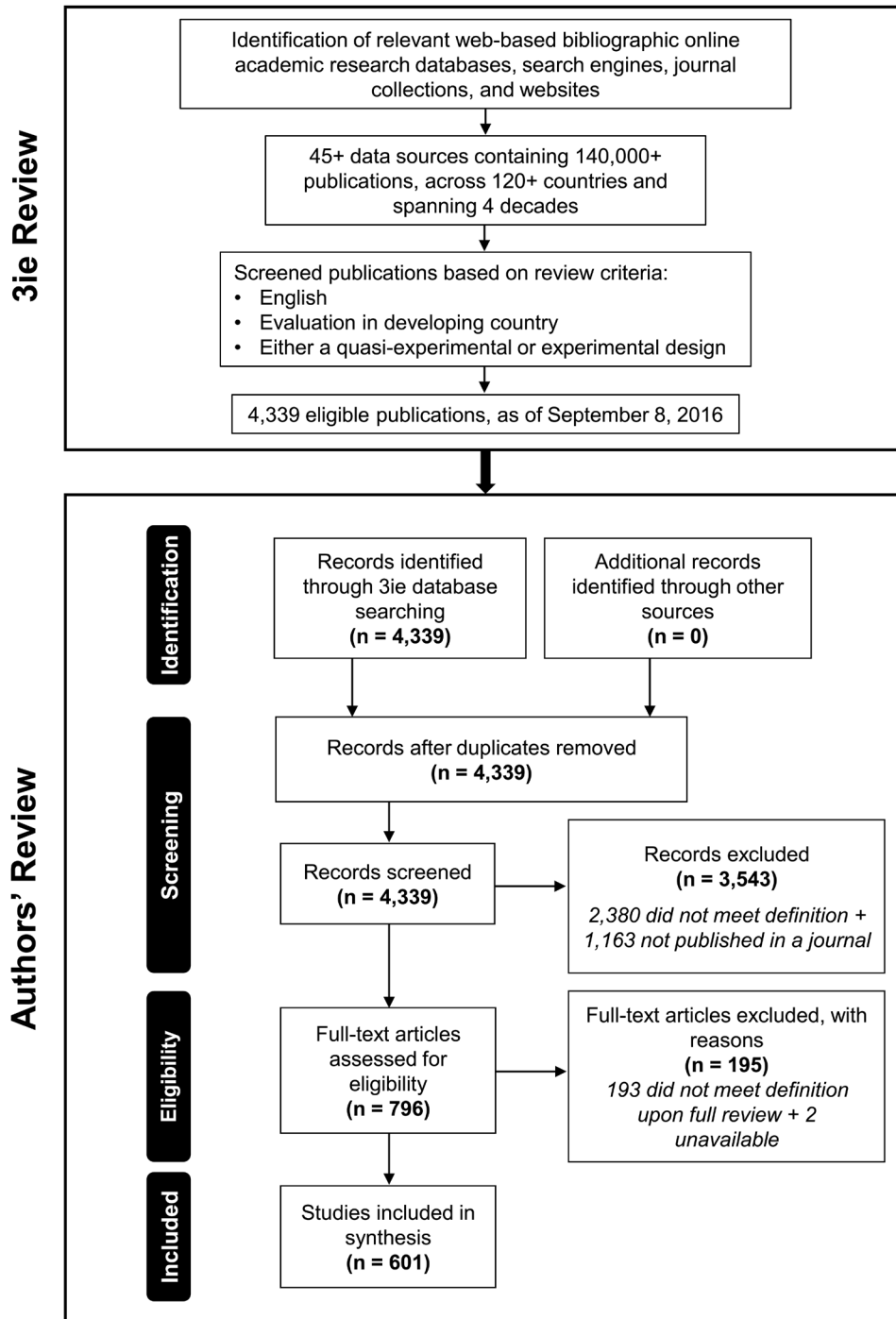
Therefore, we specifically reviewed each full factorial evaluation to determine if the authors measured or detected synergy associated with the integrated study arm. For our review, we defined synergy as a statistically significant ( $p < 0.05$ ) interaction effect between two or more intervention sectors, or instances in which the effect size of the integrated arm of a program was greater than the sum of the effect sizes among the single-sector arms.

Given the extreme heterogeneity of the types of programs evaluated and outcomes assessed, we did not seek to collectively synthesize their substantive findings. Instead, the primary objective was to determine if and how impact evaluations of integrated programs are designed to measure or systematically document the synergy and efficiency assumed in multi-sector development.

## Results

We reviewed 4,339 abstracts, comprising the entire 3ie repository as of September 8, 2016. After a two-step screening process, 601 articles were included in our final dataset for characterization ([Figure 2](#), [Supplementary File 2](#)). From the initial set of 4,339 articles from the 3ie repository, 3,543 were excluded (2,380 did not meet the definition of integrated and 1,163 were not published in a scientific journal). The full text of the remaining 796 were assessed for eligibility. One hundred and ninety-five were excluded with full text review (193 did not meet the definition of integrated and two were not available to reviewers). This left 601 studies included in the analysis. The list of articles is included here as [Supplementary File 3](#), and each article may also be found in a [searchable online database](#).

