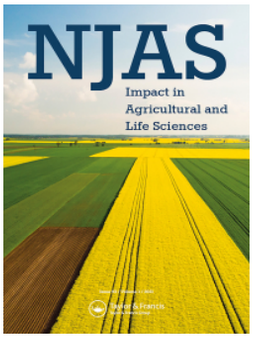

NJAS: Impact in Agricultural and Life Sciences

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tjls21

Co-innovation and Integration and Implementation Sciences: Measuring their research impact - examination of five New Zealand primary sector case studies

Bruce Small, Melissa Robson-Williams, Penny Payne, James A. Turner, Roger Robson-Williams & Akiko Horita

To cite this article: Bruce Small, Melissa Robson-Williams, Penny Payne, James A. Turner, Roger Robson-Williams & Akiko Horita (2021) Co-innovation and Integration and Implementation Sciences: Measuring their research impact - examination of five New Zealand primary sector case studies, NJAS: Impact in Agricultural and Life Sciences, 93:1, 5-47, DOI: 10.1080/27685241.2021.1957267

To link to this article: https://doi.org/10.1080/27685241.2021.1957267

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 02 Sep 2021.

Submit your article to this journal

Article views: 394

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tjls21

NJAS: IMPACT IN AGRICULTURAL AND LIFE SCIENCES
2021, VOL. 93, NO. 1, 5–47
https://doi.org/10.1080/27685241.2021.1957267

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Co-innovation and Integration and Implementation Sciences: Measuring their research impact - examination of five New Zealand primary sector case studies

Bruce Small[a], Melissa Robson-Williams[b], Penny Payne[a], James A. Turner[a], Roger Robson-Williams[c] and Akiko Horita[a]

[a]AgResearch, Hamilton, New Zealand; [b]Manaaki Whenua – Landcare Research, Lincoln, New Zealand; [c]Plant and Food Research, Lincoln, New Zealand

ABSTRACT

The Primary Innovation programme investigated co-innovation to solve complex agricultural problems in five New Zealand primary sector projects. The projects engaged diverse stakeholders using a collaborative, integrative process to co-define problems, and co-create and implement solutions. Each project included a Reflexive Monitor, who facilitated group relationships, encouraged a systems perspective, and integration of multiple disciplinary and stakeholder knowledges. Reflexive Monitors also encouraged reflexive practice and adaptive project management, while helping the team pursue the project ambition for change. This paper, with respect to the five projects, seeks to address the following research question: Is co-innovation an effective research approach for achieving societal impact from innovations? To address this question, we describe attempts to operationalise and measure co-innovation through 1) five behavioural principles of co-innovation, 2) Reflexive Monitors' focus on each principle, and 3) the presence or absence of elements of the Integration and Implementation Sciences Framework (i2S) for enhancing research impact. We evaluate the relationship between these three process measures and project success, measured by outputs and two proxy impact measures: participants' subjective comparisons with the counterfactual and anticipated achievement of desired long-term impacts. Results indicated that the five principles of co-innovation and the presence or absence of elements defined in the i2S framework were positively related to the three success measures. This suggests validity of these measurement tools, and of using a co-innovation approach and/or systematic attention to the elements of the i2S framework to enhance the processes, outcomes and impacts of projects tackling complex real-world problems.

KEYWORD Complex problems; co-innovation; transdisciplinary; integration and implementation sciences; collaboration; process measures; impact measures

CONTACT James Turner James.Turner@agresearch.co.nz

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

This paper uses five case studies from the Primary Innovation programme to assess how well these projects integrated knowledge, values and goals from a range of different scientific disciplines and stakeholder groups, across multiple interconnected systems to provide collaboratively developed solutions to complex real-world problems. Primary Innovation, trialled, evaluated and adapted a *co-innovation* approach to test the usefulness of co-innovation for addressing some complex problems in New Zealand's agricultural production system. This programme finished in September 2017 and has been the subject of a number of peer-reviewed published papers (e.g. Vereijssen et al. 2017; Botha, Turner, 2017; Fielke et al. 2018), a special issue of the journal 'Outlook on Agriculture' (Volume 46, Issue 2, 2017) and the Primary Innovation website (https://www.beyondresults.co.nz/primary-innovation/). The projects engaged a diverse range of stakeholders from science, industry, policymaking, practice and primary sector end-users, using a collaborative, integrative process to co-define problems, and co-create and implement solutions.

Primary Innovation was a research programme initiated to address the issue that "Technology transfer in the NZ biological industries is substantially under-delivering the potential impacts of new technological advances" (Turner and Botha, 2012, p. 7). Through engagement and participation of relevant actors in the agricultural innovation system, it is proposed that co-innovation may enhance the societal impact of research, increasing the uptake of improved practices and new technological innovations. Thus, a particular focus of Primary Innovation was on the efficacy of the practice of co-innovation to create targeted societal impact. In respect to the five Primary Innovation projects, we seek to shed light on the following research question: Is co-innovation an effective research approach for achieving societal impact from innovations? Another hypothesis of Primary Innovation was that co-innovation is better suited to more complex problems. Although we are unable to directly address this second question in the current paper, by examining the nature of complexity in our five case-study, real-world, research problems, we can offer insight into differences in their structural complexity, which have logical implications for an appropriate research focus to achieve societal impact.

We also examine the five co-innovation projects through the lens of the Integration and Implementation Sciences (i2S) framework developed by Bammer (2013). This is a transdisciplinary framework for conducting actionable research to solve complex real-world problems. The framework was also designed as a tool to systematise the documentation of transdisciplinary

projects and enable comparative analyses of transdisciplinary projects. We attempt to measure, for each project, the degree to which the elements of the i2S framework have been considered and addressed. Before discussing co-innovation, i2S, research problem complexity and the methods developed to measure them, we briefly introduce the five Primary Innovation projects that served as case studies. Further details about the specific nature of each case study may be found in Vereijssen et al. (2017).

2. The five co-innovation case studies

Nutrient Management (NM) Project – this project started in 2012 and from conception was designed to be a co-innovation project. The aim of the NM project was to support farmers to comply with changing, and increasingly stringent, regional water quality regulations (Pinxterhuis et al. 2018).

Tomato Potato Psyllid (TPP) Project – this project commenced in October 2013 but only became involved with Primary Innovation (and the practice of co-innovation) from June 2014. The aim of the TPP project was to develop controls for the TPP, a pest damaging potato crops, for the New Zealand potato industry (Vereijssen, Williams, 2017b).

Heifer Rearing (HR) Project – this project began in September 2013 and from conception was designed to be a co-innovation project. The aim of the HR project was to improve the New Zealand dairy herd reproductive performance by lifting the proportion of heifers entering the national herd at target live weight (Pinxterhuis et al. 2018).

Water Use Efficiency (WUE) Project – this project started in October 2012 and from conception was designed to be a co-innovation project. The aim of the WUE project was to improve on-farm irrigation decisions using better characterisation of irrigation demand and accurate short-term weather forecasts (Srinivasan, Bewsell, Jongmans, Elley 2017; 2019).

Log Segregation (LS) Project – this was a forestry sector project the aim of which was to develop cost-effective approaches to characterise and deal with variation in wood properties within and between trees to enhance value-added production (Moore 2019).

3. The nature of co-innovation

Transdisciplinary research to address complex real-world problems emphasises the need to involve disciplinary experts along with non-academic stakeholders 'to generate a shared understanding and way of tackling the problem, which then guide the investigation and the implementation of its findings' (Bammer 2013, p. 8). Co-innovation puts this transdisciplinary,

collaborative concept into practice and has gained popularity as an approach to deal with complex issues in agricultural systems science. "The intention is that problems are addressed through a co-evolution of technologies, practices, policies and market changes undertaken in processes of collaboration and negotiation involving multiple stakeholders" (Coutts et al. 2017, p. 99). A significant element of co-innovation is the involvement of appropriate processes and actors in the research programme to ensure solution implementation via increased relevance, buy-in and aligned social, institutional and market changes. To achieve the desired outcome – innovations in policy, practice or technology – the actors in co-innovation are required to co-learn about the shared problem, context and science, co-create and co-evaluate solution ideas, and co-develop and implement them, as their involvement, skills and roles are required (Botha, Klerkx, Small, Turner 2014).

In practice, there seems to be few hard and fast rules governing the concept of co-innovation. Rather, co-innovation might be described as an organically evolving toolkit of principles, practices, processes, tools and techniques designed to help innovation teams come up with practical, sustainable solutions to complex problems (Botha et al. 2017). Co-innovation aims to produce solutions with attributes of scientific credibility, political and organisational authority, and public and stakeholder legitimacy through the process of fostering dialogue, engagement and co-production with actors in each of these realms.

Numerous authors have contributed to defining different aspects of co-innovation. Klerkx et al. (2012) and Wieczorek and Hekkert (2012) discuss its systemic nature. Bitzer and Bijman (2015) emphasise the aspects of collaboration, coordination and complementarity, while Lee, Olson and Trimi (2012) focus on collaboration, convergence and co-creation. Kilelu et al. (2013) note that concepts, values, technologies, and institutional rules may need to co-evolve. Beers et al. (2016) discuss the relevance of co-learning and social learning amongst participant stakeholders, while Klerkx and Gildemacher (2012) note the importance of brokering in the practice of co-innovation, and Ingram et al. (2020) highlight the role of contextual factors and facilitation processes in shaping co-innovation processes.

The practice of reflection and reflexivity (Vereijssen et al. 2017; Ingram, Gaskell, Mills, Dwyer 2020) and the use of adaptive management (Klerkx, Aarts, Leeuwis 2010) all contribute to the concept of co-innovation. Co-innovation teams sometimes utilise a process referred to as 'Reflexive Monitoring in Action' (RMA). This is an interactive methodology to encourage reflection and learning within groups of diverse actors that seek to contribute to system change to deal with complex problems. Iterative, collective reflection on the barriers and opportunities of the current system helps to

stimulate collective learning and the design of adaptive systemic interventions. The RMA methodology also involves iterative reflection on the institutional setting (context) in relation to the long-term ambition for change (project purpose/end-goal) and project actions and activities and their effects (Van Mierlo, Arkesteijn, Leeuwis 2010; Van Mierlo, Regeer, van Amstel et al. 2010b). To encourage and help innovation project teams implement the RMA methodology, a specialist role called a 'Reflexive Monitor' is sometimes employed.

4. Co-innovation as practiced in Primary Innovation

Primary Innovation, while employing most of the elements of co-innovation noted above, nonetheless, like others before, developed its own unique version of co-innovation; focussing on the role of research in co-innovation for societal impact. It involved taking a systemic approach in each of the case studies (Turner, Klerkx, Rijswijk, Williams, Barnard 2016; Turner et al. 2017), and used Social Network Analysis along with knowledge and innovation brokering (King 2017; King, Fielke, Bayne, Klerkx, Nettle 2019). The concepts of up-scaling and out-scaling in the agri-food sector were employed (Beechener 2017; Beechener 2020) using a community of change (Turner et al. 2017b). Additionally, the programme utilised monitoring and evaluation theory and practice for learning and context-specific adaptive management in the projects (Turner et al. 2017; Botha, Coutts, 2017; Vereijssen, Williams, 2017).

In addition, each innovation project team was assigned a Reflexive Monitor, to assist them to put co-innovation into practice and help realise the projects' change ambitions/impact goals (Fielke et al. 2017; Rijswijk, Bewsell, Small, Blackett 2015). Primary Innovation also developed a set of guiding principles. Initially, nine principles were identified (Coutts et al. 2017) informed by ex-post analysis of three co-innovation projects (Fielke et al. 2018), and loosely based on Nederlof et al.'s (2011) principles for agricultural innovation platforms. During the course of the programme, these nine principles were refined to five key principles designed to guide project teams in their practice of co-innovation (Boyce et al. 2016).

