

# Research and the real world

## Analyzing the short- and long-term impact of knowledge transfer

**Working Paper****Author(s):**

Caves, Katherine Marie  Lueling, Sarah

**Publication date:**

2020-11-06

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000449607>

**Rights / license:**

In Copyright - Non-Commercial Use Permitted

**Originally published in:**

CES Working Papers 2



# Research and the real world

Analyzing the short- and long-term impact of knowledge transfer

## Authors:

Katherine Caves  
Sarah Lueling

CES Working Paper, No. 2, 2020

# Research and the real world: Analyzing the short- and long-term impact of knowledge transfer

## Abstract

Knowledge exchange interventions are an important driver of research impact in science, policy, and society. This study examines five years of intensive knowledge-exchange “reform labs” in the field of education policy. Using qualitative analysis and a regression analysis approach applied to rich case data and quantitative results, we find that the reform labs have a significant impact on both participants and their reform projects in the short term immediately after the event and in the medium term up to five years later. We also find evidence that the impact on individuals and projects drives broader social impact. In comparing the reform labs to best practices identified in the literature, we find evidence that knowledge exchange interventions of longer duration, with case-focused teamwork, and involving intensive researcher-participant interaction are more impactful. Additionally, we observe that diverse participants can drive impact, as long as their specific needs are met.

## Keywords

Knowledge exchange, knowledge transfer, research impact, evidence-based policy, evidence-based reform, education policy

## 1. Introduction

Modern research in policy- and practice-oriented fields like the social sciences is expected to resonate with and be useful for policymakers and practitioners. Research funding organizations and universities prioritize research relevance, pathways to impact, and knowledge exchange (Bandola-Gill, 2019). Universities include research impact in their mission statements and evaluation criteria, and go as far as to employ or designate knowledge exchange professionals (Knight & Lightowler, 2013). Policymakers work with scholars and knowledge brokers to use research for evidence-based policy (Bandola-Gill & Lyall, 2017; Warira et al., 2017) and to justify investments in research (Adam et al., 2018). The field of research impact evaluation assesses the progress of research toward uptake, use, and impact on society (i.e. Reale et al., 2018; Adam et al., 2018; Morton, 2015; Penfield et al., 2013; Knight & Lightowler, 2013).

Scholars in the social sciences and humanities (SSH) face the challenge of making an impact on both science and society (Bandola-Gill, 2019; Brownson et al., 2006). One approach to driving this impact is through knowledge exchange interventions, or activities specifically designed to disseminate knowledge from research or researchers to the general public, policymakers, or practitioners (Morton, 2015). In the literature on the impact of knowledge exchange activities, researchers search for the best approaches and practices to generate immediate, proximate, and long-term impacts across a variety of fields and participants (Greenhalgh et al., 2016).

This study evaluates an rigorous knowledge exchange activity that brings researchers together with policymakers and practitioners in annual two-week “reform labs” from 2015-2019. Participants join the reform lab in teams, with each team working on a specific education systems reform, usually related to vocational education and training (VET). The reform labs are operated by [redacted for blind review], which was set up explicitly for knowledge exchange.

Using a combination of qualitative and quantitative analysis based on rich case data, evaluations, and follow-up survey data, we find that the reform labs impact both participants and their team projects immediately after the institute and up to five years later. We find that the reform labs have facilitated further research collaboration and improvement of the reform projects. Evidence indicates that, through their impact on participants and their projects, the reform labs may be driving some degree of social impact. We evaluate these findings relative to the best practices described in the literature, and find support for existing best practices as well as evidence that participant diversity can drive impact.

## 2. Literature

Research can have three main types of impact: scientific, social, and political. To sum up, scientific impact is the advancement of knowledge production. Social impact is the implementation of knowledge transfer to improve societal challenges. Political impact is the result of knowledge transfer from research into policymaking and practice (Reale et al., 2018). Much of the initial literature on research impact is focused on the field of health (i.e. Milat, Bauman, & Redman, 2015; Brownson et al., 2006). Since 2010, growing interest has focused on SSH research impact, including policymaking (Oliver et al., 2014; Warira et al., 2017) and specific policy fields like education (Lingard, 2013).

As research impact becomes an increasingly important part of scientific evaluation and funding, some scholars attempt to transmit knowledge directly to policymakers, practitioners, and other stakeholders (Warira et al., 2017). These activities are known as knowledge exchange, or “activities to increase the uptake of research” (Morton, 2015, p. 406). Although most of the work on knowledge exchange has historically been in the field of public health, the necessity for collaboration between researchers and policymakers is attracting interest across a variety of fields (Oliver et al., 2014). Active knowledge

exchange activities are a promising potential solution to the research-policy knowledge gap (*ibid*), as are measures to increase the capacity of policymakers for understanding and using research (Lingard, 2013).

Empirical evidence on the impact of active knowledge exchange shows that each type of activity has its own costs and benefits. One set of approaches connects research to end users without the researcher herself being deeply involved. For example, knowledge brokers are actors that connect science and policymaking through a variety of mechanisms (Bandola-Gill & Lyall, 2017). Some universities employ dedicated knowledge exchange professionals, and although they can be effective these individuals may also have weak relationships to the university, lack access to professional development, and struggle with funding and employment mechanisms designed for traditional research (Knight & Lightowler, 2013).

Other knowledge brokers operate outside of universities as intermediaries, and play a variety of roles including knowledge transmission, research translation, and capacity building (Bandola-Gill & Lyall, 2017). Clearinghouses of data, best practices, and key findings are a common approach to passive knowledge brokering in a variety of fields including SSH, with mixed results (Soydan et al., 2010; Schoenfeld, 2006). The media is another form of knowledge broker, and plays a storytelling role that can be crucial for communicating research findings and new evidence (Davidson, 2017).

Another set of knowledge exchange activities involves direct contact between researchers and their target audience of policymakers, practitioners, or end users. Social media is one such example, and although posting on social media can create a fast and low-cost method of disseminating research and driving impact, social media can also drive or exacerbate publication bias, misunderstandings, and confidentiality concerns (Schnitzler et al., 2016).

Another approach to direct knowledge exchange is the co-production of knowledge, where researchers partner with practitioners, policymakers, or other non-academic partners to carry out policy-relevant research (Durose et al., 2017). Djenontin and Meadow (2018) review knowledge co-production interventions in the environmental sciences and management using a logic frame approach, finding that certain practical barriers can make co-production challenging. Specifically, designing the projects, working across different professional cultures, and project management might all undermine the impact of such activities (*ibid*).