Most articles (84%) did not identify the interventions being evaluated as “integrated”, or any other related term ([Table 1](#)).



**Figure 2.** PRISMA flow diagram for systematic review with 3ie impact evaluation repository review process.



**Table 1. Summary of integrated development impact evaluation characteristics.**

Characteristic (N = 601)	Frequency	
Self-identified term for the nature of intervention being evaluated (by study author, in the title or abstract)	Integrated	44 (7%)
	Combined	25 (4%)
	Multi-component	9 (1%)
	Multi-faceted	2 (<1%)
	Other	19 (3%)
	Did not self-identify	507 (84%)
General study design	Experimental	419 (70%)
	Quasi-experimental	174 (29%)
	Both	8 (1%)
Study arm combinations included in design	Integrated arm(s) + control arm	366 (61%)
	Integrated arms(s) only	68 (11%)
	Single-sector arm(s) + Integrated arm(s)	132 (22%)
	Full factorial	26 (4%)
	Other	9 (1%)
Partial factorial design (Randomized study that includes at least 1 single-sector arm — but not ALL single-sector arms — and at least 1 integrated arm and at least 1 control arm)	12 (2%)	
Full factorial design (Randomized study that include ALL single-sector arms, at least 1 integrated arm, and at least 1 control arm)	26 (4%)	
Qualitative component included in evaluation	60 (10%)	
Sectors included in study <u>intervention</u>	Agriculture & food security	65 (11%)
	Economic development	238 (40%)
	Education	433 (72%)
	Environment	25 (4%)
	Governance	12 (2%)
	Health	456 (76%)
	Nutrition	266 (44%)
	Water, sanitation, and hygiene	44 (7%)
Number of intervention sectors included in design	2	373 (62%)
	3	132 (22%)
	4	87 (14%)
	5 or more	9 (1%)
Sectors in which <u>outcomes</u> measured	Agriculture & food security	40 (7%)
	Economic development	102 (17%)
	Education	149 (25%)
	Environment	13 (2%)
	Governance	8 (1%)
	Health	373 (62%)
	Nutrition	174 (29%)
	Water, sanitation, and hygiene	23 (4%)
	Integrated outcome (bespoke)	1 (<1%)
	Other	54 (9%)
Number of sectors in which outcomes were measured	1	347 (58%)
	2	185 (31%)
	3	60 (10%)
	4 or more	9 (1%)
Geographic area of study	Middle East/North Africa	21 (3%)
	Sub-Saharan Africa	213 (35%)
	Asia	177 (29%)
	Latin American & the Caribbean	188 (31%)
	Europe	4 (<1%)
	Oceania	2 (<1%)
	Number of countries represented	70
Cost analysis conducted	43 (7%)	
Implementation/process evaluation conducted	41 (7%)	

The majority of evaluations (70%) employed a randomized controlled design to assess the effectiveness of their interventions. However, only 26 (4%) of the 601 studies reviewed used a full factorial design and only 12 (2%) employed a partial factorial design. The majority of evaluations (61%) assessed the effectiveness of an integrated intervention by comparing one or more integrated arms to a no-treatment control only. A minority of evaluations included a comparison of integrated arms only (11%), or contained single-sector arms and integrated arms but no control arm (22%). Few evaluations included qualitative (10%) or cost analyses (7%) components.

With regard to what types of interventions and desired outcomes were being assessed, the three sectors most often represented in the intervention design — in order of highest to lowest frequency — were health, education, and nutrition. The same three sectors were also most common in terms of outcomes measured, with nutrition slightly outpacing education for second most common.

For the 38 studies that represented either a partial or a full factorial design, we assessed whether the effectiveness of an integrated intervention — in terms of study outcomes — was evaluated. Of the 26 that were full factorial, seven reported findings that showed that the integrated arm was most effective (De Brauw *et al.*, 2015; Haque *et al.*, 2010; Leventhal *et al.*, 2016; Nga *et al.*, 2009; Nga *et al.*, 2011; Olsen *et al.*, 2003; Widen *et al.*, 2015). Eight demonstrated mixed findings, or did not report the effectiveness of the integrated arm as compared to the other arms (Awasthi *et al.*, 2013; Duflo *et al.*, 2015; Gilgen & Mascie-Taylor, 2001; Halliday *et al.*, 2014; Jinabhai *et al.*, 2001; Kim *et al.*, 2015; Leventhal *et al.*, 2015; Tahlil *et al.*, 2015). In some cases, the added value of integration was reported in another study, or it was stated that the combination was not intended to have effects on the outcomes of the separate sectors, so the data was not fully analyzed or reported. In terms of mixed findings, some studies demonstrated tradeoffs, where integration added value to certain outcomes, but was deleterious for others. Finally, 11 evaluations found no added value of integration (Attanasio *et al.*, 2014; Dangour *et al.*, 2011; Desai & Tarozzi, 2011; Fenn *et al.*, 2012; Friis *et al.*, 2003; Gilgen *et al.*, 2001; Gowani *et al.*, 2014; Mwaniki *et al.*, 2002; Nahar *et al.*, 2012; Rohner *et al.*, 2010; Walker *et al.*, 2006).