However, despite the above similarities between the five co-innovation projects, the teams worked independently of each other with each team applying the principles of co-innovation in its own way, and to varying degrees, to reflect the unique context of the project; team capabilities, and the nature of the problem being worked on (Pinxterhuis et al. 2018). Likewise, each group had its own Reflexive Monitor who developed a unique relationship with the rest of the project team, placing differing degrees of emphasis

on each of the five principles. Thus, there was variation amongst the co-innovation teams with respect to their practice of the principles of co-innovation and their Reflexive Monitors' focus on the principles (Coutts et al. 2017). For the purpose of this paper, we will expand on the latter two aspects of co-innovation as practiced in Primary Innovation: the five principles of co-innovation and Reflexive Monitoring.

5. The five principles of co-innovation

The nine principles (Coutts et al. 2017) were consolidated into the following five behavioural principles (Boyce et al. 2016, p. 3) by drawing on lessons from successful research projects (Turner et al. 2017):

(1) Involve partners and stakeholders – this is about being inclusive of all who may be affected by the research issue or its solution, involving them in the research process from the beginning planning stages (Boyce et al. 2016).
(2) Take a problem focus – 'The focus of your research proposal should be on addressing a problem or realising an opportunity as the outcome of an interactive [process] . . . rather than linear technology transfer of a solution from "experts" to "end-users"' (Turner, Klerkx, Rijswijk, Williams, Barnard 2016: p. 4 as cited in Boyce et al. 2016, p. 7).
(3) Assemble and nurture the right team – an interdisciplinary/transdisciplinary research team requires a diverse range of skills, not just limited to appropriate disciplinary and interdisciplinary skills but also including social process skills such as collaboration, facilitation, managing group dynamics and power hierarchies, managing interpersonal relationships, brokering, negotiating, knowledge exchange, knowledge integration and science translation. The aim is to form a transdisciplinary team capable of integrating both scientific disciplinary knowledge along with local knowledge and expertise of the lay or interested public (Boyce et al. 2016).
(4) Front up: Share results early and often – "Supporting a team using a co-innovation approach involves researchers fronting-up to research programme participants early and often in order to share data and results as they emerge, rather than waiting until the end of the research" (Boyce et al. 2016, p. 11).
(5) Plan-Do-Observe-Reflect: the action learning cycle – "The Plan – Do – Observe (Monitor) – Reflect cycle emphasises the importance of acting on what is discovered in the Observe stage. The intention of this principle is to maintain: a focus on action" (Boyce et al. 2016, p. 13).

Echoing Coutts et al. (2017), our first hypothesis regarding the five case studies was:

H1: The stronger the practice of the principles of co-innovation in a project the greater the probability of the project achieving its desired long-term impact.

To test this hypothesis a scale was developed to measure the degree to which each co-innovation project team practiced each of the five principles. A measure of project output fitness for purpose, and two proxy measures of project impact, were developed as evaluation criteria to test hypothesis 1.

6. Reflexive monitoring

In the Primary Innovation case studies, Reflexive Monitoring in Action (RMA) was an important aspect of the operationalisation of co-innovation. This was supported by the assignment of a Reflexive Monitor to each team.

The Reflexive Monitors' roles evolved during the lifetime of Primary Innovation, but it became clear across the separate projects that there was no single definition, and that the role was dependent on the group and project (Fielke et al. 2017). Most of the Reflexive Monitors had a background in social sciences, facilitation and negotiation skills as well as an understanding of social research skills, such as qualitative interviewing and analysis (Payne 2017). Van Mierlo et al. (2010b) identified two main methods, at opposite ends of a continuum, that may be used by a Reflexive Monitor: from appreciative inquiry to critical realism. The Reflexive Monitors in Primary Innovation identified the major tasks that they carried out as (Payne 2017):

- Supporting role – supporting project manager/team leader and wider project team
- Getting the project team to where it needs to go
- Identifying conflict – helping to constructively address conflict
- Data collector/evaluator – monitoring and making sure the project is on track
- Facilitating project meetings
- Providing feedback to the project team
- Making sure that the right stakeholders are involved and that everyone has a chance to participate.

The Reflexive Monitor was not necessarily deeply involved in the project research but was required to have a good understanding of the goals of the project and relationship with the project manager, who must also have an understanding of the co-innovation process and the role of the Reflexive

Monitor. A significant component of a Reflexive Monitor's role was understanding and managing group dynamics and interpersonal relationships amongst the project team. When the five principles of co-innovation were developed, it became the Reflexive Monitors' job to encourage the practice of the five principles in the projects.

Therefore, our second hypothesis was:

H2: The greater a reflexive monitor's focus on ensuring the project practiced the principles of co-innovation, the greater the probability of the project achieving its desired long-term impact.

To test this hypothesis a scale was developed to measure the degree of emphasis a Reflexive Monitor placed on the practice of each of the five co-innovation principles. The output fitness for purpose measure and the two proxy measures of project impact, were used as evaluation criteria to test hypothesis 2.

Next, we briefly overview a recently developed framework for conducting research addressing complex real-world problems, the Integration and Implementation Science Framework (i2S) and propose hypothesis 3.

7. Integration and implementation science framework

Another, perhaps more systematic, approach to dealing with complex real-world problems is the Integration and Implementation Science Framework (i2S) proposed by Gabriele Bammer (2013). The discipline of Integration and Implementation Sciences (i2S) arose as a response to the question 'How can academic research enhance its contribution to addressing widespread poverty, global climate change, organised crime, escalating healthcare costs or the myriad other major problems facing societies' (Bammer 2013, p. 3).

The i2S framework is as an effort to systematise an approach to what is variously called multi-disciplinary, interdisciplinary, transdisciplinary, action research, systemic intervention or integrative applied research. Bammer (2013), in noting that there was no systematic way of approaching, planning or describing these kinds of research projects, which tackled complex real-world problems collaboratively, sought to develop a framework that would systematise the process of integrating disciplinary and stakeholder knowledge, dealing with remaining unknowns and implementing research solutions. Bammer argued that there is a need for a common systematic framework to enable comparative description and analysis across multiple transdisciplinary studies to understand what works and what does not work, and therefore help progress the science of transdisciplinary research.

Such a framework would have to handle systems, values, contexts, unknowns, and imperfection; which reductionist disciplinary approaches are unable to competently address. Systems require a holistic understanding but may have undefined boundaries, feedback loops and non-linear processes, spatial/temporal cause-effect delays, and emergent properties. Consequently, manipulation of systems components can sometimes lead to unexpected or unintended consequences; therefore, the framework must include ongoing monitoring, evaluation and reflection regarding the research intervention and resulting system change. Value conflicts are common in transdisciplinary research projects. These arise from competing values and different desired end-goals, held by different legitimate stakeholders in the research process or intervention. Value conflicts must also be revealed and addressed through the transdisciplinary research framework (Ingram, Gaskell, Mills, Dwyer 2020; Turner et al. 2020a).

Contexts may involve relevant but complex, social, historical, cultural, political, economic and other circumstances, and issues of authorisation and institutional settings. While unknowns, such as known unknowns (some of which will be research questions), unknown knowns (e.g., tacit knowledge) and most problematically, unknown unknowns (including the influence of unknown factors, and the unanticipated or unforeseeable emergent consequences of our research or project activities), may also play an important role in the integration of knowledge and the successful implementation of research solutions.

Bammer (2013, p. 18) identifies a 'research style' that underpins the discipline of i2S, *integrative applied research*, and defines it as 'a research style that deals with complex real-world problems by bringing together disciplinary and stakeholder knowledge and explicitly dealing with the remaining unknowns, in order to use that integrated research to support practice and policy change.' The discipline of i2S has a theoretical structure of three domains 1) synthesising disciplinary and stakeholder knowledge, 2) understanding and managing diverse unknowns and, 3) providing integrated research support for policy and practice change. It provides a five-question framework to help 'flesh out' and understand how each of these three domains is considered and acted upon in a project. These five questions are applicable across all three domains. Bammer (2013, p. 21) gives the questions in a generalised form (i.e., applicable across the three domains), as:

(1) What is the integrative applied research aiming to achieve and who is intended to benefit?
(2) What is the integrative applied research dealing with – that is, which knowledge is synthesised, unknowns considered and aspects of policy and practice targeted?

(3) How is the integrative applied research undertaken (the knowledge synthesised, diverse unknowns understood and managed, and integrated research support provided), by whom and when?
(4) What circumstances might influence the integrative applied research?
(5) What is the result of the integrative applied research?

These questions may be asked in any order and may be used iteratively throughout the research programme to plan, describe/observe, and reflect. Integrative applied research places a clear 'emphasis on the exchange of methodological insights between research groups working on very different problems' (Bammer 2013, p. 19). Each question in each domain has a series of probes, which seek to ascertain if elements of the framework have been considered or applied. The i2S framework with its three domains and five questions provides a mechanism for planning, describing and documenting the methodology, analysing the results of integrative applied research projects, and comparing across multiple transdisciplinary research projects. From the above descriptions we may conclude that the co-innovation process is an example of integrative applied research to find solutions to complex real-world problems. In our five case studies the research problems lay in the agricultural innovation systems realm.

Since the i2S framework is specifically designed to increase the value and usefulness of research at solving complex real-world problems, and therefore create impact, we proposed hypothesis 3:

H3: The more a project considers/addresses/applies the questions and elements of the i2S framework, the greater the probability of the project achieving its desired long-term impact.

To test this hypothesis a process and a measure were developed to test for the presence or absence of i2S elements in the co-innovation case studies. The output fitness for purpose measure and the two proxy measures of project impact are used as evaluation criteria to test hypothesis 3.

Before proceeding to describe the various independent and dependent measures developed for use in this study, we first consider the concept of 'complexity' in real-world problems.

8. The dimensionality of complexity in real world problems

The standard treatment of problem complexity (e.g. Patton 2011) considers the interaction between two dimensions – the degree of (un)certainty about how to solve the problem and the degree of (dis)agreement amongst stakeholders about how to solve the problem. From these two dimensions, Patton plots the location of problems in 5 spaces: 1) simple, 2) socially complicated,

3) technically complicated, 4) complex and 5) chaotic. This treatment of complexity resonates with potential differences and disagreements between multiple stakeholders and scientific and knowledge uncertainty characterising complex real-world problems but does not address the layer of real-world complexity associated with evolving interconnected systems.

Arkesteijn et al. (2015) argued that complex problems, in addition to the dimensions of (dis)agreement and (un)certainty, have a third dimension. Namely, systemic stability, which they define as 'the historically grown mechanisms that provide stability for the current system and favour existing, undesirable but normalised social practices, and helps to explain why many interventions fail to improve the situation' (p. 100). This is similar to the concepts of 'lock-in' (Verbong and Geels, 2010) or from a positive perspective 'resilience', in that a highly resilient system tends to return to its original state after perturbation (Wilson 2014).

This is a logical way of thinking about complexity for real-world problems that involve 1) multiple affected stakeholders, who potentially disagree on values and end goals, and may also disagree on beliefs about how the world works (socially complex), 2) a degree of scientific uncertainty in that there are known unknowns that can be scientifically resolved (at the lower end of uncertainty, i.e. simple – if disagreement is low) or known unknowns the solution to which is less certain (more complex perhaps – technically complicated) and also potentially unknown unknowns (complex emergent properties) and 3) interventions required in a range of interacting systems that may vary in their degree of stability. Figure 1, from Arkesteijn et al. (2015) illustrates the concept.

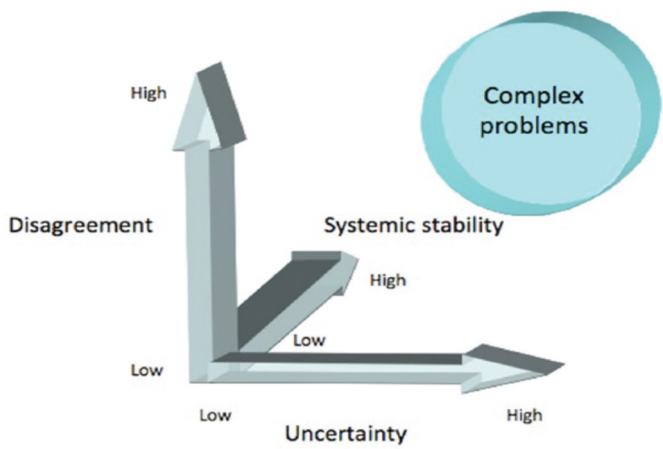

Figure 1. Source: Arkesteijn et al. (2015).