One of the most common examples of knowledge exchange involving direct contact between researchers and end users is the seminar or workshop. Evidence about this type of knowledge exchange activity is mixed. For example, a series of public health workshops in the UK succeeded at facilitating knowledge transfer, translation, and exchange during the workshops but failed to generate longer-term impacts on participants' actions 4-6 months later (Rushmer, Hunter, & Steven, 2014). However, this kind of knowledge exchange activity is generally considered the gold standard (Reed et al., 2014).

Empirical research has yielded a number of best practices and recommendations for high-impact knowledge exchange events. Huntington et al. (2002) compare three workshops in the same field and find varying levels of success. The most successful event was longer, occurred regularly, included preparation of case studies by practitioners in advance, and involved cooperation and exchange among researchers and non-researchers. Rushmer, Hunter, and Steven (2014) recommended taking a comprehensive approach to knowledge exchange events that includes elements of co-creation, iteration, teamwork among end users, and mentoring. Reed et al. (2014) recommend five principles for effective knowledge exchange, which include a strong focus on stakeholder interaction and relationship building to increase engagement and sharing tangibly useful information to foster impact.

### 3. Evaluation frameworks

A number of frameworks are commonly used to evaluate research impact. For example, the Payback Framework is widely applied in public health research (Reale et al., 2018; Milat, Bauman, & Redman, 2015), and comprises three impact stages: secondary outputs, adoption by the public and practitioners, and final outcomes (Donovan & Hanney, 2011). The RAPID framework focuses on the impact of context in the dissemination and uptake of research results (Court & Young, 2006). Knox, Hill, and Berlin (2018) emphasize the cyclical nature of research impact. In general, the utility of different approaches depends on their ability to address the methodological challenges of assessing research impact (Adam et al., 2018). Greenhalgh et al. (2016) review and compare various frameworks for research impact assessment, concluding that varying methods are useful in different contexts.

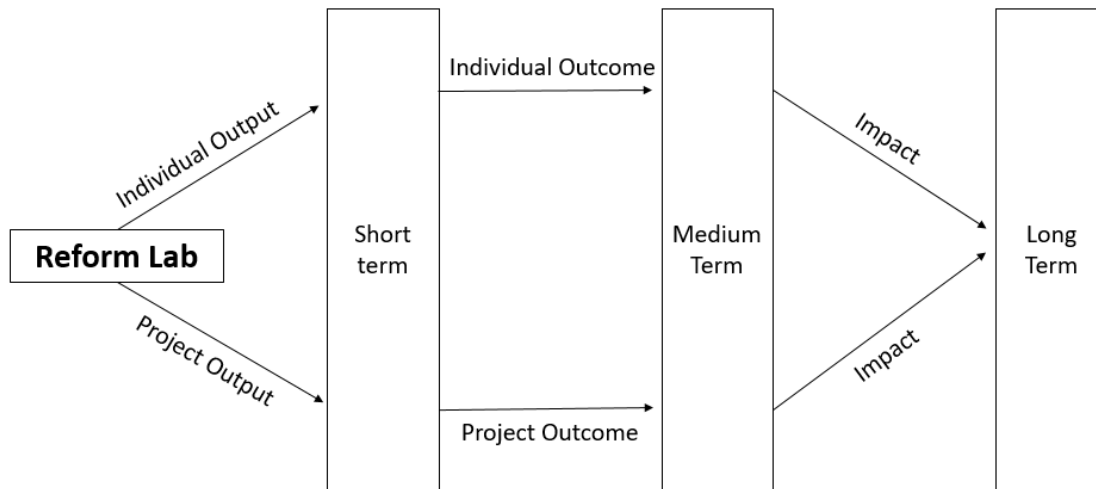
The literature on research impact assessment includes a number of models for how research impact develops. Reale et al. (2018) identify seven steps from research to impact in SSH, including inputs, outputs/activities, engagement/involvement, change in awareness/reaction, changes in knowledge/skills, changes in behavior/practices, and final outcomes or impacts. Solans-Domenech et al., (2019) identify six processes in three stages: first, researchers disseminate their results and build further research capacity; second, they interact with end users and the research influences behavior and practice; finally, research is taken up and has impact on society.

Morton (2015) takes knowledge exchange activities as the unit of analysis and develops a modified version of the Research Contribution Framework (Bell et al., 2011) with its three phases of research uptake, research use, and research impact. Morton (2015) further defines indicators and proposes methods for evaluating the impact of knowledge exchange in each process. This approach is based on a theory of change approach drawn from the evaluation literature on logic models and results-chain analysis (Rogers, 2008). We build on Morton's (2015) approach, using a logic model in conjunction with an intervention-specific theory of change.

We apply the results chain logic model, which is one of the most commonly used approaches in the literature (Cummings, 1997). It has five stages: input, activity, output, outcome, and impact. Inputs are the elements that go into the knowledge exchange activity to enable change. Activity is the knowledge exchange intervention itself. Outputs are short-term changes brought about by the interventions, while outcomes are medium-term effects and impact is long-term change.

The theory of change we apply is based on the reform labs' combination of individual and team-level intervention as shown in Figure 1. In the short term—between the start and end of the program—we expect the reform lab to have an effect on both individuals' knowledge and capacity and an effect on the quality of the team's reform plan. We expect these effects to persist into the medium term (one to five years after the reform lab), and the changes for individuals and teams eventually drive long-term impact. The reform lab itself does not directly drive long-term impact, but individuals and teams with more knowledge, higher capacity, and better reform concepts should drive that impact.

Figure 1: Theory of change



Based on the logic model and theory of change, we investigate five specific empirical questions that address this process. Table 1 summarizes the research questions. Research questions A1 and B2 address outputs on the individuals who attend in the short term, or the change from before to after the reform lab. Research questions A2 and B2 deal with the medium-term outcomes for individuals and projects, in this case one to five years after the reform lab. Finally, research question C3 deals with the impact on the places participants represent, including their education systems and key indicators related to the goals of the reform labs. A framework like this one could apply to a similar intervention in any field.