Only three of the full factorial studies incorporated cost analysis, and two of the three found that the integrated arm was cost-effective. The third did not perform a cost analysis specifically on the integrated arm.

We also reviewed the findings reported in the 60 studies that included a qualitative component. We found only 1 study that intentionally documented synergy via the qualitative inquiry — the others used the method to investigate other aspects of the intervention.

## Discussion

Our screening of 4,339 records in the 3ie Impact Evaluation Repository identified 601 journal articles that describe studies of

programs we defined as integrated. Our full text review of these 601 articles revealed several interesting trends. First, researchers do not use standardized terms for describing integrated development programs. In fact, the majority of authors did not use any term at all to indicate the integrated or multi-sector nature of the interventions they were evaluating. This finding validates our manual screening methodology. Had we used a key term search strategy, we likely would have missed many relevant studies. Interestingly, 46% of the full factorial evaluations addressed integration or synergy in their abstracts, as compared to only 16% of all studies identified as integrated.

Next, only 26 evaluations employed a full factorial design. Though randomized controlled designs are sufficient to confidently detect the impact of these types of programs, a full factorial design is the only design that truly enables researchers to measure whether the impact is related to the synergy presumed to result from integrated, multi-sector programming.

We recognize that full factorial designs are often costly and time-consuming, and may not be feasible in many, or even most, contexts. They may not be necessary for the types of integrated approaches that have been confirmed in the past to deliver synergy. However, for new or less-researched multi-sector models, robust factorial studies will help sufficiently determine the role of integration in their implementation and results. Factorial or not, examining cost efficiencies and qualitatively assessing synergies are other valuable methods that can help determine how integration factors in to the program findings.

## Limitations

Creating and applying a definition of integrated development was a subjective process. To address this, we utilized two independent coders and employed inter-coder agreement procedures to enhance reliability in our screening process. Assigning core sectors to studies was also a subjective process, and in some cases assigning sectors to an intervention was difficult (e.g., depending on its particular aim, aid to small-scale farmers could conceivably be an economic development/livelihoods, agriculture, or nutrition intervention). We attempted to mitigate this by providing definitions and examples of core sectors to both reviewers, and once a type of intervention was categorized in one way it remained consistent across studies.

Another potential limitation is that the 3ie repository may not be exhaustive; eligible publications in regional or small databases could have been overlooked. Given the size of the repository, however — more than 4,300 publications — including a small number of studies that may be absent in the 3ie repository would not have changed the substance of our findings. Furthermore, there are many studies in the impact evaluation repository (and therefore this review) that focus on the health sector, but that does not necessarily mean that a majority of integrated programs are focused on health. Due to different evaluation cultures within different sectoral communities, and the ease to which some interventions lend themselves to certain types of evaluations, health is almost certainly overrepresented here than if this review had a broader methodological



sampling frame. Other groups have significant evidence bases that are not captured here as they do not fall into the inclusion criteria of the impact evaluation repository, and therefore this review.

We also recognize that in the past 4–5 years an increasing proportion of impact evaluations are being written up as working papers (Cameron *et al.*, 2016) and may never be published. Although our review would have missed unpublished reports, the conclusions we draw in the paper – that synergistic effects are rarely measured in evaluations of integrated development programs – would not likely have changed had unpublished papers been included.

## Conclusions

Our systematic review is not intended to determine whether or not integrated development approaches work. We know from the high number of randomized evaluations included here that in many contexts integrated, multi-sector interventions have produced positive impacts. What our systematic review does indicate, however, is that very few evaluations to date were designed to specifically examine the synergistic and interaction effects that are potentially associated with integrated programming. In other words, to what extent is the integration itself producing the results versus other factors? Impact evaluations of new or yet-to-be-proven integrated programs need to be better designed to intentionally assess not only their impact but the explicit value added of linking two or more development sectors, in

terms of service delivery outcomes, participant perspectives, and cost.

Addressing these gaps is essential as the international community pivots toward a more cross-cutting global development agenda. Implementing this agenda will likely deploy more promising and innovative silo-breaking programs. We must ensure that our research designs and measurement strategies keep pace accordingly.

---

## Competing interests

No competing interests were disclosed.

## Grant information

Bill and Melinda Gates Foundation [OPP1130126] and the FHI Foundation.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

## Acknowledgements

The authors would like to thank Merywen Wigley and Greg Beck, for providing programmatic support and leadership on this grant. The authors would also like to thank Carol Manion, the library specialist who audited the 3ie Impact Evaluation Repository methodology and comprehensiveness.

## Supplementary material

Supplementary File 1: PRISMA checklist.

[Click here to access the data.](#)

Supplementary File 2: PRISMA flowchart, showing the number of records identified, included and excluded.

[Click here to access the data.](#)

Supplementary File 3: Full list of 601 references included in the review.