From this analysis of problem complexity, a simple three-item measure of project complexity was developed and used to assess the dimensional complexity of each project. The results of the five projects' complexity analyses are presented and implications of the analyses are noted for research project focus on achieving impact addressing real-world problems.

9. Project evaluation measures

The evaluation of whether the co-innovation projects achieved their end goals effectively, is problematic, in that the purposes/end goals/change ambitions (desired impact) of projects addressing complex real-world problems may be expected to take many years to achieve. As the Primary Innovation projects were just finishing at the time these data were collected, research impacts have not yet achieved full implementation. Therefore, the most robust project evaluation criterion, empirical measurement of impact by end-users, is not yet available. As a substitute, or proxy, for direct impact measurement with end-users we developed three subjective measures of project success assessed by research project team members:

(i) A *fitness of purpose* measure of project outputs
(ii) A *counterfactual* measure of probable project impacts
(iii) An *anticipatory impact* measure: the project teams' expectations that the projects' intended outcomes and impacts would be achieved

The development of the independent measurement variables and the validity criteria of dependent variables' measurements are described below in the methods section.

10. Methods

10.1. Overview and participants

For the three evaluation measures of effectiveness of the co-innovation projects in achieving their desired long-term impact (dependent variables' measures), the measure of the extent to which the five co-innovation principles were enacted and the measure of the emphasis placed on the co-innovation principles by the Reflexive Monitor (independent variables' measures), data were collected using an on-line survey questionnaire, hosted on the SurveyMonkey© platform. The survey was approved by the AgResearch Human Research Ethics process and consisted of 60 questions. After gaining participants' informed consent and ascertaining their innovation projects and roles, a series of questions examined the project teams' adherence to the five principles of co-innovation and the role of the reflexive monitor in promoting

the principles. Space was given for participants to make qualitative comments about any of the principles and their enactment in their particular project. The remaining questions comprised the three evaluation (independent variables) measures of project impact. The survey was completed by participants in the last month of the project – September 2017.

A total of 17 Primary Innovation participants (co-innovation project team members) across the five projects were invited to complete the questionnaire – 3 from Heifer Rearing, 3 from Nutrient Management, 2 from Water Use Efficiency, 3 from Log Segregation, and 5 from the Tomato Potato Psyllid project team. Sixteen of the 17 invitees responded. Participants spent an average of 30 minutes completing the survey, with a 100% completion rate.

The data for the problem complexity measure and the measure of absence/presence of i2S elements in each project were gathered at a day-long workshop in February 2017, seven months before the end of Primary Innovation. The same 17 participants invited to answer the Survey Monkey© co-innovation questionnaire attended this earlier workshop plus a few additional relevant persons taking the total sample size across the five groups up to 23. The complexity measure was passed out to all workshop participants as a 1-page paper-based questionnaire. With only three items, this was completed in a few minutes and collected immediately for later analysis.

Collecting the i2S data took the remainder of the workshop day. Members from each of the five project teams entered narrative data into a spreadsheet itemising the three domains and five questions and the various probes regarding the different elements of the i2S framework. Each group was assisted by a facilitator trained in the i2S framework. Effectively, the spreadsheet was a semi-structured interview schedule, which was jointly completed by each project team's members with guidance from the facilitators. The result of this data collection process was qualitative data describing the projects research process with respect to the domains, questions, and elements of the i2S framework.

10.2. Complexity measure

A 3-item measure for project complexity was developed from the Arkesteijn et al. (2015) model of complexity. One item for each of the three complexity dimensions. Each dimension was rated on a 12-point scale. Two items, scientific uncertainty and system stability were rated 1 = very low and 12 = very high. The third item, stakeholder disagreement was worded such that a very high degree of agreement was rated 1 and a very low degree of agreement was rated as 12. Therefore, all three items have the same directionality – the higher the score, the greater the contribution of that dimension to project problem complexity. The project complexity scale dimensions were scored by each project team in a consensus decision. The overall

complexity of a project was defined as the mean of the three-dimensional scores. Following Patton (2011), we categorise the scale's levels of complexity as 1–3 = simple problem, 4–6 = a complicated problem, 7–9 = a complex problem, and 10–12 = a chaotic problem (a seriously complex problem).

10.3. Five principles of co-innovation compliance measure

Participants were asked to rate the extent to which they agreed that each principle had been adhered to in their innovation project, with between five and seven question items per principle. Seven-point Likert scale response sets were used to measure project teams' adherence to each of the co-innovation principle question items (1 = strongly disagree to 7 = strongly agree). The ratings of each principle's question items were averaged to obtain a rating for each principle. These five principle ratings were averaged to obtain an overall rating for the five principles of co-innovation measure. Thus, each principle was unit weighted in the overall five principle rating. Table 1 presents the five principles and their respective question items.

10.4. Measure of reflexive monitors' focus on five principles of co-innovation

Five statements items were used to measure the degree (7-point Likert scale – 1 = strongly disagree to 7 = strongly agree) to which team members considered the Reflexive Monitors' focused on the five principles of co-innovation throughout their projects' lives. Table 2 presents the five question items in the Reflexive Monitor measure.

10.5. The i2S element measure

To collect the narrative i2S data for turning into a quantitative measure, we first deconstructed the domains, questions and probes of the i2S framework into discrete elements. We did this for the first four questions in each domain. We did not do this for the fifth question in each domain because these questions do not introduce new elements. Rather, the fifth questions were evaluative questions in which the project teams were asked about the degree to which they thought they had been successful in addressing the elements of the domain. Initially, we identified 59 elements across the three domains. However, after trialling the instrument we removed a number of elements either on the basis that they were not practical to answer or that they seemed to largely be repetitive of other identified elements. Thus, the instrument was refined to 43 elements across the three domains, with domain 1 containing 19 (ex 25), domain 2 containing 9 (ex 12) and domain three containing 15 (ex

Table 1. The five principles of co-innovation compliance measure: principles and scale items.

Principle	Focus of principle probes/question items
I. Involve partners and stakeholders	1) Identification of all relevant partners and stakeholders, 2) Identification of all the relevant science disciplines, 3) Building relationships between researchers, partners, and other stakeholders – respecting each other's culture, values, and goals, 4) Introducing new participants into the project as required, and 5) Partners and stakeholders involved with researchers in an on-going capacity throughout the life of the project
II. Take a problem focus	1) Clear identification and definition of the research problem at the beginning of the project, 2) Initial problem consultatively and inclusively defined and considered from all relevant disciplinary perspectives, 3) Stakeholders adequately and meaningfully included in the problem definition from an early stage, 4) Preparedness of the team to redefine the problem if/when required, 5) Specification of the problem was used to define the project approach (not vice versa).
III. Assemble the right team	1) Specification of the problem was used to determine an appropriate mixture of the relevant disciplines required in the project team, 2) The project team included adequate representation by members of relevant stakeholder groups affected by the problem and potential solutions, 3) The project team collaborated well as a group, with complementary skills, knowledge and personalities, 4) The project team had the necessary soft skills (e.g., facilitators, systems thinkers, science translators and brokers etc.) to be able to effectively bridge between researchers, partners, stakeholders and end-users, and 5) The team leader managed the project in a fair, open, and collaborative manner.
IV. Front up: Share results early and often	1) ,The team leader encouraged and facilitated good communication among the project team (disciplinary scientists, partners, and stakeholders), 2) During the project, communication between project team members (disciplinary scientists, partners, and stakeholders) was very good, 3) Throughout the course of the project open dialogue and discussion were encouraged and maintained between project team members (disciplinary scientists, partners, and stakeholders), 4) Information and progress reports were regularly disseminated to project team members (disciplinary scientists, partners, and stakeholders), and 5) There was an adequate plan developed and implemented for early and ongoing dissemination of project findings.

(*Continued*)

Table 1. (Continued).

Principle	Focus of principle probes/question items
V. Use the action learning cycle	1) The project team took an iterative and adaptive approach in order to involve research disciplines, partners, and stakeholders as and when needed during the project, 2) The project team regularly reflected on progress, considering how the research approach could be adaptively managed to maintain focus on the research problem and better achieve long-term project impact, 3) The project team regularly reflected on their collaborative interactions, considering how to modify behaviour and/or group processes to better achieve project impact, 4) The project team regularly reflected on its internal and external communication and information dissemination practices in order to enhance the sharing of results, 5) The project team reflected on and modified project activities based on feedback from sharing project results, 6) The project team regularly took time to apply the action learning/research cycle (plan, do, observe, reflect, plan . . .) throughout all phases of the project, 7) The project team regularly reflected on its practice of the principles of co-innovation, considering how to improve their application of the principles.

Table 2. Reflexive Monitors' focus on five principles of co-innovation measure: scale items.

Measure	Question Item
Reflexive monitor focus on co-innovation principles	1) Throughout the project the Reflexive Monitor placed strong focus on involving partners and stakeholders (i.e., ensuring disciplinary, partner, and stakeholder representation, inclusion, and participation), 2) Throughout the project the Reflexive Monitor encouraged the team to take a problem focus, 3) The project Reflexive Monitor helped to enhance team processes and positive interactions amongst team members (ensuring procedural fairness regarding contribution and influence, and helping build and maintain relationships between team members), 4) Throughout the project the Reflexive Monitor continuously encouraged early dialogue and discussion of project progress and results, and 5) Throughout the life of the project the Reflexive Monitor encouraged the project team to take a reflective, action learning approach to the research problem and solution generation.

22). Appendix A presents a table in which the three Domains, the four questions in each Domain, and the elements associated with each of the question probes are listed.

Three analysts independently assessed each project team's qualitative responses and rated the degree to which each of the 43 elements appeared to be present, or were addressed in the narrative descriptions of the research. The analysts used a 5-point scale with descriptive anchors at the scale mid-point and the two end-points where 0 = the element was absent or not addressed in the data describing the project, the mid-point of the scale 2 = the element was present or addressed to a moderate degree, and 4 = the element was strongly present or addressed to a high degree.

The three analysts then compared their ratings for each of the elements for each project. If there was more than 1 scale point difference between their ratings of an element, analysts went back to the original narrative data, and reached agreement (within 1 scale point) about that element. The final rating for each element was the mean of the three analysts' revised ratings. i2S domain scores for each project were calculated by averaging element scores within a domain. For each project, the three domain scores were averaged to create an overall i2S framework score (i.e., an overall rating of how well the project addressed the whole i2S framework). Figure 2 Illustrates the research process and methodology for the three research hypotheses.

10.6. Proxy measures of effectiveness of co-innovation projects in achieving desired impact

The latter portion of the survey collected the performance criteria used to assess the independent variables; participants were asked to evaluate the use of the co-innovation process in each of their projects. As previously noted, due to the fact that a true empirical measure of project impact from an end-user perspective was not available at the time of data collection, an outputs measure and two proxy impact measures were developed to provide the project teams' subjective assessments of the probable impact of each project. Question items in these three measures took the form of statements which participants were asked to indicate their level of agreement with using a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree).