Table 1: Specific research questions by target group and time point

		Time points		
		T1: Short-term outputs	T2: Medium-term outcomes	T3: Long-term impact
Targets	A: People	A1	A2	-
	B: Projects	B1	B2	-
	C: Places	-	-	C3

Table 2 summarizes our specific hypotheses for each research question along with the data we apply to that question. We describe the data in the next section.

**Table 2: Summary of hypotheses and data**

RQ	Specific hypothesis	Data
A1	In the short term, Individuals who attend the reform lab have an improved knowledge of VET, understanding of the education system and feel more prepared to lead the education system reform after the reform lab compared to before.	Evaluation Survey
A2	In the medium term, individuals attending the reform lab have an improved knowledge of VET, understanding of education and feel more prepared to lead the education reform compared to before the reform lab.	Survey
B1	In the short term, the reform lab has a positive impact on developing and shaping the reform project specifically the goals and the strength of the project.	Case study Evaluation Survey
B2	In the medium term, the reform lab has a positive impact on developing the implementation plan and on shaping the reform project (strength and goals) in participants' home base.	Case study Survey Final report
C3	The reform lab has improved VET outcomes in participants countries/states (also referred to as "places") in the following areas <ul style="list-style-type: none"> <li>A. Program quality</li> <li>B. Program size/scope</li> <li>C. Laws passed</li> <li>D. People working on reform</li> <li>E. Funding raised</li> <li>F. Changed population attitudes</li> </ul>	Case study Survey Final report

## 4. Data

The reform labs are based on best practices in knowledge exchange activities like workshops and seminars where researchers and stakeholders interact directly. The program is long, lasting ten full days. It combines scientific input, extensive researcher-participant interaction, and a variety of knowledge exchange types including lectures, site visits, workshops, and discussions. The reform lab centers around pre-developed case studies for each team, which participants and researchers use to co-create reform implementation plans. The content is geared towards tangibly useful information for reform leaders, and the format encourages intensive relationship building within teams. The labs' goals are to disseminate research results, to improve participants' knowledge, to help teams improve their reform cases, and to drive improvement in the teams' education and systems.

In total, 42 teams attended the reform lab in the 5 years we study, including 185 individual participants. Table 3 shows descriptive statistics. Almost half of all teams were part of a case that attended the reform lab more than once, although participants generally do not overlap. The group size ranges from one to 18 people. Teams come from 14 different countries and 10 different states both in developing and developed countries.



**Table 3: Descriptive statistics**

Characteristic	Categories	Representation
		Total, case study, evaluations
National income level	Low- or middle-income country	50%
	High-income country	50%
Native English	Yes	50%
	No	50%
Time since attendance	1-2 years (2018-2019 cohorts)	33%
	3-5 years (2015-2017 cohorts)	67%
Dosage	Team attending for the first time	55%
	Team is repeating	45%
Team size	Small team (1-3)	64%
	Large team (4-18)	36%
Preparation	Less extensive preparation	26%
	Extensive preparation	74%
Gender	Female	48%
	Male	52%

We use data from three sources to analyze impact: a case study, course evaluations, and a survey. First, we developed a case study of each team (Yin, 2005). This includes their reform lab application, case description, notes from intensive case discussions, plans developed during the reform lab, interview data on key implementation factors, quantitative data on key indicators, a report on the team’s progress three months after the reform lab, and field notes from site visits in roughly one-third of cases.

Second, individual participants answered a course evaluation at the end of the reform lab they attended. This captures participants’ impressions of the reform lab’s impact immediately at the end of their participation.

Third, we sent all participants a survey in 2020, one to five years after participation. The survey includes information on individual progress and changes, team or project progress, and changes in the reform context. From the 185 participants who attended the reform lab, 88 participants completed the survey. The most common reason participant did not fill out the survey was that they had changed employers so their email address was no longer valid. There is an increasing amount of responses each year. For example, 2015 has a 15% response rate and 2019 has a 60% response rate. We combine this data with the other two data sources, which represent all participants.

## 5. Method

Measuring the output, outcomes, and impact of knowledge exchange activities is complicated by a number of key methodological challenges. Adam et al., (2018) summarize the challenges of evaluating research impact: time lags, attribution and contribution, marginal differences, transaction costs, and the unit of assessment (Adam et al., 2018). Choosing a knowledge exchange activity rather than a research project simplifies some of the issues, but we still lack a counterfactual and we did not randomly select participants to attend the reform lab.

We use regression models that test the average difference in outcomes before and after the reform lab, as well as before the reform lab and one to five years later. This model allows us to assess the impact of the reform lab in a given year and over time on specific goals (i.e. to improve participants’ knowledge and understanding of VET and education systems, to help improve teams’ reform cases, and, indirectly,

to improve teams' local education systems). Our model tests both the immediate impact of the reform lab and its extended impact.

To assess whether certain key characteristics had a significant effect on the reform labs' impact, we use seven time-independent control variables including native language, national income, timing of attendance, repetition of team attendance, team size, amount of preparation, and participants' gender. We include team size, repeated attendance at the team level, and the teams' amount of preparation to capture variation across teams. Finally, we include a control for when the individual or team attended the reform lab to capture variation across cohorts.

## 6. Results

We present our results by research question, starting with individuals' short-term outputs and medium-term outcomes, then moving cases' short-term outputs and medium-term outcomes, and finally examining longer-term social impact.

### 6.1. Results for participants

We have two main data sources for individual impact: course evaluations and survey data. Course evaluations, filled out on the last day of the reform lab every year, capture short-term outputs contemporaneously. Three questions on the evaluation address the lab's immediate impact. These deal with participants' understanding of VET, their belief that they can apply what they learned in their own contexts, and the extent to which they feel prepared to work on their projects. On average, participants report very strong subjective impact, stating that understanding of VET improved (4.88 out of 5 points), that they can apply what they learned to their own contexts (4.60), and that they feel more prepared for their projects after the institute (4.58). These scores do not vary dramatically by year, with all years' averages at or above 4.50 in all three questions..

All five years show very high values in all three areas, indicating that the reform lab had a positive impact on key indicators for participants. However, these results are simple averages and, although they imply growth during the institute, it is difficult to quantify the institute's effects. Therefore, we turn to the survey data where we asked participants to report on key indicators both before and after the reform lab.