[Click here to access the data.](#)

## References

Attanasio OP, Fernández C, Fitzsimons EO, *et al.*: **Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial.** *BMJ*. 2014; **349**: g5785.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Awasthi S, Peto R, Read S, *et al.*: **Population deworming every 6 months with albendazole in 1 million pre-school children in North India: DEVTA, a cluster-randomised trial.** *Lancet*. 2013; **381**(9876): 1478–1486.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Cameron D, Mishra A, Brown A: **The growth of impact evaluation for international development: how much have we learned?** *J Dev Effect*. 2016; **8**(1): 1–21.

[Publisher Full Text](#)

Dangour AD, Albala C, Allen E, *et al.*: **Effect of a nutrition supplement and physical activity program on pneumonia and walking capacity in Chilean older people: a factorial cluster randomized trial.** *PLoS Med*. 2011; **8**(4): e1001023.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

De Brauw A, Eozenou P, Moursi M: **Programme participation intensity and children's nutritional status: evidence from a randomised control trial in Mozambique.** *J Dev Stud*. 2015; **51**(8): 996–1015.

[Publisher Full Text](#)

Desai J, Tarozzi A: **Microcredit, family planning programs, and contraceptive behavior: evidence from a field experiment in Ethiopia.** *Demography*. 2011; **48**(2): 749–782.

[PubMed Abstract](#) | [Publisher Full Text](#)

Duflo E, Dupas P, Kremer M: **Education, HIV, and early fertility: experimental**

evidence from Kenya. *Am Econ Rev*. 2015; **105**(9): 2757–2797.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Fenn B, Bulti AT, Nduna T, *et al.*: **An evaluation of an operations research project to reduce childhood stunting in a food-insecure area in Ethiopia.** *Public Health Nutr*. 2012; **15**(9): 1746–1754.

[PubMed Abstract](#) | [Publisher Full Text](#)

Friis H, Mwaniki D, Omondi B, *et al.*: **Effects on haemoglobin of multi-micronutrient supplementation and multi-helminth chemotherapy: a randomized, controlled trial in Kenyan school children.** *Eur J Clin Nutr*. 2003; **57**(4): 573–579.

[PubMed Abstract](#) | [Publisher Full Text](#)

Gilgen D, Mascie-Taylor CG: **The effect of anthelmintic treatment on helminth infection and anaemia.** *Parasitology*. 2001; **122**(Pt 1): 105–110.

[PubMed Abstract](#) | [Publisher Full Text](#)

Gilgen DD, Mascie-Taylor CG, Rosetta LL: **Intestinal helminth infections, anaemia and labour productivity of female tea pluckers in Bangladesh.** *Trop Med Int Health*. 2001; **6**(6): 449–457.

[PubMed Abstract](#) | [Publisher Full Text](#)

Gowani S, Yousafzai AK, Armstrong R, *et al.*: **Cost effectiveness of responsive stimulation and nutrition interventions on early child development outcomes in Pakistan.** *Ann N Y Acad Sci*. 2014; **1308**: 149–161.

[PubMed Abstract](#) | [Publisher Full Text](#)

Halliday KE, Okello G, Turner EL, *et al.*: **Impact of intermittent screening and treatment for malaria among school children in Kenya: a cluster randomised trial.** *PLoS Med*. 2014; **11**(1): e1001594.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Haque R, Ahmed T, Wahed MA, *et al.*: **Low-dose beta-carotene supplementation and deworming improve serum vitamin A and beta-carotene concentrations in preschool children of Bangladesh.** *J Health Popul Nutr*. 2010; **28**(3): 230–237.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Jinabhai CC, Taylor M, Coutsooudis A, *et al.*: **A randomized controlled trial of the effect of anthelmintic treatment and micronutrient fortification on health status and school performance of rural primary school children.** *Ann Trop Paediatr*. 2001; **21**(4): 319–333.

[PubMed Abstract](#) | [Publisher Full Text](#)

Kim DA, Hwong AR, Stafford D, *et al.*: **Social network targeting to maximise population behaviour change: a cluster randomised controlled trial.** *Lancet*. 2015; **386**(9989): 145–153.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Le Blanc D: **Towards integration at last? the sustainable development goals as a network of targets.** Working Paper No. 141. New York: Department of Economic and Social Affairs, United Nations. 2015.

[Reference Source](#)

Leventhal KS, DeMaria LM, Gillham JE, *et al.*: **A psychosocial resilience curriculum provides the “missing piece” to boost adolescent physical health: A randomized controlled trial of Girls First in India.** *Soc Sci Med*. 2016; **161**: 37–46.

[PubMed Abstract](#) | [Publisher Full Text](#)

Leventhal KS, Gillham J, DeMaria L, *et al.*: **Building psychosocial assets and wellbeing among adolescent girls: A randomized controlled trial.** *J Adolesc*. 2015; **45**: 284–295.

[PubMed Abstract](#) | [Publisher Full Text](#)

Mishra A, Cameron D: **Impact Evaluation Repository Search and Screening Protocol.** Washington, DC: International Initiative for Impact Evaluations (3ie). 2014.

[Reference Source](#)

Moher D, Liberati A, Tetzlaff J, *et al.*: **Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.** *PLoS Med*. 2009; **6**(7): e1000097.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Mwaniki D, Omondi B, Muniu E, *et al.*: **Effects on serum retinol of multi-micronutrient supplementation and multi-helminth chemotherapy: a randomised, controlled trial in Kenyan school children.** *Eur J Clin Nutr*. 2002; **56**(7): 666–673.