For the first measure, consisting of five items, participants rated their project outputs to date, assessing their degree of fitness-for-purpose. The second measure, using seven items, asked participants to rate the extent to which they agreed that using a co-innovation approach enhanced project success over the counterfactual of not using a co-innovation approach. Participants were informed that 'This is a very subjective question and we are interested in your best guess/intuitive response. It may help if you try to compare your co-innovation project against another similar project where

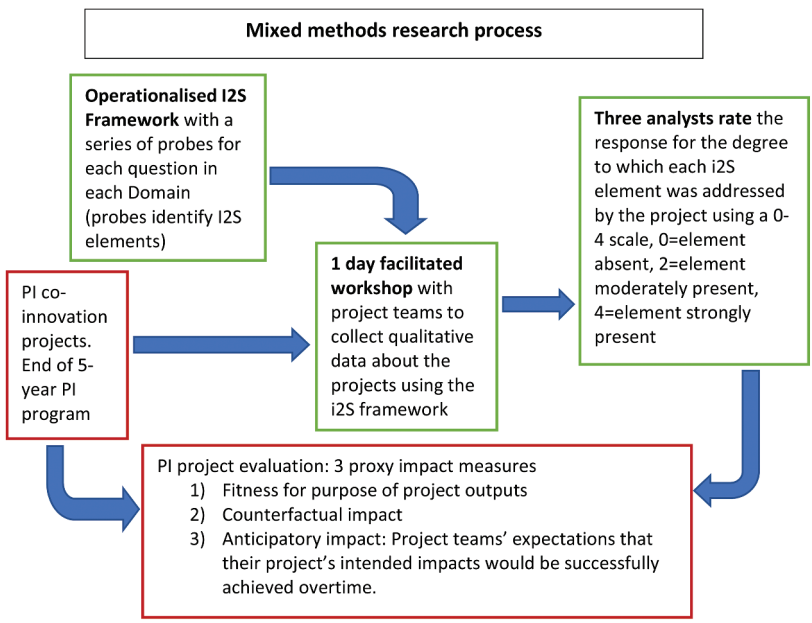

Figure 2. Process and methodology for addressing three study hypotheses.

the principles of co-innovation were not practiced.' For the final four item measure, participants were asked to look forward, and in lieu of elapsed time, estimate the success of project outcomes and impacts, according to the research problem they were addressing. Participants were informed that 'Often the elapse of a significant time period is necessary before introduced innovations or system interventions achieve their desired (long-term) impacts. In lieu of elapsed time, this proxy measure of project outcome/impact success is your subjective assessment (best guess/intuition) of the likely long-term outcome/impacts of the project.' They were asked to 'Think about the wider influence of the project's outputs and outcomes in regard to addressing the identified research problem.' For each of these three proxy impact measures, the overall measurement score was the mean of each measure's items' ratings. Table 3 presents the three proxy impact measures and their statement items.

11. Results and discussion

11.1. Project complexity

Table 4 presents the results of the project dimensional complexity analysis. All five projects have an overall complexity score ranging between 7.0 and 7.7 placing them just in the complex range. However, our analysis shows that the dimensional structure of complexity is different for each of the projects. The implication from this result is that dimensions of greatest complexity indicate a focal point for research emphasis for the creation of impact.

Scale: All three complexity dimensions rated on a 1–12 scale. Scientific uncertainty: 1 = very low scientific uncertainty, 12 = very high scientific uncertainty. Stakeholder disagreement: 1 = very low stakeholder disagreement, 12 = very high degree of stakeholder disagreement. Systemic lock-in: 1 = very low degree of systemic lock-in, 12 = very high degree of systemic lock-in. Overall complexity score = mean of three items.

For example, the TPP project has a high degree of scientific uncertainty but a moderate degree of stakeholder disagreement and systemic lock-in, which perhaps implies a need for the primary research focus for impact to be on resolving scientific problems in the project. In stark contrast, HR with a low scientific uncertainty score, a moderately high stakeholder disagreement score and a very high systemic lock-in score, has a quite different dimensional complexity structure. This structure implies that to achieve impact the research focus will need to be on resolving social differences and systemic implementation aspects. The WUE project exhibits a somewhat different structural dimensionality, with scientific uncertainty moderately high, stakeholder disagreement

Table 3. Question items for three proxy measures of project impact/success.

Proxy impact measure	Question/statement items for measure
I. Fitness of purpose measure of project outputs	1) Outputs which met policy, industry and end-user expectations of needs, 2) Outputs which achieved innovative solutions to the problem issue, 3) Outputs which achieved an evidence-based solution to the issue, 4) Outputs which achieved the immediate project goals (provided viable solutions to the problem issue), 5) Outputs that were fit-for-purpose for adoption by next and end-users.
II. Counterfactual measure of probable project impact	1) I believe that using a co-innovation approach enhanced the integration of local stakeholder knowledge throughout the project more than would otherwise have been the case, 2) I believe that using a co-innovation approach has resulted in better outputs, with greater fitness for purpose than would otherwise have been the case, 3) I believe that using a co-innovation approach has increased the likelihood of successfully achieving the project's desired long-term goal than would otherwise have been the case, 4) I believe that using a co-innovation approach has resulted in a more clearly identified implementation pathway than would otherwise have been the case, 5) I believe that using a co-innovation approach has resulted in a solution that is more readily acceptable to industry stakeholders than would otherwise have been the case 6) I believe that using a co-innovation approach has resulted in a solution that will be more acceptable to end-users (farmers, producers, growers) than would otherwise have been the case, and 7) I believe that using a co-innovation approach has resulted in a solution that will achieve greater adoption and implementation by end-users than would otherwise have been the case
III. Project teams' expectations that the projects' intended outcomes and impacts would be achieved	1) An adequate plan was developed, together by the project team (scientists, partners, stakeholder participants), for the next steps after project completion (e.g., a process for the dissemination and uptake of research findings by the appropriate actors, end-users and groups was outlined for implementation), 2) The outputs of the project defined improved practices, processes policies or technologies, the adoption of which could have a significant positive impact on the problem in the long term, 3) I believe that the outcome of this project will definitely be the adoption of improved practices and processes that have a significant long-term positive impact on solving the problem, and 4) New social networks were formed during the course of the co-innovation project, which added social capital and will strengthen future-sector network collaboration.

Table 4. Dimensional complexity analysis of five co-innovation projects.

Complexity dimension	HR	NM	WUE	TPP	LS
Scientific uncertainty	4.5	9	8	9.5	8
Stakeholder disagreement	8	7	4	6	7
Systemic lock-in	10	7	11	6	8
Overall Complexity score	7.5	7.7	7.7	7.2	7.0

quite low but very high systemic lock-in. Thus, in this case the primary focus for project impact should be on the science problem and the systemic implementation aspects.

This analysis suggests that the dimensional complexity structure of a real-world research problem is an important consideration to the research focus of the project. Figure 3 shows the overall project complexity, and the dimensional complexity structure of the five case study projects. The focus of the research process for real-world impact for each of the dimensions of structure and the overall degree of project complexity are also indicated.

11.2. Independent variables

11.2.1. Five principles of co-innovation measure

Table 5 presents participants' perceptions of the degree to which their project team adhered to each of the principles of co-innovation. Scores are averages of the question item responses (7-point Likert scales, 1 = strongly disagree, 7 = strongly agree) for each principle. The table also shows the average score for each project team across all five principles (i.e. that project team's five principles of co-innovation score) and the average five principles of co-innovation score across all five co-innovation projects (i.e., total average).

Across all five project teams the average of the five principles of co-innovation measures equalled 5.1, with individual project team's average scores ranging from a low score of 3.2 (TPP) to a high score of 6.3 (WUE). This spread indicates that the different project teams applied the various principles of co-innovation to different extents. This variation is useful for our purposes. The question becomes: is this variation in the dependent variable positively reflected in variation in the proxy impact criteria measures? Similarly, there was considerable variation in the ratings given to the different principles by the different project teams, ranging from a low of 2.6 (TPP, 4. Front up: Share results early and often) to a high of 6.8 (WUE, 4. Front up: Share results early and often).

We now consider the results for each of the five principles in each project.

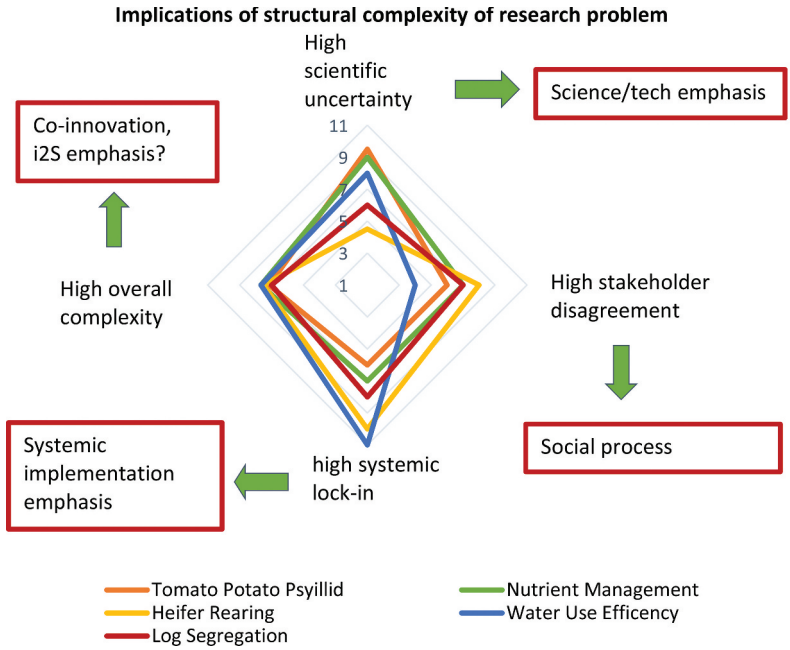

Figure 3. Structural complexity of primary innovation projects and suggested emphasis for different dimensional complexity.

Table 5. The extent to which participants agree that their project adhered to the five principles of co-innovation.

Principle	HR	NM	WUE	TPP	LS	Average
Principle 1: Involving partners and stakeholders	5.1	6.1	6.5	2.9	5.7	5.3
Principle 2: Taking a problem focus	6.1	5.9	5.2	3.8	5.7	5.3
Principle 3: Assembling the right team	4.2	6.4	6.4	3.4	5.2	5.1
Principle 4: Fronting up: Sharing results early and often	4.7	6.4	6.8	2.6	4.9	5.1
Principle 5: Using the action learning cycle	4.1	5.7	6.4	3.1	5.2	4.9
Average	4.8	6.1	6.3	3.2	5.3	5.1

Scale: 1 = strongly disagree, 4 = neither agree nor disagree, 7 = strongly agree

11.2.2. Principle 1: Involving partners and stakeholders

11.2.2.1. Quantitative data. All project members, excepting those from TPP, felt there was a relatively strong focus on identifying and involving all relevant stakeholders and science disciplines from the projects' beginnings (mean scores between 5 and 6.3). However, responses were stratified in regard to building stakeholder relationships, ongoing stakeholder involvement, and the introduction of new stakeholders when needed. This stratification saw WUE consistently scored as most positive, followed by NM and LS as moderately positive, and HR as less positive. Interestingly, this pattern was consistent with how participants scored their Reflexive Monitors' application

of this principle, suggesting participants considered ongoing adherence to principle 1 as the Reflexive Monitors' responsibility. The exception to this pattern was TPP, who scored their Reflexive Monitor as performing moderately well at focusing on this principle, despite scoring overall performance on principle 1 as poorest for all other questions. This indicates that considering principle 1 at the very beginning of the project is critical, and TPP did not have this opportunity as their project started before their involvement in Primary Innovation. This conclusion is consistent with the findings of Vereijssen et al. (2017) and Vereijssen, Williams et al. (2017b).

11.2.2.2. Qualitative data. Participants' comments suggested that project leader outlook and personality interactions in the project team strongly influenced whether this principle could be implemented and continue to be applied effectively throughout the course of the project '. . . attempts were futile as the Programme Leader just wanted to run [their] fundamental science project as per contract with MBI' and 'The various personalities of the project team did not always facilitate ongoing and full utilisation of co-innovation principles.'