We have detailed survey data on medium-term outcomes for individuals. The survey asks participants whether they apply lessons from the reform lab to their work and whether the reform lab has affected the way they work. Participants stated that they do apply what they learned to their work (4.33 on a five-point Likert scale) and that the reform lab has affected the way they work (4.49). We also observe that participants maintain a positive attitude towards VET (4.71). 87% of participants are still working in a related field. 63% are still working in the same organization and same position that they were when they attended the reform lab, 17% have moved to a new role in the same organization, and 16% are in a different organization (4% unobserved).

We measure the key indicators for short- and medium-term impact on participants through the survey. We ask respondents to report their knowledge of VET, understanding of education systems, and readiness to lead reforms both before and after attending the reform lab. Table 4 shows the results of our models testing the effect of the reform lab overall and for the control variables. We find that the reform lab has a significant impact on participants on all three indicators in both timeframes ( $p < 0.01$ ).

**Table 4: Results for participants**

	Short-term outputs			Medium-term outcomes		
	VET knowledge	Education systems knowledge	Readiness to lead	VET knowledge	Education systems knowledge	Readiness to lead
Reform Lab Impact <sup>A</sup>	1.38*** <i>(0.10)</i>	0.91*** <i>(0.10)</i>	0.99*** <i>(0.11)</i>	1.57*** <i>(0.10)</i>	1.01*** <i>(0.10)</i>	1.22*** <i>(0.10)</i>
Female <sup>B</sup>	0.20 <i>(0.20)</i>	0.07 <i>(0.19)</i>	-0.08 <i>(0.21)</i>	0.22 <i>(0.21)</i>	0.19 <i>(0.20)</i>	-0.03 <i>(0.21)</i>
2015-2017 cohorts <sup>B</sup>	0.06 <i>(0.20)</i>	0.27 <i>(0.19)</i>	0.25 <i>(0.21)</i>	-0.10 <i>(0.21)</i>	0.20 <i>(0.20)</i>	0.33 <i>(0.21)</i>
First-time team <sup>B</sup>	0.08 <i>(0.20)</i>	0.18 <i>(0.19)</i>	0.09 <i>(0.22)</i>	-0.10 <i>(0.21)</i>	0.14 <i>(0.20)</i>	0.25 <i>(0.21)</i>
Large Team <sup>B</sup>	0.17 <i>(0.20)</i>	0.21 <i>(0.20)</i>	0.10 <i>(0.22)</i>	-0.31 <i>(0.22)</i>	-0.04 <i>(0.21)</i>	-0.32 <i>(0.22)</i>
High-income country <sup>B</sup>	0.36* <i>(0.20)</i>	0.01 <i>(0.19)</i>	0.31 <i>(0.22)</i>	-0.11 <i>(0.21)</i>	-0.20 <i>(0.20)</i>	-0.03 <i>(0.21)</i>
English speaker <sup>B</sup>	0.34* <i>(0.20)</i>	0.11 <i>(0.19)</i>	0.25 <i>(0.21)</i>	-0.21 <i>(0.21)</i>	-0.20 <i>(0.20)</i>	-0.12 <i>(0.21)</i>
Extensive preparation <sup>B</sup>	-0.150 <i>(0.26)</i>	0.24 <i>(0.24)</i>	-0.37 <i>(0.27)</i>	0.08 <i>(0.27)</i>	0.37 <i>(0.25)</i>	-0.35 <i>(0.26)</i>

*Notes: Short-term output is the difference between before and immediately after the reform lab. Medium-term outcome is the difference between before the reform lab and the survey.*

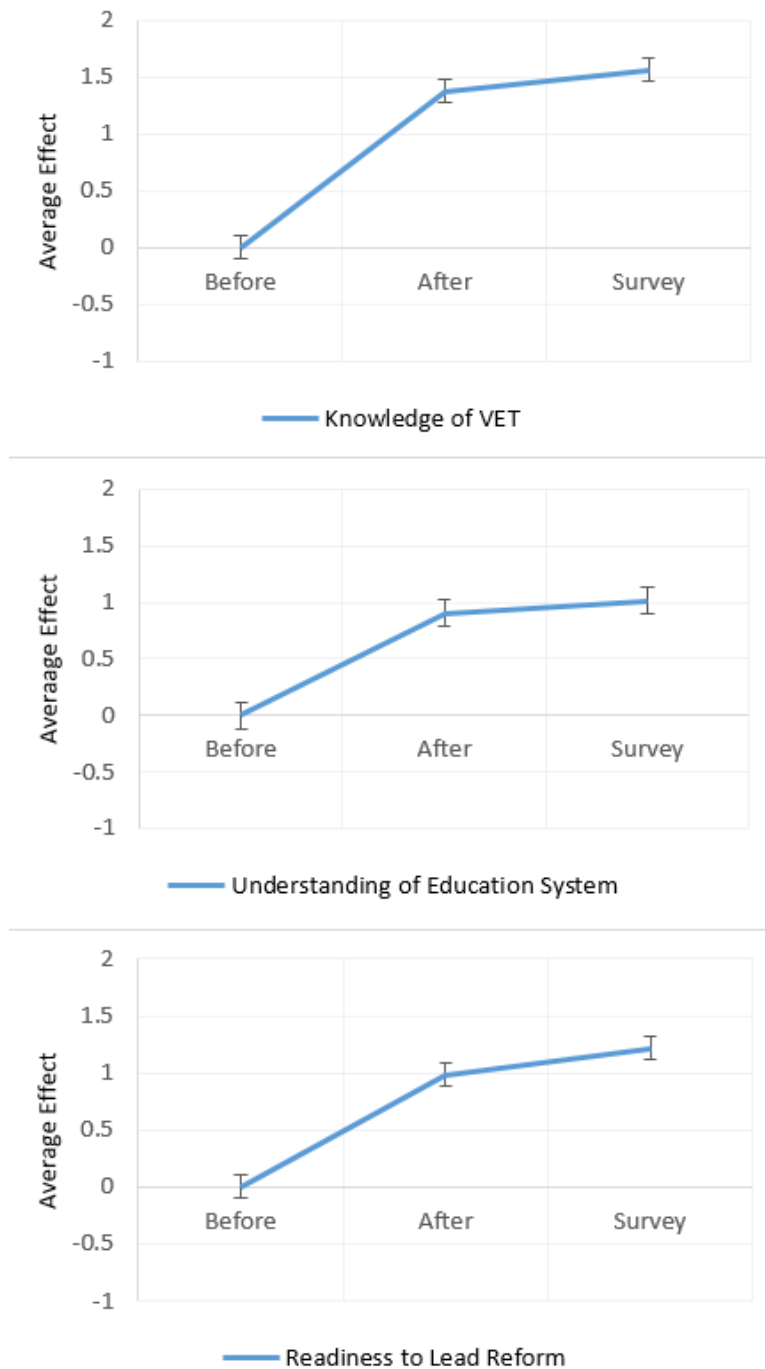
*\* Indicates statistical significance at  $p < 0.10$ , \*\* for  $p < 0.05$ , \*\*\* for  $p < 0.01$ .*