[PubMed Abstract](#) | [Publisher Full Text](#)

Nahar B, Hossain MI, Hamadani JD, *et al.*: **Effects of a community-based approach of food and psychosocial stimulation on growth and development of severely malnourished children in Bangladesh: a randomised trial.** *Eur J Clin Nutr*. 2012; **66**(6): 701–709.

[PubMed Abstract](#) | [Publisher Full Text](#)

Nga TT, Winichagoon P, Dijkhuizen MA, *et al.*: **Multi-micronutrient-fortified biscuits decreased prevalence of anemia and improved micronutrient status and effectiveness of deworming in rural Vietnamese school children.** *J Nutr*. 2009; **139**(5): 1013–1021.

[PubMed Abstract](#) | [Publisher Full Text](#)

Nga TT, Winichagoon P, Dijkhuizen MA, *et al.*: **Decreased parasite load and improved cognitive outcomes caused by deworming and consumption of multi-micronutrient fortified biscuits in rural Vietnamese schoolchildren.** *Am J Trop Med Hyg*. 2011; **85**(2): 333–340.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Olsen A, Thiong'o FW, Ouma JH, *et al.*: **Effects of multimicronutrient supplementation on helminth reinfection: a randomized, controlled trial in Kenyan schoolchildren.** *Trans R Soc Trop Med Hyg*. 2003; **97**(1): 109–114.

[PubMed Abstract](#) | [Publisher Full Text](#)

Rohner F, Zimmermann MB, Amon RJ, *et al.*: **In a randomized controlled trial of iron fortification, anthelmintic treatment, and intermittent preventive treatment of malaria for anemia control in Ivorian children, only anthelmintic treatment shows modest benefit.** *J Nutr*. 2010; **140**(3): 635–641.

[PubMed Abstract](#) | [Publisher Full Text](#)

Tahlil T, Woodman RJ, Coveney J, *et al.*: **Six-months follow-up of a cluster randomized trial of school-based smoking prevention education programs in Aceh, Indonesia.** *BMC Public Health*. 2015; **15**: 1088.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

United Nations: **Transforming our world: the 2030 agenda for sustainable development.** 2015; (accessed 9 January 2017).

[Reference Source](#)

Waddington H, White H, Snijlsteit B, *et al.*: **How to do a good systematic review of effects in international development: a tool kit.** *J Dev Effect*. 2012; **4**(3): 359–387.

[Publisher Full Text](#)

Walker SP, Chang SM, Powell CA, *et al.*: **Effects of psychosocial stimulation and dietary supplementation in early childhood on psychosocial functioning in late adolescence: follow-up of randomised controlled trial.** *BMJ*. 2006; **333**(7566): 472.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Widen EM, Bentley ME, Chasela CS, *et al.*: **Antiretroviral Treatment Is Associated With Iron Deficiency in HIV-Infected Malawian Women That Is Mitigated With Supplementation, but Is Not Associated With Infant Iron Deficiency During 24 Weeks of Exclusive Breastfeeding.** *J Acquir Immune Defic Syndr*. 2015; **69**(3): 319–328.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Peer Review Status: ? ? ?

Version 1

Reviewer Report 20 December 2017

<https://doi.org/10.21956/gatesopenres.13815.r26082>

© 2017 de Hoop T. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Thomas de Hoop** 

American Institutes for Research, Washington, DC, USA

This systematic review aims to identify and provide an overview of studies that measure the synergistic effects of integrated development approaches. To achieve this goal, the study screened and reviewed articles included in The International Initiative for Impact Evaluation's (3ie) Impact Evaluation Repository. The review addresses an important research question. The approach is rigorous and the objectives of the study are clear. However, the authors also exclude studies that were not published in peer-reviewed journals in the screening and review of the articles. I do not think this was an appropriate methodological choice. In addition, the authors do not take into consideration trade-offs between internal validity and construct validity in their conclusions. I highlight these issues in more detail below in addition to some other smaller comments. The authors make an important contribution to the literature but I would appreciate it if they would take these comments into consideration when they resubmit the paper.

## Excluding studies that are not published in peer-reviewed journals:

The authors explain that they exclude studies that were not included in peer-reviewed journals. It is unclear why they make this choice. It is well known that the findings of peer-reviewed papers can systematically differ from the findings of articles that are not published in peer-reviewed journals. This may be either because of quality differences or publication bias. Risk of bias assessments can mitigate concerns regarding the former. Therefore, it is unclear why the authors exclude the studies that did not appear in peer-reviewed journals. The authors highlight that of the 26 full factorial papers, seven demonstrated that the full factorial arm was most effective. It is very well possible that this percentage may be systematically different in articles that were not published in peer-reviewed journals because of publication bias. So excluding non-peer reviewed studies may well have led to a bias in the results. It may have been more appropriate to conduct a risk of bias assessment.