11.2.3. Principle 2: Take a problem focus

11.2.3.1. Quantitative data. Although TPP was consistent in being scored lowest, HR and NM ratings of 'taking a problem focus' were high, and WUE was somewhat lower. The HR and NM project teams scored themselves highly on having a clearly identified and defined problem, which was consultatively defined and considered all disciplines. Conversely, WUE rated themselves only moderately on these aspects, but considered that the project team was prepared and able to redefine the problem to a much greater degree than HR and NM. Interestingly, however, WUE outperformed HR and NM in adaptive use of the principle 2, and in regard to scoring the Reflexive Monitor as encouraging the project team to implement the principle. These findings suggest that benefitting from the use of principle 2 is contingent on not only having a clear problem definition at the beginning, but also adaptive consideration of the problem. Vereijssen et al. (2017) and Coutts et al. (2017) also point to the need for adaptive problem definition in co-innovation projects.

11.2.3.2. Qualitative data. Politics between people appeared to be the most significant barrier to the successful and ongoing implementation of principle 2 'There is quite a degree of contested knowledge between growers and processors as this is the basis of their transactional relationship. Finding common ground can be challenging'. Pre-defined agendas, differing viewpoints and positions taken by organisations were mentioned as critical issues and barriers, which project teams raised 'Due to externally imposed resource constraints and team dynamics utilisation of co-innovation principles was

greater in the earlier parts as compared to the latter parts of the project.' Similar issues and barriers to co-innovation were also noted by Pinxterhuis et al. (2018).

11.2.4. Principle 3: Assembling the right team

11.2.4.1. Quantitative data. While both the NM and WUE project teams rated their performance on principle 3 as relatively high and consistent across the questions, the HR project team demonstrated greater variation. While HR considered that the project team included an adequately diverse group of stakeholders, as a group, they rated collaborative performance as low in regard to having complementary skills, knowledge and personalities in the team. While the LS team was also relatively consistent across principle 3, they noted their performance in regard to having the necessary social process skills for bridging stakeholder relationships to be weaker. This suggests there is a complex multidimensionality to assembling the right team, from diverse stakeholders and disciplines, to complimentary personalities, skills and effective team leadership.

11.2.4.2. Qualitative data. Several participants commented that their project team was unbalanced in some regard, such as having too many technical skills or too many social process skills 'We probably could have involved more facilitators and systems thinkers into the conversation, e.g., members of the Value Chain Optimisation team at Scion. The project team is probably biased towards technical skills' and 'The project team was too heavily slanted to soft skills rather than having technical experts in the focus area.'

11.2.5. Principle 4: Fronting up: Sharing results early and often

11.2.5.1. Quantitative data. Efforts of the project leader to encourage and facilitate good communication among the project team appeared to be an important factor in determining how much communication occurred in reality. Interestingly, there was a large amount of variation in how participants' rated team leaders' attitudes towards good communication, with between 2 and 6 points of variation, in three of the projects. Regular dissemination of information and progress reports was rated at a maximum (7) for NM and WUE, while TPP participants disagreed that there was regular information sharing, and LS was neutral. These findings did not map directly onto whether participants felt there was an adequate plan developed for early and ongoing dissemination of project findings. That is, LS rated this question relatively high, while information sharing in the project team was not seen as regular. This suggests that two separate plans may be needed for internal sharing within the project team, and external information sharing with stakeholders.

11.2.5.2. Qualitative data. Based on participants' comments, a barrier to sharing information among and outside of the project team was the perception that findings other than those ready for implementation were not useful to share. Researchers were described as reluctant to share results before they are definite, and this was recognised as an issue when early engagement and feedback were critical components within co-innovation 'Researchers are in general hesitant to bring out results while doing the work, because preliminary results do not tell the whole story, and messages may change.' One research participant noted that it is very important stakeholders understand that 'untested knowledge holds risks if acted on, [and] this must be off-set against the value of early dialogue/debate of early results'.

11.2.6. Principle 5: Using the action learning cycle

The seven questions for principle 5 were distributed throughout the survey, as using the action learning cycle entails promoting reflexivity under all co-innovation principles. Furthermore, the project team needed to take time to apply action learning throughout the entire project, and regularly reflect on their practice of the principles, while considering how to improve their application of the principles.

11.2.6.1. Quantitative data. The WUE, NM, and LS project teams (in that order) showed the greatest focus on reflection and using the action learning cycle. HR was relatively neutral with, on average, participants neither agreeing nor disagreeing with the statement while TPP participants slightly disagreed that their team used the action learning cycle. This pattern was reflected in both the statements considering the action learning cycle and reflection with respect to each of the other four principles of co-innovation and the more general statements of reflection regarding the entire project, the teams' overall practice of the principles of co-innovation, and preparedness to modify project activities on the basis of reflection.

11.2.6.2. Qualitative data. Participants considered that the action learning cycle was easier to consciously put into practice if co-innovation was embedded in the project from the beginning and if the science team leader was supportive of the concept 'Much easier to do this if this culture is embedded in a project from the start. Must have support from the Project Leader to enable this to work.' These qualitative findings are generally supported by the quantitative results. The TPP project, with the lowest rating on this principle, started before becoming involved in the Primary Innovation programme and co-innovation did not enjoy the full support of the science team leader. The other project, which started before becoming involved with Primary Innovation, the LS project, was rated lower on this aspect than NM, and WUE but higher than HR. We note that, although HR was originally

designed as a co-innovation project, the co-innovation component did not enjoy the full support of the science team leader, whereas LS, although not originally designed as a co-innovation project, did experience full support of the science team leader.

11.3. Reflexive monitors support for co-innovation principles

Under each of the co-innovation principles, participants were asked to evaluate the degree to which their Reflexive Monitors encouraged the project team to implement the principle in practice. Table 6 displays the average score that each project team's Reflexive Monitor was given across each of the five principles, the average score of the project teams across the five principles, and the overall average score of Reflexive Monitors across all principles and projects.

Total scores of the extent to which the Reflexive Monitors encouraged implementation of the principles of co-innovation ranged from 3.1 (HR) to 6.2 (WUE) with the average of the Reflexive Monitors' total scores being 4.9. Interestingly, the ratings given to Reflexive Monitors regarding their encouragement of each of the co-innovation principles maps relatively closely onto the scores given for overall performance on the co-innovation principles.

However, several key differences emerge; TPP felt their Reflexive Monitor encouraged the co-innovation principles to a larger degree than they were adhered to generally. These ratings appeared to be an acknowledgement that the programme had already been operating when Primary Innovation became involved, meaning embedding the principles post-hoc was going to be difficult. Furthermore, a project member commented that, as the programme leader was not on-board with co-innovation, much of the Reflexive Monitor's efforts 'were futile'.

On the other hand, HR felt their Reflexive Monitor encouraged the co-innovation principles to a lesser degree than they were adhered to generally. In particular, the project team felt that the Reflexive Monitor did not encourage implementation of principle 3, assembling the right project team. They felt their Reflexive Monitor encouraged implementation of principle 2, take a

Table 6. The extent to which participants' agree that their Reflexive Monitor encouraged the project team to implement each co-innovation principle in practice.

Principle of co-innovation	HR	NM	WUE	TPP	LS	Average
Principle 1: Involving partners and stakeholders	3.3	5.3	6.5	4.8	5.7	5.1
Principle 2: Taking a problem focus	4.0	5.0	6.5	4.2	5.3	5.0
Principle 3: Assembling the right team	2.3	6.3	6.0	4.0	5.7	4.9
Principle 4: Fronting up: Sharing results early and often	3.0	4.7	6.0	4.4	5.7	4.8
Principle 5: Using the action learning cycle	3.0	5.3	6.0	4.0	5.0	4.7
Average (= RM's total score)	3.1	5.3	6.2	4.3	5.5	4.9

Scale: 1 = strongly disagree, 4 = neither agree nor disagree, 7 = strongly agree.

problem focus, to a larger degree, which was reflected in their overall adherence to the principle as a team. However, the project leader had reservations regarding the co-innovation approach, and early in the project, a personality conflict surfaced between the Reflexive Monitor and the team leader. This conflict remained unresolved at project completion and may help explain why this Reflexive Monitor despite, according to one participant, having "more of a focus on Primary Innovation project outputs than on this project's work" received the lowest score (3.1) of all the Reflexive Monitors for encouraging the team to practice the five principles of co-innovation.

There were evidently many different perspectives about the role that Reflexive Monitors were expected to play, and the roles that they occupied in reality. One project team felt the Reflexive Monitors' role was as a process coach, and to convince the team about the value of co-innovation. Another felt it was the project team's role to co-support their Reflexive Monitor in aiding the team to implement the principles effectively. One project team felt it was their Reflexive Monitor's role to aid in conflict resolution to contribute 'to a fair and positive working environment', while another felt it was their Reflexive Monitor's responsibility to share results, to engage with stakeholders and gain feedback. Finally, one project team felt it was their Reflexive Monitor's role to look out for 'blind spots' and provide a warning.

While some project teams were generally positive about their Reflexive Monitor, others appeared to have very high expectations about their performance across a large range of areas, such as being able to relate well to every member of the project team. This implies that clearer roles and responsibilities are needed surrounding Reflexive Monitors, and the project team needs to understand the role they play in also adhering to the principles of co-innovation (see also: Fielke et al. 2017). Both quantitative data and qualitative comments were consistent with the proposition that a supportive project team leader enabled the Reflexive Monitors to carry out their role responsibilities more effectively.

11.4. i2S element presence measure

Table 7 presents results of the analysis of the five co-innovation projects through the lens of the i2S framework. Table 7 shows, for each co-innovation project, the average score for each domain and the average of the three

Table 7. Absence or presence of i2S elements in each domain and overall framework.

Domain/overall framework	HR	NM	WUE	TPP	LS
Average domain 1 (19 elements)	2.7	3.1	3.0	1.3	2.7
Average domain 2 (9 elements)	1.7	2.9	2.2	1.4	2.3
Average domain 3 (15 elements)	2.6	3.3	3.4	1.5	2.6
Average of three domains	2.3	3.1	2.9	1.4	2.5

domain scores. These latter scores are the overall i2S framework scores for the projects. Note that this method of calculating the project i2S framework scores unit weights the three domains for each project.

Note: Scores range from 0 to 4, 0 = absence of evidence for consideration of elements, 2 = moderate evidence for consideration of elements, 4 = strong evidence for consideration of elements

TPP received the weakest ratings across all three domains and the overall framework, followed by HR then LS. NM received the highest ratings across the framework followed closely by WUE. Higher ratings indicate that a project showed evidence of having considered more of the elements of the i2S framework during the project research phase. Somewhat similarly to the project teams' application of the five principles of co-innovation and the Reflexive Monitors' focus on the five principles, we hypothesised that because the framework was specifically designed to help produce research that is fit for implementation, greater consideration of the domains and elements of the i2S framework would lead to better real-world impact.

11.5. Project impact evaluations – dependent variable measures

As previously noted, there were three dependent variable measures developed for research impact. Table 8 presents the results regarding participants' evaluations of their co-innovation projects' outputs. Participants were asked to rate their level of agreement or disagreement on a 7-point Likert scale with 5 statement items, where 1 = strongly disagree, 4 = neutral, and 7 = strongly agree.

Regarding project outputs, participants were asked to rate how successful their project was in delivering project outputs that met policy, industry and end-user expectations of needs. They were also asked how successful the project was in achieving an innovative, evidence-based solution, and in achieving its immediate goals. Finally, participants rated the overall success of the project in producing outputs that were fit for purpose and for adoption by end-users.

Table 8. Evaluation of co-innovation projects' outputs' fitness for purpose.