<sup>A</sup> *The overall reform lab impact comes from model (1) where the reform lab's impact is the constant.*

<sup>B</sup> *Variation along control variables comes from model (2) where each control is tested individually to detect a significant impact on the overall impact of the reform lab. The value reported is the coefficient of the control. Each cell shows the constant (model 1, overall impact) or coefficient (model 2, control variables) along with its standard error in italics and parenthesis.*

The effect size is typically around one to 1.5 points on a five-point Likert scale, with VET knowledge the largest in both timeframes followed by readiness to lead and then knowledge of education systems in general. As shown in Figure 2, we observe that all effects not only persist over time, but also grow slightly.

Figure 2: Participant results over time



The control variables are significant when the individual variable changes the overall impact of the reform lab. No control variable is significant for any outcome or timeframe except for high-income countries and English speakers in the short-term VET knowledge case. In this case, participants from high-income countries and those from native-English-speaking countries learn more than those from lower-income countries and those that do not have English as a native language. These controls are significant only at the  $p < 0.10$  level, and they are no longer significant in the medium term.

Based on the evidence listed above, including the evaluation results, the high percentage of alumni working in the field, their agreement that they apply what they learned to their work, and the significant effects of the reform lab on all three key indicators, we accept Hypotheses A1 and A2.

## 6.2. Results for projects

The short-term outputs for team's reform cases focus on the changes in teams' reform plans and implementation strategies. Medium-term outcomes for cases include project persistence, the continued use of the implementation plan made during the reform lab, the current strength of the reform case, and the current appropriateness of the reform case's goals.

One question in the evaluation deals with the short-term output from the reform lab on reform plans. The question asks whether participants believe that their reform plans have improved. The average score over all cohorts is 4.60 on a five-point Likert scale, where a score of five indicates strong improvement. Qualitatively, based on the case study data, we observe that teams tend to change the orient of their planning during the reform lab more towards engagement with employers and certain key governance principles that are discussed during the reform lab as best practices. At the end of the reform lab, their implementation plans are more focused, more evidence-based, and more practical than teams' initial plans tend to be. These qualitative results indicate that teams are taking up the key principles of the reform lab in their planning, but they are not conclusive.

Survey data includes a number of medium-term outcomes for cases. Of the 42 cases that attended the reform lab over all 5 years, 81% are still ongoing and 7% are completed—only 3% have stopped before completion (9% unobserved). We also find that 23% of teams are still using the plans they developed at the reform lab, and an additional 50% are using the plan with alterations. In 11% of teams, the plan is no longer in use (16% unobserved). We also find that 29% of teams are still using the plans they developed at the reform lab, and an additional 45% are using the plan with alterations. In 5% of teams, the plan is no longer in use (21% unobserved).

The survey includes two key indicators related to short-term outputs and medium-term outcomes for projects: reform case strength and the reform case having the right goals. We measure both of these on a five-point Likert scale before and after the reform lab. Table 5 shows the results of the impact models, which demonstrate significant effects for both key indicators in both timeframes.

**Table 5: Short-term outputs for cases**

	Short-term output		Medium-term outcome	
	Strength of Reform case	Appropriateness of reform case goals	Strength of Reform case	Appropriateness of reform case goals
Reform Lab Impact <sup>A</sup>	1.07*** <i>(0.11)</i>	0.95*** <i>(0.09)</i>	0.95*** <i>(0.14)</i>	0.88*** <i>(0.12)</i>
Female <sup>B</sup>	-0.18 <i>(0.22)</i>	-0.11 <i>(0.18)</i>	-0.44 <i>(0.29)</i>	-0.26 <i>(0.23)</i>
2015-2017 cohorts <sup>B</sup>	-0.101 <i>(0.23)</i>	0.02 <i>(0.18)</i>	-0.04 <i>(0.29)</i>	0.01 <i>(0.24)</i>
First-time team <sup>B</sup>	0.06 <i>(0.23)</i>	-0.04 <i>(0.19)</i>	-0.09 <i>(0.30)</i>	-0.04 <i>(0.24)</i>
Large team <sup>B</sup>	-0.07 <i>(0.23)</i>	-0.14 <i>(0.19)</i>	-0.44 <i>(0.30)</i>	-0.54** <i>(0.24)</i>
High-income country <sup>B</sup>	0.28 <i>(0.23)</i>	0.13 <i>(0.18)</i>	0.08 <i>(0.30)</i>	0.02 <i>(0.24)</i>
English speaker <sup>B</sup>	0.01 <i>(0.23)</i>	-0.10 <i>(0.19)</i>	-0.21 <i>(0.29)</i>	-0.10 <i>(0.24)</i>
Extensive preparation <sup>B</sup>	-0.09 <i>(0.26)</i>	0.11 <i>(0.23)</i>	-0.16 <i>(0.37)</i>	0.03 <i>(0.30)</i>

*Notes: Short-term output is the difference between before and immediately after the reform lab. Medium-term outcome is the difference between before the reform lab and the survey.*