## Internal validity versus construct validity

The authors argue in favor of designing and implementing impact evaluations with full factorial

designs. I do not disagree with this view. It is important to conduct studies with full factorial designs. However, the reality is that such impact evaluations are very hard to implement in practice in settings without full control of researchers. It is much more challenging to design and implement impact evaluations with a full factorial design when working within government systems or when the intervention is implemented at scale. In contrast, it may be relatively easy (with an emphasis on relatively) to implement a full factorial design when the intervention is designed and implemented by researchers.

The impact evaluation with full control of the intervention may have a high internal validity but the construct validity of the latter design may be much more limited. Estimated effect sizes from full factorial designs can probably not be credibly extrapolated to settings in which the government has full control over the program. So we need to be careful when we recommend impact evaluations with full factorial designs. Their construct validity may be limited if the government or another large implementing partner is unlikely to implement the program with a full factorial design in practice. Or the take-up of the program is much more limited. For example, Bold et al. (2013)<sup>1</sup> demonstrate that the effects of remedial education programs can differ significantly depending on whether they are implemented by an NGO or by the government.

Similarly, full factorial designs may sometimes not be relevant. For example, we conducted a study on the effects of a teacher training program that aimed to promote positive gender socialization in Karamoja, Uganda<sup>2</sup>. This cluster-randomized controlled trial included two treatment arms and a control group. The first treatment arm only included the teacher training. The second treatment arm included the teacher training + an intervention that included text messages related to the contents of the training<sup>2</sup>. In practice, it would not have been useful to design and implement the latter intervention without the teacher training. But as a result the evaluation does not qualify as a full factorial design according to the definition of this study. I do not think this is a limitation of the study by Chinen et al. (2017). Instead, it shows that full factorial designs are not always the best solution to measure synergistic effects of development programs.

## Other Comments

The researchers also highlight the option of analyzing synergistic effects using qualitative research. It would be helpful to learn a bit more about the specific qualitative research the authors would have in mind to measure synergistic effects.

The 3ie Impact Evaluation repository gives a comprehensive overview of impact evaluations in international development. However, it is very well possible that including studies from other databases would have changed the results. It would be good to acknowledge this.

The authors do not assess the conditions under which integrated interventions are more effective. However, the authors highlight this as one of the goals of the study. I would remove this objective.

## References

1. Bold T, Kimenyi M, Mwabu G, Alice Ng'ang'a A, et al.: Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education. *SSRN Electronic Journal*. 2013. [Publisher Full Text](#)
2. Chinen M, Coombes A, De Hoop T, Castro-Zarzur R, et al.: Can Teacher Training Programs Influence Gender Norms? Mixed-Methods Experimental Evidence from Northern Uganda. *Journal on Education in Emergencies*. 2017; **3** (1): 44-78

**Are the rationale for, and objectives of, the Systematic Review clearly stated?**

Yes

**Are sufficient details of the methods and analysis provided to allow replication by others?**

Yes

**Is the statistical analysis and its interpretation appropriate?**

Partly

**Are the conclusions drawn adequately supported by the results presented in the review?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Development Economics, Systematic Reviews in International Development, Impact Evaluations

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Reviewer Report 04 December 2017

<https://doi.org/10.21956/gatesopenres.13815.r26084>

© 2017 Trasi R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Reshma Trasi**

Pathfinder International, Washington, DC, USA

The paper and its subject is extremely timely and relevant given the integrated development focus of the SDGs. I found the methodology, rationale for the sampling frame, phases of the literature review to be well reasoned and laid out. The subject matter of this paper, especially the 38 studies, could prove to be extremely helpful for implementation practitioners, researchers, and policy makers as sector-specific communities grapple with the complexity of designing and delivering integrated development programs in support of the SDGs.

I submit, below, a few constructive suggestions for the authors' consideration to tighten and focus (in some places) or expand (in others) the paper to make it even more useful for the broader development community.

- The authors may want to consider adding the word "impact" into the title itself. "Evaluating the Impact of Integrated Development: are we asking the right questions? A Systematic Review", perhaps?