Output 'fitness for purpose' measure items	HR	NM	WUE	TPP	LS	Mean
Project outputs met policy/industry/end-user expectations and needs	3.0	6.0	6.5	3.4	5.3	4.8
Project achieved an innovative solution to the problem	5.0	6.3	6.0	3.2	6.0	5.3
Project achieved an evidence-based solution to the problem	4.7	6.7	6.5	3.2	6.0	5.4
Project achieved its immediate goals (i.e. viable solutions to research problem)	2.7	6.7	6.5	3.8	5.7	5.1
Project produced outputs fit-for-purpose for adoption by end-users	4.0	5.7	6.5	3.0	4.7	4.8
Mean (i.e. scale score)	3.9	6.3	6.4	3.3	5.5	5.1

Scale: 1 = strongly disagree, 4 = neither agree nor disagree, 7 = strongly agree

Overall, participants were mostly positive about their project delivering innovative and evidence-based solutions to the problem. Interestingly, participants felt slightly less confident that their projects were successful in producing outputs that were fit for purpose for adoption by end users. The HR project team felt they were relatively unsuccessful at delivering on the project's immediate goals of delivering outputs that were viable solutions. The rationale provided for this was that the time-intensive nature of the process of co-innovation slowed down the delivery of the immediate outputs. They were also less confident that their project successfully delivered outputs that met policy, industry and end-user expectations and needs.

11.6. Project counterfactual (proxy) impact evaluation: Using co-innovation versus alternative approaches

For this 'counterfactual' evaluation, participants were asked a set of questions about their project processes, outputs, outcomes and impacts, **if a co-innovation approach had *not* been used in their project**. That is, did they believe that using a co-innovation approach improved the project more than an alternative method would have? Aspects considered were integration of local stakeholder knowledge, degree to which outputs were fit for purpose, identification of an implementation pathway and acceptability of the end solution. Table 9 presents the results of the counterfactual evaluation for the five projects.

Project mean scores for the counterfactual impact measure ranged from 4.1 (neither agree nor disagree) to 6.2 (reasonably strong agreement). The TPP and HR had the lowest scale mean scores while NM and WUE had the highest.

Table 9. Counterfactual (proxy) impact evaluation of the five co-innovation projects.

Counterfactual measure items	HR	NM	WUE	TPP	LS	Mean
Co-innovation enhanced integration of local stakeholder knowledge	5.3	6.0	5.5	5.6	5.7	5.6
Co-innovation resulted in better outputs, with greater fitness for purpose	3.3	6.7	6.0	4.4	5.7	5.2
Co-innovation increased likelihood of achieving desired long-term impact	4.0	6.7	6.0	3.6	5.0	5.1
Co-innovation resulted in a more clearly identified implementation pathway	3.0	6.3	5.5	4.4	4.3	4.7
Co-innovation resulted in a solution more acceptable to industry stakeholders	4.7	5.7	6.0	3.8	6.7	5.4
Co-innovation led to a solution more acceptable to end-users (farmers, producers, growers)	5.5	6.0	6.0	3.4	6.7	5.5
Co-innovation led to a solution that will achieve greater adoption and implementation by end-users	4.0	6.3	6.0	3.6	6.0	5.2
Mean (counterfactual score for each project)	4.3	6.2	5.9	4.1	5.7	5.2

Scale: 1 = strongly disagree, 4 = neither agree nor disagree, 7 = strongly agree

Across the case study projects, the mean value of the counterfactual measure was 5.2 – indicating that, generally, participants believed that using a co-innovation approach would create greater impact than if it had not been used.

Participants were most positive about 'using a co-innovation approach enhanced the integration of local stakeholder knowledge throughout the projects, than would otherwise have been the case' (mean = 5.6). Similarly, participants were unanimous in agreeing to a moderate to high degree that co-innovation resulted in a solution that is more acceptable to industry stakeholders (mean = 5.5) and end-users (mean = 5.4), than would otherwise have been the case.

In regard to creating a solution that will achieve greater adoption and implementation by end-users, NM, LS and WUE were all in strong agreement (mean scores between 6 and 7). However, TPP (mean = 3.6) and HR (mean = 4.0) were less convinced. Overall, participants were slightly less positive that a co-innovation approach had resulted in a more clearly defined implementation pathway than would otherwise have been the case (mean across case study project = 4.7).

As is well documented in the literature, the down-sides of using a co-innovation approach are the time and energy-intensive processes required. Participants also noted that a dedicated Reflexive Monitor and project leader are needed, to 'make it happen' in practice. One project team found co-innovation slowed every step down, particularly in regard to disseminating knowledge, as internal review processes did not acknowledge non-scientific contributions. Additionally, the use of co-innovation principles was found to increase the number and type of stakeholders and engagement methods used in the research process, resulting in increased management and time requirements (Vereijssen et al. 2017). Thompson et al. (2017, p. 34) in an analysis of scientist and stakeholder perspectives of transdisciplinary research claimed that 'The most widely identified pragmatic concern was a perception of the process being highly time-consuming and resource-intensive.' Similarly, Turner et al. (2020b) estimated that working as a multi-disciplinary team had an extra 40% time overhead. For further reflections on the operationalisation of the i2S framework and the process of conducting this transdisciplinary research project, see: Robson-Williams et al. (2019; 2020a, 2020b).

Participants had mixed perspectives about what co-innovation provided as a methodology. One participant felt co-innovation provided 'structured guidance/checklist[s] to gauge progress throughout the life of project towards achieving its desired outcomes'. Another felt that co-innovation provided more structure and linked the implementation pathway from activities, to outcomes and impacts. Another participant commented that co-innovation 'opened the eyes' of the project team to some degree, but did not fundamentally affect the implementation pathway or delivery of outputs.

Finally, one participant believed that the critical value of co-innovation was engaging stakeholders regarding the problem and solution, not the project contract and milestones.

11.7. Anticipatory (proxy) evaluation of co-innovation projects' impacts

As insufficient time had passed for the full impacts of the Primary Innovation projects to be realised, research participants were asked to estimate the likely long-term outcomes and impacts of their projects. This included improvements in practices, processes, policies or technologies that could have significant impact on the problem in the long term. It also included the adoption of these improved processes and practices, and the formation of new social networks that may strengthen future collaboration potential. Finally, participants were asked whether they had or were likely to develop a plan, following project completion, for disseminating and promoting uptake of the research findings.

Participants were moderately positive in anticipating that the outputs of their project would definitely improve practices, processes, policies or technologies, and the adoption of these would have significant positive impacts in the long term. They were also moderately positive that the new social networks that were formed through their projects would add social capital and strengthen future, sector collaborative work. While NM, WUE, and LS felt relatively sure that their projects would achieve adoption of improved practices and processes, HR was less certain, with two thirds of the participants from the project team indicating that they disagreed. An HR participant's comment suggested this was a contextual issue due to organisational restructuring that was preventing a plan being created for use of the research findings. One TPP project participant indicated that their similarly low score on estimated adoption was because the project was focused on knowledge generation that would then be developed and lead to greater impact. Securing ongoing funding to ensure project outcomes were achieved and extended was an important factor mentioned by several participants. Table 10 presents the results of the anticipatory impact evaluation of the five projects

11.8. Correlations between the independent variable measures' scores

We examined the correlations between the three independent variables. The five principles score had a strong but non-significant correlation of 0.6 ($p = 0.3$) with the Reflexive Monitors score – note that a significant role of the Reflexive Monitors was to encourage the team to enact the five principles. Interestingly, a very high and significant correlation of 0.98 ($p = 0.003$) was found between the five principles score and the i2S framework score. This reflects the earlier noted similarities between the concept of co-innovation

Table 10. Anticipatory (proxy) impact evaluations of the five co-innovation projects.

Impact expectation measure items	HR	NM	WUE	TPP	LS	Mean
Adequate plan co-developed for project solution implementation	4.0	6.0	6.0	3.0	4.0	4.6
Adoption of research outputs (practices, policies, processes, technologies) could have positive long- term impact on problem	5.0	6.3	6.0	3.6	6.3	5.5
Outcome will definitely be adoption of practices or processes with positive long-term impacts on problem	3.3	6.3	6.0	3.2	6.0	5.0
New social networks will increase sector social capital and strengthen future sector collaborations	5.0	5.7	6.5	4.6	6.0	5.6
Mean	4.3	6.1	6.1	3.6	5.6	5.1

Scale: 1 = strongly disagree, 4 = neither agree nor disagree, 7 = strongly agree.

and the i2S framework and suggests that when teams enact the co-innovation principles they are also likely to be considering the elements of the i2S framework. There was a non-significant correlation of 0.53 ($p = 0.35$) between the Reflexive Monitor score and the i2S framework score.

11.9. Correlations between the dependent variables' measurement scores

Next, we examined the correlations between the three dependent variables. As all three dependent variable measures (output fitness score, counterfactual score, and anticipatory impact score) were designed to be project evaluation measures, we expected them all to be reasonably strongly correlated. In fact, all three correlations were greater than 0.98 ($p = 0.003$).

11.9.1. Comparison of independent and dependent measures' overall scores for the five projects

From Table 11 it may be observed that project teams that achieved higher scores on the independent variables also tended to have higher scores on the dependent variables. This result suggests that compliance with the five

Table 11. Comparison of independent[1] and dependent[2] measures' overall scores for the five projects.

Measure	Project teams' overall scores on each measure				
	HR	NM	WUE	TPP	LS
Five principles[1]	4.9	6.1	6.3	3.2	5.3
Reflexive monitor focus[1]	3.1	5.3	6.2	4.3	5.5
i2S framework[1]	2.3	3.1	2.9	1.4	2.5
Outputs fit for purpose[2]	3.9	6.3	6.4	3.3	5.5
Counterfactual[2]	4.3	6.2	5.9	4.1	5.7
Outcome/impact[2]	4.3	6.1	6.1	3.6	5.6

Note: with the exception of the i2S framework score, which used a 0–4 scale, where 0 = absence of evidence of elements' consideration, 2 = moderate evidence for consideration of elements, 4 = strong evidence for consideration of elements, all other measures used a 1–7 Likert scale where 1 represents a very poor evaluation and 7 represents an excellent evaluation.

principles, considering the i2S framework domains and elements, and using a Reflexive Monitor, may be useful practices for enhancing real-world impact from research.

11.10. Correlational analysis between independent and dependent variable measurement scores

Finally, Table 12 presents the correlations (and their p values) between the independent variable measurements and the dependent variable measurements. All correlations between the three independent variables and the three dependent variables are positive and large, though not necessarily significant at the 0.05 level. Each row of the table addresses one of our hypotheses.

Our first alternate hypothesis was: **The stronger the practice of the principles of co-innovation in a project the greater the probability of the project achieving its desired long-term impact**. The five principles compliance measure correlates positively and significantly ($p < 0.05$) with all three evaluation criteria and with the mean of the three evaluation measures. Thus, the null hypothesis is rejected at the 0.05 level suggesting our results are consistent with the alternative hypothesis.

Our second alternate hypothesis was: **The greater a Reflexive Monitor's focus on ensuring the project practiced the principles of co-innovation, the greater the probability of the project achieving its desired long-term impact**. While all the correlations between the Reflexive Monitor scores and the three evaluation measures and their mean were positive and large, none of them were significant at the $p < 0.05$ level. Therefore, the null hypothesis cannot be rejected and our second alternative hypothesis is not supported at the 0.05 level by the current data set.

Our third alternate hypothesis was: **The more a project considers/addresses/applies the domains and elements of the i2S framework, the greater the probability of the project achieving its desired long-term impact**. The i2S framework score correlated positively and significantly ($p < 0.05$) with all three evaluation criteria and with the mean of the three

Table 12. Correlations between independent and dependent variables' measurement scores.

Variables	Outputs evaluation score	Counterfactual evaluation score	Anticipatory Outcome evaluation score	Mean evaluation score
Five principles score	0.93 (p = 0.02)*	0.88 (p = 0.048)*	0.95 (p = 0.01)*	0.93 (p = 0.02)*
Reflexive monitor score	0.83 (p = 0.08)	0.83 (p = 0.08)	0.80 (p = 0.10)	0.82 (p = 0.09)
i2S elements score	0.91 (p = 0.030)*	0.88 (p = 0.047)*	0.94 (p = 0.020)*	0.89 (p = 0.018)*

* correlation significant at $p < 0.05$.

evaluation measures. Thus, the null hypothesis is rejected at the 0.05 level suggesting our results are consistent with the alternative hypothesis. Thus, at the arbitrary $p < 0.05$ level of statistical significance, our correlational results strongly support hypotheses one and hypothesis three, however hypothesis two is rejected. If the significance level had been set at the less stringent, $p < 0.1$, all four of hypothesis two's correlations would have been significant and null hypothesis two would have been rejected and the alternative hypothesis supported.