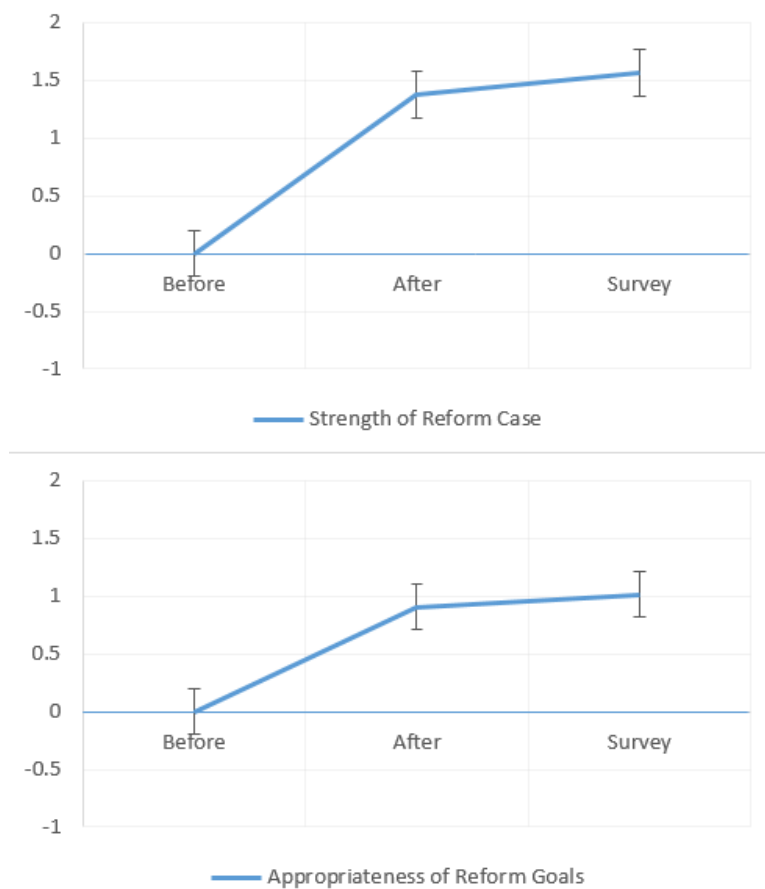
*\* Indicates statistical significance at  $p < 0.05$ , \*\* indicates statistical significance at  $p < 0.01$*

*<sup>A</sup>The overall reform lab impact comes from model (1) where the reform lab's impact is the constant.*

*<sup>B</sup>Variation along control variables comes from model (2) where each control is tested individually to detect a significant impact on the overall impact of the reform lab. The value reported is the coefficient of the control. Each cell shows the constant (model 1, overall impact) or coefficient (model 2, control variables) along with its standard error in italics and parenthesis.*

As shown in Figure 3, reform lab cases are significantly stronger and have significantly better goals at the end of the event compared to the beginning, and this effect persists one to five years after the event ( $p < 0.01$ ). Effects sizes are close to one point on a five-point scale, and get very slightly smaller over time.

Figure 3: Project outcomes over time



Again, there are almost no significant control-variable tests. Only the dummy for larger teams is significant ( $p < 0.05$ ) in the medium-term case for whether the reform has the right goals. Teams that brought larger teams are less likely to have the right reform goals in the medium term, although this effect is not present immediately after the institute. No other control test is significant, indicating that there are no other cases where the differences we test are associated with major differences in impact.

This evidence shows that projects change during the reform lab and that participants believe they improve. Projects also tend to be persistent and they tend to continue using the implementation plans they developed at the reform lab. Therefore, we accept Hypotheses B1 and B2 based on that evidence and the finding that the reform lab has a significant impact on teams' reform cases in the short and medium terms.

### 6.3. Results for long-term impact

The overall impact of the reform labs on science, participants, and society is the most difficult research question to address. For the impact on science, we summarize the collaborations and research projects that have arisen from the reform lab. For the impact on participants, we examine the ongoing contact among individuals. Finally, we use a number of indicators to capture potential impact of the reform labs on society through the reforms and participants.

Approximately half of the cases that have attended the reform lab have engaged in further research projects. These research relationships help reform leaders maintain contact with researchers and develop further evidence-based policy. They have also yielded new research evidence, publications, and theory testing that feeds into the continued development of the reform lab's content.

One way for the institute to have a long-term impact on individuals is to develop networks among professionals. We asked survey respondents whether they are still in contact with other participants, and find that that 67% of participants are still in contact with other alumni. 47% of participants remain in contact with other alumni from their own case teams, and 19% have maintained relationships with participants from other countries. This kind of international networking among reform leaders indicates a result that may generate further impact in the future.

The most difficult impact to assess is that of the reform lab’s cases on their education systems and society. We apply a number of indicators that approach various facets of this potential impact. Table 6 shows the results of a number of statements about specific reform impacts that participants would have targeted based on what they learned at the reform lab. These are measured on a five-point Likert scale, where one indicates no effect at all, three is a moderate effect, and five is a very strong effect. We find that all effects are in the moderate range, with legal changes slightly weaker and the linkage between education and employment slightly stronger.

**Table 6: Strength of reforms’ effects**

Case-level indicator	Strength
Reform increased linkage between employment and education	3.55
Reform increased the number of students doing work-based learning overall	3.35
Reform increased the amount of work-based learning in VET	3.13
Reform has improved the quality of VET	3.23
Reform has resulted in any new laws or legal changes	3.06
Reform improved education governance, either in one program or overall	3.01
Reform has increased shared ownership	3.14

*Notes: strength is measured on a 1-5 Likert scale where 1= strongly disagree, 3= moderately agree, 5 = strongly agree*

The broadest form of impact is overall social change. This is, of course, also the most difficult to measure. To approximate this social impact, we asked participants whether they agree with three statements about how things in their countries or states have changed since they attended the reform lab. These statements cover trends in attitudes related to VET, linkage between education and employment, and change in education systems—all potential main impacts of the reform labs. In all cases, the effects are moderate to strong, with participants reporting moderately strong impact on attitudes toward and awareness of VET (3.75) and moderately strongly increased linkage or shared ownership of VET (3.73), but only moderate degrees of change in their education systems (3.12). We do not interpret these effects as direct impacts of the reform lab, and it is very likely that the same forces attracting teams to a reform lab focused on education systems development also drove these changes.

Qualitatively, a large number of participants state in comments that earlier local views of VET as a second-class education option are gradually dissipating. They also indicate that more people are aware of how VET can become an alternative to traditional upper-secondary education in their countries. We also see that participants agree that the reform lab has led to positive impacts on the relationships between key decision makers in the education and employment systems in their home countries. One participant noted that reform lab alumni have created roundtables to share their experiences with stakeholders and enhance awareness of a VET system’s potential. Others have created quarterly “Community of Practice” meetings that bring together employers, educators, and intermediaries to discuss apprenticeships as vehicles for creating greater opportunity for youth.



## 7. Discussion

Although the knowledge exchange intervention analyzed in this paper is specific to education policy, and our key outcomes refer to impact in that arena, the findings are relevant to knowledge exchange in general. In this section, we discuss how our results relate to specific best practices related to designing and delivering knowledge exchange interventions. We also propose potential additional best practices based on our analyses.