- In the **Introduction** section, the authors correctly note, “To effectively advocate for integrated, multi-disciplinary approaches to development, it behooves us to understand **under which circumstances** integrating two or more development sectors enhances impacts in amplified or synergistic ways.” (emphasis added). I did not find this addressed in the Findings section. I strongly recommend that the Findings section be bolstered to focus on the findings from the 38 studies to discuss this critical element. Or, consider deleting this as a rationale for the paper to focus it on the objectives mentioned.
- The authors also state, “In this paper we present the results of a systematic review designed to identify if and how synergies and interaction effects between sectors in integrated interventions are being measured.” It will help to clarify that “how” refers to the methods used (experimental vs quasi experimental; # of sectors for which outcomes are measured, etc.) rather than the specific outcomes being measured and evaluated in the integrated programs. [That said, ‘diversity of outcome measures used in integrated programs’ is a paper that I would urge the authors to consider writing next.]
- **Definitional clarity:**
  - I recommend including a little more discussion around the definition that the authors choose for “integrated development” – what other definitions did you consider? Why did you find them lacking? How did you arrive at this definition? How did you define “delivery” of programs? How did you assess, in your review of the papers, “intentional linkage”? How did you assess, in your review, if the interventions were linked by “design” (which I would consider being important because it demonstrates a planned approach)?
  - I struggled with semi-equating integrated development with “multi-disciplinary”. Multi-sectoral is probably the term the authors may want to stick closer to. Multi-disciplinary is more of a research term with definitions of its own.
  - I would imagine that there is wide diversity in the outcomes that each sector uses. It may help to give an example for what counted as an outcome for each sector. It will help ground the paper a bit more for the reader.
- In the **Methods** and **Limitations** section, there are a few steps and limitations missing – language, focus on LMICs, etc. This has been well covered by another reviewer, so I will not repeat these here.
- In the **Discussions** section, I recommend two things for the authors’ consideration:
  - An emphasis on the 38 studies. I would have liked to see a separate table that digs into these studies a little more. – what did these show? Where and on what type of integrated development do these focus vis-à-vis the 601 studies? What are they measuring? What synergistic outcomes did you see? Why do you think this is the case?
  - A discussion on gaps - what gaps did you see?
- In the **Conclusions** section, consider adding a paragraph on the utility value of this important endeavor for various audiences. What are the key takeaways for this audience? Or, perhaps, consider adding an impact evaluation agenda? What are the implications for reaching the SDGs from your viewpoint?



I trust these are helpful. Thank you for the opportunity to review this paper.

**Are the rationale for, and objectives of, the Systematic Review clearly stated?**

Partly

**Are sufficient details of the methods and analysis provided to allow replication by others?**

Yes

**Is the statistical analysis and its interpretation appropriate?**

Yes

**Are the conclusions drawn adequately supported by the results presented in the review?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Health Systems, Interdisciplinary research design, Gender

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Reviewer Report 22 November 2017

<https://doi.org/10.21956/gatesopenres.13815.r26081>

© 2017 Cole D et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Leah Shipton**

Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

**Erica Di Ruggiero**

Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

**Donald C. Cole**

Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

The authors present an interesting review of integrated development approaches and are right to highlight this as an important area of research. The structure of the review is clear and easy to follow the argument and ideas of the authors.

## Introduction - Rationale

- It would be helpful to have “integrated development intervention” defined earlier in the manuscript, especially because authors mention that there has yet to be consensus on a definition in the literature. Providing an example of an integrated development intervention (including a brief description of which features make it an integrated development

intervention) would be helpful in explaining why they are so important in the context of the SDGs.

- The statement below was part of the rationale for conducting this study, but the objectives do not reflect this rationale and neither do the results or discussion sections of the study. Specifically, the authors do not analyze the circumstances when an integrated intervention works. I recommend that this sentence be removed or rephrased to reflect the objectives and results/discussion of the study.
  - "To effectively advocate for integrated, multi-disciplinary approaches to development, it behooves us to understand under which circumstances integrating two or more development sectors enhances impacts in amplified or synergistic ways."
  - The authors might consider circumstances of evaluation in terms of implementation, commitment by participating sectors, funding available, or context (sociocultural, political, institutional, geographical, economic, etc – and this context needs to be defined). For example, how did the institutional structure of sectors involved in an integrated intervention influence the interventions success or failure? How does the geographic context of the integrated intervention influence its success or failure? Etc.

### **Introduction - Objectives**

- The secondary objective of the review seems unnecessary because the authors should state the characteristics of the included studies in the results section as part of the systematic review, therefore this does not need to be an objective.
- Regarding the first objective, it would be helpful to have clarification of what the authors mean by "how." By the end of the article it was clear that the authors mean "how" in terms of evaluation study design (e.g. experimental, factorial) rather than "how" in terms of type of evaluation (e.g. process, outcomes, impact). However, it would be helpful to have this clarified in the objective statement. Specifically, the authors should clarify that they mean to summarize how impact evaluations have been conducted on integrated development programs.

### **Methods**

- Selecting the International Initiative for Impact Evaluation (3ie) Impact Evaluation Repository, which only has impact evaluations, as the sole source of studies for this review is another reason why the authors should consider clarifying that the review aims to summarize how impact evaluations have been conducted on integrated development programs.
- In response to criticism that the MDGs only focused on LMICs, the SDGs were supposed to apply to all countries, irrespective of "development" status. Based on the definition of integrated development interventions/approaches provided by the authors, these approaches are not dependent on location, i.e. an integrated development approach can take place in any country as long as it "intentionally links the design and delivery of programs across more than one core sector." 3ie search strategy limited the scope to LMICs and LMIC regions. If the SDGs – which are meant to apply to all countries – are invoked as rationale for this systematic review, then it might be helpful to have a comment from the authors for excluding integrated approaches in LMIC contexts.
- The 3ie repository search strategy limited inclusion criteria to English language articles. The authors make no comment on how this could have excluded potentially relevant

evaluations for their review, especially if evaluation reports were published by governments and especially considering that these evaluations were conducted in LMIC contexts.