12. Study limitations

This study has several limitations that need to be considered whilst interpreting the results. Case study sample size – due to the small number of case studies involved (n = 5) in the correlational analyses testing the hypotheses (Table 12) there was very low statistical power to find a real effect at the 0.05 level. Due to the relationship between sample size, statistical significance, effect size, and statistical power, to achieve a statistically significant result for our sample size, a correlation (or effect size) of at least 0.88 is required. In the behavioural sciences, a correlation of 0.2 is considered a small effect, 0.5 is considered a medium effect, and 0.8 is considered a large effect size (Cohen 1988). Thus, all point correlations in Table 12 are large effects (in the case of the non-significant correlations, the large confidence interval around the point correlation includes 0), in the cases where significance was reached, the point size effects are very large. Nonetheless, with a sample size of n = 5, we are very reluctant to draw any generalisations, but rather consider that our results are suggestive of relationships, and act as positive encouragement for more research along these lines.

A second limitation is that there were a small number of respondents (from a small population pool) from each project who were making the assessments of the independent and dependent variables. Across the five case studies and the six variable measures, participants numbered between 2-7. With these small sample sizes psychometric analysis of the variable measures would be inappropriate. Because such cases studies often only have a few people in them, obtaining large enough sample sizes is difficult. Integration of results (e.g. meta-analysis) from multiple repetitions of small sample size projects like this might help develop reliable and valid measures.

The dimensional structure of complexity scale was a three-item measure with only one item to measure each of the three dimensions. More items to measure each dimension would improve the reliability of the measure, and consequently, its potential validity.

As has been previously noted, due to the time lag between project completion and project impact, the three dependent variable impact measures are not measures of empirical impact by end-users – the most

appropriate type of impact measure. One was an output fitness for purpose measure while the other two were subjective proxies impact measures – a counterfactual impact measure and an anticipatory impact measure. Additionally, the participants in this research were members of the project team, rather than end-users of the project outputs, the most appropriate group for assessing real-world impacts. Since these are subjective rather than empirical measures – they may be subject to individual bias on the part of the participants. To help overcome these limitations an important next step is to conduct evaluations of the projects' impacts with end-users after a suitable period of time has elapsed from project completion. Such a study could also confirm whether or not research project team members are able to judge the likely effectiveness and impact of their research outputs.

A fifth potential limitation is in the operationalisation of the i2S framework and the identification of the framework elements. This was primarily developed by the first two authors (with some guidance from Gabriele Bammer, the creator of the i2S framework – though any problems with this deconstruction are the responsibility of the current authors). Although we adapted our original element identification, on reflection, as a result of our experience of using the current instrument (Appendix A) to collect the data presented above, we consider that considerable improvement may be possible.

Finally, the translation of narrative descriptions of the projects into quantitative scores for the absence or presence of i2S elements was a subjective process and as such was susceptible to individual bias. Additionally, although the original i2S narrative data was collected with the aid of facilitators knowledgeable about the i2S framework, the validity of quantitative ratings was limited by the quality of the narrative data. To help counter the potential of individual bias, we used an inter-subjective process with three rating judges as previously described.

13. Conclusions

First, we have demonstrated that different real-world problems may be complex to differing degrees along three structural dimensions of complexity. From this empirical result, we have analytically inferred that problems with different structural dimensionality may require different degrees of focus on different aspects of the research process to create the desired long-term impacts (see Figure 3).

Second, our results supported alternate hypothesis one and are suggestive of a positive relationship between project team compliance with the five principles of co-innovation and success of a project in terms of our proxy measures of impact; output's fitness for purpose, the counterfactual, and project team anticipation of impact.

Third, our results supported alternate hypothesis three and are suggestive of a positive relationship between consideration of the i2S framework domains and elements, and three of the project evaluation measures: output's fitness for purpose, the counterfactual, and the project teams' anticipation of impact.

Our results did not support hypothesis two that the Reflexive Monitors focus on the five principles of co-innovation would be positively associated with project impact. However, we note the large positive correlations between Reflexive Monitors' focus on the principles and the proxy impact evaluations and speculate that the lack of statistical significance could be a study artefact due to the very small sample size.

We note that these results must be interpreted with considerable caution as outlined in the study limitations. With the current data limitations regarding sample size, wide confidence intervals around the correlation effect size, and low statistical power, we emphasise that these results must only be considered provisional. However, we believe that the implication of our results is that the hypotheses considered in this paper are worthy of further empirical investigation to better understand the relationship between the independent variables measured and the dependent impact.

We acknowledge that the best dependent variable to test the efficacy of research impact from a co-innovation or an i2S approach would be an empirical measure of impact, as judged by end-users, taken after a suitable temporal lapse. However, we believe that the degree of consistency between our three dependent measures is indicative of the validity of our proxy approach to impact measurement.

Finally, we hope that our operationalisation of the i2S framework and our methodological approach to the assessment of impact from different research approaches provides an example of one way of testing the impact efficacy of research approaches and perhaps an inkling of assurance for researchers using co-innovation approaches or using the i2S framework to address real-world problems.

Acknowledgments

We would like to acknowledge the Ministry of Business, Innovation and Employment and Our Land and Water National Science Challenge for funding this research. We would also like to thank Gabriele Bammer for support and encouragement. A special thanks is due to the many research participants who responded to surveys and/or who graciously gave their time in personal interviews. Lastly, we wish to thank Susanna Finlay-Smits, Margaret Brown and the annonymous reviewers for their careful critical reviews of the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Funding for this research was provided by two Ministry of Business, Innovation and Employment grants, the Collaboration Laboratory project of the Our Land and Water National Science Challenge, MBIE contract [C10X1507] and the Primary Innovation project [CONT-30071-BITR-AGR]. While providing funding, the Ministry of Business, Innovation and Employment did not play any role in the study design, collection, analysis and interpretation of the data and in the writing of the report and the decision to submit for publication.

ORCID

James A. Turner http://orcid.org/0000-0003-4396-3280

References

Arkesteijn, M., van Mierlo, B., & Leeuwis, C. (2015). The need for reflexive evaluation approaches in development cooperation. *Evaluation*, *21*(1), 99–115. https://doi.org/10.1177/1356389014564719

Bammer, G. (2013). *Disciplining interdisciplinarity: Integration and implementation sciences for researching complex real-world problems*. Australian National University E Press, The Australian National University, Canberra. Open access online http://press.anu.edu.au?p=222171 .

Beechener, E. S. (2020). *How co-innovation anticipates scaling: The modulating function of a co-innovation space*. Thesis submitted in fulfilment of the requirements for a PhD in Science and Technology Studies at the University of Edinburgh. University of Edinburgh, Edinburgh, Scotland. http://dx.doi.org/10.7488/era/448

Beechener, S. (2017). How co-innovation anticipates scaling: Does putting co-innovation into practice support up/out-scaling in the agri-food sector. *Webinar presentation to Primary Innovation programme 6th July 2017*. https://www.youtube.com/watch?v=BomSe-GIO7s&list=PLGJuUj2ZWqonmmOKjVgujhzbjdxh4VQon&index=3

Beers, P., van Mierlo, B., & Hoes, A.-C. (2016). Toward an integrative perspective on social learning in system innovation initiatives. *Ecology and Society*, *21*(1), 33. http://dx.doi.org/10.5751/ES-08148-210133

Bitzer, V., & Bijman, J. (2015). From innovation to co-innovation? An exploration of African agrifood chains. *British Food Journal*, *117*(8), 2182–2199. https://doi.org/10.1108/BFJ-12-2014-0103

Botha, N., Coutts, J., Turner, J. A., White, T., & Williams, T. (2017). Evaluating for learning and accountability in system innovation: Incorporating reflexivity in a logical framework. *Outlook on Agriculture*, *46*(2), 154–160. https://doi.org/10.1177/0030727017707406

Botha, N., Klerkx, L., Small, B., & Turner, J. A. (2014). Lessons on transdisciplinary research in a co-innovation programme in the New Zealand agricultural sector. *Outlook on Agriculture*, *43*(3), 219–223. https://doi.org/10.5367/oa.2014.0175

Botha, N., Turner, J. A., Fielke, S., & Klerkx, L. (2017). Using a co-innovation approach to support innovation and learning: Cross-cutting observations from different settings and emergent issues. *Outlook on Agriculture*, *46*(2), 87–91. https://doi.org/10.1177/0030727017707403

Boyce, W., Percy, H., Turner, J. A., Fear, A., Mills, T., & Craven, C. (2016, September). *Building Co-innovation into your research proposal: What is Co-innovation and How to Apply it to Research Proposals*. *Internal AgResearch report produced for Primary Innovation*. AgResearch, Hamilton, New Zealand. Retrieved March 11, 2019, from http://www.beyondresults.co.nz/SuccessPrinciples/Documents/Guide%20to%20Co-innovation%20FINAL.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). L Erlbaum Associates.

Coutts, J., White, T., Blackett, P., Rijswijk, K., Bewsell, D., Park, N., Turner, J. A., & Botha, N. (2017). Evaluating a space for co-innovation: Practical application of nine principles for co-innovation in five innovation projects. *Outlook on Agriculture*, *46*(2), 99–107. https://doi.org/10.1177/0030727017708453

Fielke, S., Nelson, T., Blackett, P., Bewsell, D., Bayne, K., Park, N., Rijswijk, K., & Small, B. (2017). Hitting the bullseye: Learning to become a reflexive monitor in New Zealand. *Outlook on Agriculture*, *46*(2), 117–124. https://doi.org/10.1177/0030727017708490

Fielke, S. J., Botha, N., Reid, J., Gray, D., Blackett, P., Park, N., & Williams, T. (2018). Lessons for co-innovation in agricultural innovation systems: A multiple case study analysis and a conceptual model. *The Journal of Agricultural Education and Extension*, *24*(1), 9–27. https://doi.org/10.1080/1389224X.2017.1394885

Ingram, J., Gaskell, P., Mills, J., & Dwyer, J. (2020). How do we enact co-innovation with stakeholders in agricultural research projects? Managing the complex interplay between contextual and facilitation processes. *Journal of Rural Studies*, *78 (August 2020)*, 65–77. https://doi.org/10.1016/j.jrurstud.2020.06.003

Kilelu, C. W., Klerkx, L., & Leeuwis, C. (2013). Unravelling the role of innovation platforms in supporting co-evolution of innovation: Contributions and tensions in a smallholder dairy development programme. *Agricultural Systems*, *118 (June 2013)*, 65–77. http://dx.doi.org/10.1016/j.agsy.2013.03.003

King, B. (2017). *Final report: Reach and impact of primary innovation: Social network analysis longitudinal analysis 2014, 2015, and 2017 (Researchers)*. Internal Report to Primary Innovation Programme.

King, B., Fielke, S., Bayne, K., Klerkx, L., & Nettle, R. (2019). Navigating shades of social capital and trust to leverage opportunities for rural innovation. *Journal of Rural Studies*, *68 (May 2019)*, 123–134. https://doi.org/10.1016/j.jrurstud.2019.02.003

Klerkx, L., van Mierlo, B., & Leeuwis, C. (2012). Evolution of systems approaches to agricultural innovation: Concepts, analysis and interventions. In I. Darnhofer, D. Gibbon, & B. Didieu (Eds.), *Farming Systems research into the 21st century: The new dynamic* (pp. 457–483). Springer.

Klerkx, L., Aarts, N., & Leeuwis, C. (2010). Adaptive management in agricultural innovation systems: The interactions between innovation networks and their environment. *Agricultural Systems*, *103*(6), 390–400. https://doi.org/10.1016/j.agsy.2010.03.012

Klerkx, L., & Gildemacher, P. (2012). The role of innovation brokers in agricultural innovation systems. In *Agricultural innovation systems: An investment sourcebook, module 3* (pp. 221–230). World Bank, Washington DC, United States.