The design of the reform labs aligns with the best practices in knowledge exchange interventions, specifically a long duration, regular recurrence, case-focused work (Huntington et al., 2002), intensive interaction and co-creation between participants and researchers, teamwork, mentoring (Rushmer, Hunter, & Steven, 2014), stakeholder interaction, relationship-building, and a focus on tangibly useful information (Reed et al., 2014). All of these best practices apply across fields the same way they do to this intervention. The results of our analysis show that the program is very effective at its key goals in the short, medium, and long terms. The results found in this study endorse the findings in the literature on the key features of effective knowledge exchange activities. Although we do not vary the characteristics of the reform lab to test specific best practices individually, we can draw some cautious insights from the results related to each practice.

A long duration is necessary simply because it is impossible to convey the full content of the reform labs and work with every case in a shorter time. The specific length of a knowledge exchange intervention will vary by the complexity and depth of the knowledge in question, and we believe that this case is probably at the higher end of that scale. A two-week institute is a major investment for participants, so it may also facilitate impact by selecting for highly interested participants and teams. Therefore, we agree with the literature that long intervention duration facilitates both deeper knowledge exchange and higher participant engagement due to selection.

Reform labs recur annually, creating an opportunity for teams (and in a very few cases, individuals) to repeat participation. Approximately half of teams choose to repeat attendance, sending a new group of individuals for at least two years. We include a control for whether a team is attending for the first time or repeating, but that control is never significant. Therefore, we find that teams see the value in repeating attendance but their results are not different based on sending new teams for multiple years. We do not disagree with the literature and we find qualitative support for the value of repeating knowledge exchange activities regularly, but we do not find empirical support for this particular practice.

One of the most important features of the reform lab is its focus on developing each team's reform case. This requires each team to prepare and work on a case—aligning with the casework best practice—as well as orienting the content of the institute towards tangibly useful information. Both of these are best practices, and we expect them to drive case-level impact. We do find highly significant impact at the case level in the short- and medium terms. We do not find a significant difference between the teams that prepared more extensively and those that were less extensive, but all teams were required to do a certain degree of preparation. Therefore, we align with the literature on the best practices of working directly on cases and on providing applicable information.

One set of best practices has to do with intensive researcher-participant interaction, co-creation, and mentoring between researchers and participants. These are all key features of the reform labs, although effects may vary if participants are native English speakers since that would enable them to communicate more efficiently with the researchers. We do find that native English speakers get more

out of the institute, at least in the medium-term. We also report a number of post-institute follow-up research projects and relationships. Therefore, our findings align strongly with the literature that these intentional and intense interactions between researchers and practitioners drive knowledge exchange impact.

Finally, the last set of best practices focuses on within-practitioner relationships: teamwork, relationships, and stakeholder interaction. One indication of these interactions is team size, with bigger teams potentially bringing a broader range of stakeholders to the table and smaller teams potentially having more intense teamwork experiences. We find that larger teams have significantly lower medium-term outcomes in terms of the reform case's goals being appropriate, but are not significantly different from small teams in any other outcome. This result could indicate that the diversity of perspectives makes everyone harder to please, or it could arise from within-team conflict after the reform lab's conclusion. We also observe that many participants stay in contact with their fellow participants—both from their own teams and from other teams. These relationships, especially those with education leaders from other countries, may not exist without the reform lab and might help participants access new perspectives and find new ideas. Therefore, we tentatively agree with the literature that participant relationships and teamwork drive knowledge exchange impact.

In addition to the best practices described in the literature, we find that participants coming from high-income countries and that speak English get more out of the reform labs. Qualitatively, participants report great value from working across teams—in addition to within teams—and from participating as part of a very diverse group in the same field of practice. Taken together, these findings indicate that participants from lower-income countries that may not have English as an official language are valuable contributors to knowledge exchange interventions, but may not get as much out of them as their high-income, English-speaking colleagues. Therefore, we add to the knowledge exchange best practices that diverse participants drive impact, but their specific needs must be considered.

Although the reform labs are an example of best-practice knowledge exchange, implementing a program like this is very costly. Recruiting teams that can bear the costs or raise the funds necessary to attend an off-site program is not trivial. Organizing the event and investing two full weeks in intensive exchange with practitioners takes a great deal of researchers' time away from normal work and is frankly exhausting. The program pays for itself both financially and in terms of future research ideas and collaborations, but it is not something that can be taken on lightly. Although the investment is so high, we argue that it is worthwhile because of the high return in terms of impact.

## 8. Conclusions

We find that active a long, intensive, case-based intervention that brings researchers and policymakers together to work on concrete problems is an impactful model of knowledge exchange. The impact of this program works on both individual attendees and the cases they bring to the institute, and that impact is clearly evident in the short term immediately after the event and persists into the medium term up to five years later. We see no evidence of deteriorating results over the longer term. Although it is difficult to assess very long-term and society-level impacts, some evidence indicates that these may exist as well.

We make four main contributions to the literature. First, we test a knowledge exchange intervention that aligns with best practices specified by previous research. Second, we test the impact of knowledge exchange in the short, medium, and long terms. Third, we evaluate impact for the cases participants work on as well as for the participants themselves. Fourth, we apply a variety of rich data sources and

rigorous methods. The results imply that the best practices already identified in the literature for knowledge exchange events—including long duration, holding regularly repeated events, case-centered programming with a focus on actionable content, intensive participant-researcher interaction and co-creation, and participant teamwork and relationship-building. We also add that including diverse participants can add to knowledge exchange impact for all participants, especially when organizers take steps to ensure that all participants have equal access to the knowledge exchanged.

Testing the impact of knowledge exchange interventions on individuals and cases and at multiple time points lets us examine how impact emerges, persists, and spreads over time. Social impact arises not just from individuals but also from their reform projects, so—building on Morton’s (2015) approach—we test outcomes for both cases and individuals. We find significant effects for both impact drivers, which indicates that future research evaluating case-based knowledge exchange interventions should also include the case or project effects to evaluate impact fully. We find clear short-term effects for individuals and cases, in alignment with the literature. Although previous empirical results for medium- and long-term results are more mixed, we find that the results persist up to five years after the institute for individuals and for cases. We also find some evidence that the effects for individuals and cases drive broader social impact long-term.