- This is not a typical systematic review because the authors did not design the search protocol or conduct the literature search stages of the systematic review. The authors provide sound justification for their decision to use the 3ie repository to cover these aspects of the systematic review process. However, the authors might consider repositioning this article as a modified approach to a systematic review. The authors provide a link to the inclusion criteria and review methodology. However, it may be helpful if the authors provided brief details of the search protocol and search strategy, for example: language criteria, key terms, databases chosen, inclusion of grey literature. These are aspects of the systematic review search that helps the reader assess the appropriateness of the search and the articles included for analysis.
- It would be helpful for the following sentence on page 4 to be rephrased: "The 3ie review process has no restriction on publication date; however, the systematic review upon which it is based was completed in July, 2016." I needed to review the 3ie repository to understand what this sentence means – that a search of 45 databases was conducted in August 2016, which is where the authors sourced their articles for this review.
  - The authors should also comment on the limitations of this strategy in their limitations section of the manuscript.
- The authors consulted a library information science specialist to audit the methodology. The specialist affirmed that the 3ie repository search included the most relevant databases, however, there were lesser known databases excluded from the search. The authors do not search these databases separately. It would be beneficial if the authors commented on/justified their decision to exclude these databases, despite feedback from a specialist that they may contain relevant articles for the review. The search in these databases could have used the same terms as the 3ie repository search strategy.
- The authors state that there is no universal agreement on an integrated development approach definition and then they present their definition (presumably FHI 360 because of author affiliations, although this should be stated more clearly), which is used for the review. If space is available, it would be helpful to see other definitions of integrated development approaches and justification by the authors as to why their definition was chosen for the review. It would be beneficial to comment on why the authors support a definition that excludes interventions that integrate subsectors of a core sector or interventions that measure multiple sector outcomes for interventions that do not have a multi-sector component. Arguably, these two intervention approaches could also have amplified impacts worth understanding. This discussion around definition may be most appropriate in the introduction section of the manuscript.
- The authors should be more explicit on how the definition of integrated development approaches influenced the review process and inclusion/exclusion criteria.
- The authors should clarify how they decided on the list of sectors for their review such that another set of researchers would come up with the same list of sectors if they replicated the study.

- The authors excluded grey literature articles, and based on the results, setting this criterion excluded 1,163 potentially relevant articles. The authors should justify their decision to exclude this grey literature, especially considering that many impact evaluations may not be published in peer reviewed literature.
  - This is another point that the authors should comment on in the limitations section.

## Results

- It would be interesting to know who conducted the evaluations included in the review (e.g. NGOs, government, UN agencies) and for whom the evaluations were conducted (e.g. consultants hired by government). This information would provide additional insight on the state of integrated development approaches and provide an understanding of who is interested in these evaluations and who is applying the findings of these evaluations.
- It would be interesting to know if there were any sectors that were commonly integrated (e.g. health and nutrition sectors). The commonality of pairings may provide interesting insight on the current state of integrated approaches to development in terms of which sectors are most often paired.
- In the table, the authors present “sectors included in a study intervention” (e.g. 238/601 (40%) of studies included the economic development sector in the intervention) and “sector with outcomes measured” (e.g. 102/601 (17%) of studies had economic development outcomes measured). It may be interesting to have the authors present the numerator as “sector with outcomes measured” over the denominator as “sectors included in a study intervention.” This would illustrate, of the articles that had X development sector involved, what proportion measured X sector outcomes. In the case of the economic development sector, 43% of studies (102/238) that included the economic development sector also had economic development outcomes measured.
- It seems unnecessary for the authors to comment on whether the included articles had a qualitative component considering that the eligibility criteria for the review excluded qualitative evaluation design methods.

## Discussion and Conclusion

- The results show that of the 26 full factorial evaluations, seven reported the integrated arm as more effective, eight had mixed findings, and 11 reported no added value of the integrated arm. This finding relates to the author’s objective to “document if synergies are detected.” The authors should comment on this finding in the discussion section, rather than only commenting on the number of full factorial evaluations conducted and the feasibility considerations associated with this evaluation study design.
  - Specifically, what synergies were detected/how are those synergies defined? How does this finding relate back to the authors definition of integrated development approaches?
- The authors comment that cost efficiency evaluations and qualitative assessments of synergies are also valuable methods to understand the impact of integrated interventions, yet these methods to evaluation were excluded from the review. The authors might consider expanding on why they excluded studies that used these methods and/or comment on whether future research should include evaluation of integrated approaches using these methods.

- It might be interesting for the authors to comment on future directions for research on integrated development interventions.
- The authors state that “Our systematic review is not intended to determine whether or not integrated development approaches work. We know from the high number of randomized evaluations included here that in many contexts integrated, multi-sector interventions have produced positive impacts.” However, the authors did not provide readers with any results on the effectiveness of integrated approaches using RCT evaluations in the results section of the manuscript.
  - Also, what do the authors mean by context? Location?

**Are the rationale for, and objectives of, the Systematic Review clearly stated?**

Partly

**Are sufficient details of the methods and analysis provided to allow replication by others?**

Yes

**Is the statistical analysis and its interpretation appropriate?**

Yes

**Are the conclusions drawn adequately supported by the results presented in the review?**

Yes

**Competing Interests:** No competing interests were disclosed.

**We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.**