Lee, S. M., Olson, D. L., & Trini, S. (2012). Co-innovation, convergenomics, collaboration and co-creation for organisational values. *Management Decision*, *50*(5), 817–831.

Moore, J. (2019). Bridging the gap between research and implementation – A personal view. *NZ Journal of Forestry*, *63*(4): 36-39. http://www.nzjf.org.nz/abstract.php?volume_issue=j63_4&first_page=36

Nederlof, S., Wongtschowski, M., & Van Der Lee, F. (2011). *Putting heads together. Agricultural innovation platforms in practice. Bulletin 396, development, policy and practice*. KIT. Retrieved March 24, 2014, from https://www.lmpublishers.nl/assests/PuttingHeads.pdf

Patton, M. Q. (2011). *Developmental evaluation. Applying complex concepts to enhance innovation and use*. The Guilford Press.

Payne, T. (ed.) (2017). *Reflexive monitoring in New Zealand: A guide for reflexive monitors*. AgResearch Internal Report. AgResearch, Hamilton, New Zealand.

Pinxterhuis, I., Dirks, S., Bewsell, D., Edwards, P., Brazendale, R., & Turner, J. A. (2018). Co-innovation to improve profit and environmental performance of dairy farm systems in New Zealand. *Rural Extension & Innovation Systems Journal*, *14*(2), 23–33. https://search.informit.org/doi/epdf/10.3316/informit.248755871113556

Rijswijk, K., Bewsell, D., Small, B., & Blackett, P. (2015). Reflexive monitoring in New Zealand: Evaluation lessons in supporting transformative change. *Evaluation Journal of Australasia 15 (4):* 38–43. https://doi.org/10.1177/1035719X1501500405

Robson-Williams, M., Small, B., & Robson-Williams, R. (2020a). Chapter 8 in Interdisciplinary and Transdisciplinary Failures: Lessons Learned from Cautionary Tales. In D. Fam & M. O'Rourke (Eds.), *A week in the life of a transdisciplinary researcher: Failures in research to support policy for water quality management in New Zealand's South Island*. (pp. 131-146) Routledge. https://doi.org/10.4324/9780367207045

Robson-Williams, M., Small, B., & Robson-Williams, R. (2019). Evaluating the Integration and Implementation Sciences Framework. *ITD 2019 Joining forces for Change Conference*, International Science Council, Gothenberg, 10-13th September.

Robson-Williams, M., Small, B., & Robson-Williams, R. (2020b). The importance of asking the right questions for transdisciplinary research design: Lessons from the Selwyn-Waihora project. *GAIA - Ecological Perspectives for Science and Society*, *29*(3), 170–175. https://doi.org/10.14512/gaia.29.3.7

Srinivasan, M. S., Bewsell, D., Jongmans, C., & Elley, G. (2017). Just-in-case to justified irrigation: Applying co-innovation principles to irrigation water management. *Outlook on Agriculture*, *46*(2), 138–145. https://doi.org/10.1177/0030727017708491

Srinivasan, M. S., Bewsell, D., Jongmans, C., & Elley, G. (2019). Research idea to science for impact: Tracing the significant moments in an innovation based irrigation study. *Agricultural Water Management*, *212*(1), 181–192. https://doi.org/10.1016/j.agwat.2018.08.045

Thompson, M. A., Owen, S., Lindsay, J. M., Leonard, G. S., & Cronin, S. J. (2017). Scientists and stakeholder perspectives of transdisciplinary research: Early attitudes, expectations, and tensions. *Environmental Science and Policy*, *74 (August 2017)*, 30–39. https://doi.org/10.1016/j.envsci.2017.04.006

Turner, J. A., Allen, W., Fraser, C., Fenemor, A., Horita, A., White, T., Chen, L., Atkinson, M., & Rush, M. (2020b). Navigating institutional challenges: Design to enable community participation in social learning for freshwater planning. *Environmental Management*, *65 (March 2020)*, 288–305. https://doi.org/10.1007/s00267-020-01256-x

Turner, J. A., & Botha, N. (2012). *Co-learning and co-innovation to achieve impact – A new era in technology transfer for New Zealand's biological industries*. AgResearch MBIE Proposal.

Turner, J. A., Horita, A., Fielke, S., Klerkx, L., Blackett, P., Bewsell, D., Small, B., & Boyce, W. M. (2020a). Revealing power dynamics and staging conflicts in agricultural system transitions: Case studies of innovation platforms in New Zealand. *Journal of Rural Studies*, *76 (May 2020)*, 152–162. https://doi.org/10.1016/j.jrurstud.2020.04.022

Turner, J. A., Klerkx, L., Rijswijk, K., Williams, T., & Barnard, T. (2016). Systemic problems affecting co-innovation in the New Zealand agricultural innovation system: Identification of blocking mechanisms and underlying institutional logics. *NJAS - Wageningen Journal of Life Sciences*, *76 (March 2016)*, 99–112. https://doi.org/10.1016/j.njas.2015.12.001

Turner, J. A., Klerkx, L., White, T., Nelson, T., Everett-Hincks, J., Mackay, A., & Botha, N. (2017). Unpacking systemic innovation capacity as strategic ambidexterity: How projects dynamically configure capabilities for agricultural innovation. *Land Use Policy*, *68 (November 2017)*, 503–523. https://doi.org/10.1016/j.landusepol.2017.07.054

Turner, J. A., Williams, T., Nicholas, G., Rijswijk, F. J., Barnard, K., Beechener, T., & Horita, A. (2017b). Triggering system innovation in agricultural innovation systems: Initial insights from a community of change in New Zealand. *Outlook on Agriculture*, *46*(2), 125–130. https://doi.org/10.1177/0030727017708500

van Mierlo, B., Arkesteijn, M., & Leeuwis, C. (2010). Enhancing the reflexivity of system innovation projects with system analyses. *American Journal of Evaluation*, *31*(2), 143–161. https://doi.org/10.1177/1098214010366046

van Mierlo, B., Regeer, B., van Amstel, M., Arkesteijn, M., Beekman, V. (2010b). *Reflexive monitoring in action. A guide for monitoring system innovation projects*. Communication and Innovation Studies, WUR; Athena Institute, VU.

Verbong, G. P., & Geels, F. W. (2010). Exploring sustainability transitions in the electricity sector with socio-technical pathways. *Technological Forecasting and Social Change*, *77*(8), 1214–1221. https://doi.org/10.1016/j.techfore.2010.04.008

Vereijssen, J., Srinivasan, M. S., Dirks, S., Fielke, S., Jongmans, C., Agnew, N. M., Klerkx, L., Pinxterhuis, I., Moore, J., Edwards, P., Brazendale, R., Botha, N., & Turner, J. A. (2017). Addressing complex challenges using a co-innovation approach: Lessons from five case studies in the New Zealand primary sector. *Outlook on Agriculture*, *46*(2), 108–116. https://doi.org/10.1177/0030727017712321

Vereijssen, J., Williams, T. A., Park, N. M., Nielsen, M.-C., & Agnew, N. M. (2017b). Evaluating co-innovation principles in a fundamental bioprotection research programme addressing challenges to potato production. *New Zealand Plant Protection*, *70 (2017)*, 16–24. https://doi.org/10.30843/nzpp.2017.70.20

Wieczorek, A. J., & Hekkert, M. P. (2012). Systemic instruments for systemic innovation problems: A framework for policy makers and innovation scholars. *Science and Public Policy*, *39*(1), 74–87. https://doi.org/10.1093/scipol/scr008

Wilson, G. A. (2014). Community resilience: Path dependency, lock-in effects and transitional ruptures. *Journal of Environmental Planning and Management*, *57*(1), 1–26. https://doi.org/10.1080/09640568.2012.741519

Appendix A. The i2S element measure: Domains, questions, and elements

Domain	Question	Element
I. Synthesising disciplinary and stakeholder knowledge	(1) What was the synthesis of disciplinary and stakeholder knowledge aiming to achieve and who is intended to benefit?	1. Purpose of combining disciplinary and SH knowledge documented, 2. Disciplinary and SH beneficiaries documented
	2) Which disciplinary and stakeholder knowledge was considered?	3. Systems tools used to help define the 'problem' and investigate links between problem elements 4. A recognised method used to develop a system view 5. Identification of different views on defining the problem 6. Identification of all the disciplines that could be relevant 7. Identification of all the stakeholders that could be relevant 8. A recognised methodology used for identification of relevant stakeholders and disciplines 9. Relevant science disciplines included and clarity on how decisions were made on what research was undertaken 10. Relevant stakeholders included and clarity on how decisions were made on which SH were considered relevant to include 11. All the relevant perspectives within each stakeholder group identified 12. Consideration given to how the problem was described or framed 13. Differences or conflicts managed
	3) How was the disciplinary and stakeholder knowledge synthesised, by whom, and when?	14. Knowledge from the stakeholder and science participants synthesised using either formal or informal methods 15. Consideration given to who was the most appropriate person (s) to lead the integration and synthesis 16. Synthesising of knowledge (stakeholder with stakeholder, discipline with discipline, stakeholder with disciplinary) occurs at multiple stages in project
	4) What circumstances influenced the synthesis of disciplinary and stakeholder knowledge?	17. Overall context (and changes in context) influencing research project taken into account 18. Consideration given to the authorisation and legitimacy of the project 19. Project barriers and facilitators managed

(*Continued*)

(Continued).

Domain	Question	Element
II. Understanding and managing diverse unknowns	1) What was the understanding and management of diverse unknowns trying to achieve?	20. Clarity within research team, research stakeholders and target audience on why managing for unknowns is important
	2) Which unknowns/uncertainties/ risks considered?	21. Systematic consideration of unknowns (at outset and on-going through project life)
		22. Opportunities and risks associated with undertaking the project managed
	3) How were the recognised unknowns and uncertainties managed or responded to?	23. Tools or processes used to help cope with uncertainty in research or implementation process (For example, the precautionary principle, scenarios, sensitivity analysis, hedging, adaptive management (acceptance approach))
		24. Identification and management of unknowns that were irreducible within the scope of the project
		25. Surprises or the unexpected managed
		26. Consideration of participants' tacit knowledge (researchers, research stakeholders and target audience) and management of any conflicts
		27. Communication of uncertainty in research outputs and implications for policy or practice change
	4) What circumstances influenced the management of unknowns	28. Management of project barriers and facilitators to understanding and managing unknowns within research and target audience organisations

(*Continued*)

(Continued).

Domain	Question	Element
(I) Providing integrated research support for policy and practice change	1) What is the integrated research support aiming to achieve, and who is intended to benefit?	29.Practice or policy change intent of the project clearly stated
		30. Target audience (e.g., next-user, end-user) government, business, or civil society clearly stated
	2) Which aspects of policy and practice are targeted by the provision of integrated research support?	31. Consideration given to the target change as part of a system (e.g., if it was a change of regional council policy, was the way the RC makes policy decisions considered, position in the policy cycle, reaction of lobby groups and community groups)
		32. Identification of different systems views
		33. Consideration given to how change happens – explicit or implicit (e.g., programme logic, causal chain modelling, theory of change, etc)
		34. Identification of all the opportunities and target audiences relevant for the targeted change
		35. Consideration of how research findings were framed or described for target audience
		36. Identified differences or conflicts in values between research team and target audience managed
	3) How was integrated research support provided, by whom, and when?	37. Consideration of what methods used to deliver research findings to the target audience
		38. Consideration given to who was involved in the delivery of the research findings
		39. Interaction with the target audience occurs at multiple stages in project
	4) What circumstances might/did influence the provision of integrated research support for policy and practice change?	40. Overall context (and changes in context) influencing target change and target audience managed
		41. Consideration given to the authorisation and legitimacy of the project to interact with target audience
		42. Identification and management of difficulties in moving from doing the research (domain 1) in to trying to implement (domain 3)
		43. Project barriers and facilitators with target audience managed