We use a mixed-methods approach with multiple data sources to examine impact from a variety of perspectives and ensure the robustness of our results. We use rich case data on every team over the full time period from institute attendance to the writing of this paper to assess impact qualitatively. Quantitatively, we use evaluation and survey results, testing the significance of impact and look for variation across key case characteristics.

The main limitation of this study is that our results are not causal. We cannot make any causal claims because participants were not randomly selected into the reform lab. We have no counterfactual comparison of similar cases that did not attend the institute. Future research using oversubscribed knowledge exchange events or another randomization opportunity can make a major contribution in that area. In addition, we did not vary the treatment, so our findings related to best practices are also non-causal. We report subjective measures and participants’ self-reported learning. Although we attempt to mitigate bias in these scores by using multiple sources collected at different times, the self-reported data may introduce some bias into our results.

## Acknowledgements

[Redacted for blind review]

## References

- Adam, P., Ovseiko, P. V., Grant, J., Graham, K. E., Boukhris, O. F., Dowd, A. M., Sued, O. (2018). ISRIA statement: ten-point guidelines for an effective process of research impact assessment. *Health research policy and systems*, 16(1), 1-16.
- Bandola-Gill, J. (2019). Between relevance and excellence? Research impact agenda and the production of policy knowledge. *Science and Public Policy*, 46(6), 895-905.
- Bandola-Gill, J., & Lyall, C. (2017). Knowledge brokers and policy advice in policy formulation. In *Handbook of policy formulation*. Edward Elgar Publishing.
- Brownson, R. C., Royer, C., Ewing, R., & McBride, T. D. (2006). Researchers and policymakers: travelers in parallel universes. *American journal of preventive medicine*, 30(2), 164-172.
- Court, J., & Young, J. (2006). Bridging research and policy in international development: an analytical and practical framework. *Development in Practice*, 16(1), 85-90.
- Cummings, F. H. (1997). Logic models, logical frameworks and results-based management: contrasts and comparisons. *Canadian Journal of Development Studies/Revue canadienne d'études du développement*, 18(sup1), 587-596.
- Davidson, B. (2017). Storytelling and evidence-based policy: lessons from the grey literature. *Palgrave Communications*, 3(1), 1-10.
- Djenontin, I. N. S., & Meadow, A. M. (2018). The art of co-production of knowledge in environmental sciences and management: lessons from international practice. *Environmental Management*, 61(6), 885-903.
- Donovan, C., & Hanney, S. (2011). The 'payback framework' explained. *Research Evaluation*, 20(3), 181-183.
- Durose, C., Needham, C., Mangan, C., & Rees, J. (2017). Generating 'good enough' evidence for co-production. *Evidence & Policy: A Journal of Research, Debate and Practice*, 13(1), 135-151.
- Greenhalgh, T., Raftery, J., Hanney, S. et al. Research impact: a narrative review. *BMC Med* 14, 78 (2016). <https://doi.org/10.1186/s12916-016-0620-8>
- Huntington, H. P., Brown-Schwalenberg, P. K., Frost, K. J., Fernandez-Gimenez, M. E., Norton, D. W., & Rosenberg, D. H. (2002). Observations on the workshop as a means of improving communication between holders of traditional and scientific knowledge. *Environmental Management*, 30(6), 0778-0792.
- Knight, C., & Lightowler, C. (2010). Reflections of 'knowledge exchange professionals' in the social sciences: emerging opportunities and challenges for university-based knowledge brokers. *Evidence & policy: a journal of research, debate and practice*, 6(4), 543-556.
- Knox, V., Hill, C. J., & Berlin, G. (2018). Can evidence-based policy Ameliorate the nation's social problems?. *The Annals of the American Academy of Political and Social Science*, 678(1), 166-179.
- Lingard, B. (2013) The impact of research on education policy in an era of evidence-based policy, *Critical Studies in Education*, 54:2, 113-131, DOI: 10.1080/17508487.2013.781515

- Milat, A.J., Bauman, A.E. & Redman, S. A narrative review of research impact assessment models and methods. *Health Res Policy Sys* 13, 18 (2015). <https://doi.org/10.1186/s12961-015-0003-1>
- Morton, S. (2015). Progressing research impact assessment: A 'contributions' approach. *Research Evaluation*, 24(4), 405-419. <https://doi.org/10.1093/reseval/rvv016>
- Oliver, K., Innvar, S., Lorenc, T. et al. A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC Health Serv Res* 14, 2 (2014). <https://doi.org/10.1186/1472-6963-14-2>
- Reale, E. , Avramov, D., Canhial, K., Donovan, C., Flecha, R., Holm, P., Larkin, C., Lepori, C., Mosoni-Fried, J., Oliver, E., Primer, i E., Puigvert, L., Scharnhorst, A., Schubert, A., Soler, M., Soòs, S., Sordé, T., Travis C., Van Horik R., (2018). A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research. *Research Evaluation*, 27(4), 298-308.
- Reed, M. S., Stringer, L. C., Fazey, I., Evely, A. C., & Kruijssen, J. H. (2014). Five principles for the practice of knowledge exchange in environmental management. *Journal of environmental management*, 146, 337-345.
- Rogers, P. J. (2008) 'Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions', *Evaluation*, 14/1: 29–48.
- Rushmer, R. K., Hunter, D. J., & Steven, A. (2014). Using interactive workshops to prompt knowledge exchange: a realist evaluation of a knowledge to action initiative. *Public Health*, 128(6), 552-560.
- Schnitzler, K., Davies, N., Ross, F., & Harris, R. (2016). Using Twitter™ to drive research impact: a discussion of strategies, opportunities and challenges. *International Journal of Nursing Studies*, 59, 15-26.
- Schoenfeld, A. H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher*, 35(2), 13-21.
- Soydan, H., Mullen, E. J., Alexandra, L., Rehnman, J., & Li, Y. P. (2010). Evidence-based clearinghouses in social work. *Research on Social Work Practice*, 20(6), 690-700.
- Warira, D., Mueni, E., Gay, E., & Lee, M. (2017). Achieving and Sustaining Evidence-Informed Policy Making: Effective Communication and Collaboration Can Go a Long Way. *Science Communication*, 39(3), 382-394.
- Yin, R. K. (2015). Causality, generalizability, and the future of mixed methods research. In *The Oxford handbook of multimethod and mixed methods research inquiry*.