Evaluating a Quality Improvement Collaborative: A hybrid approach

Abstract

Purpose:

Quality Improvement Collaboratives (QICs) are a popular approach to improving healthcare services and patient outcomes. This paper evaluates a QIC implemented by a large, integrated healthcare organisation in Wales in the UK.

Design/Methodology/Approach:

This evaluation study draws on two well-established evaluation frameworks: Kirkpatrick's approach to gather data on participant satisfaction and learning, and Stake's approach to gather data and form judgements about the impact of the intervention. A mixed methods approach was taken which included documentary analysis, surveys, semi-structured interviews, and observation of the QIC programme.

Findings: Together the two frameworks provide a rounded interpretation of the extent to which the QIC intervention was fit-for-purpose. Broadly the evaluation of the QIC was positive with some areas of improvement identified.

Originality: Evaluation studies are limited on QICs and if present tend to adopt one framework. Given the complexities of undertaking quality improvement within healthcare this study uniquely takes a hybrid approach.

Research limitations/implications: This study is limited to a QIC conducted within one organisation. Further testing of the hybrid framework is needed that extends to different designs of QICs.

Practical implications: A hybrid framework is provided to assist those charged with designing and evaluating quality improvement collaboratives.

Keywords: Quality Improvement, Collaboratives, Breakthrough Collaboratives, Evaluation, Kirkpatrick, Stake

Introduction

In response to national efforts to improve the quality of care, policymakers and healthcare leaders have increasingly turned to Quality Improvement Collaboratives (QICs) as an efficient approach to improve provider practices and patient outcomes. QICs can be loosely defined as multi professional teams coming together from multiple organisations over a 9–18-month period to improve performance on a given topic supported by improvement experts (Kilo, 1998). Common components of QICs can be identified, however, including in person learning sessions, Plan-Do-Study-Act cycles (PDSAs), multi-disciplinary quality improvement (QI) teams, and data collection for QI (Nadeem et al., 2013). However, use of the term Collaboratives often hides confusion about form and focus, such that judgements about effectiveness are often difficult to make.

Collaboration and co-production emphasise the role of inclusivity and meaningful partnerships between all stakeholders (e.g., Rycroft-Malone et al., 2016; Cooke et al., 2017). Looking at the origins of co-production, Osborne and Strokosch (2013) identify two, often unconnected, areas of work – one from public administration and management (PAM) and the other notably from service management theory. Elinor Ostrom's (1972) seminal work is cited as the originator of co-production from a PAM perspective. From a service management perspective, which regards coproduction as an essential and inalienable core component of service delivery (Osborne et al., 2016), this can be linked to the work of Normann (1991) who refers to the 'moments of truth' in service provision. Collaborative models also appear within other theoretical perspectives, e.g., sociology of education (see Vygotsky, 1978) on legitimate peripheral participation; Wenger (1998) on communities of practice; and Engestrom (1987) on activity theory).

Alongside the concept of collaboration equal significance has grown around QI, as over the last two decades interest and use of QI methodologies has emerged in healthcare (Burgess and Radnor, 2013). They have been built on practices first used in industry (Deming 1986; Ishakawa 1985; Juran 1951, 1964) and transferred and adopted into healthcare settings by for example, Institute for Healthcare Improvement (IHI) in the USA under the leadership of Berwick (1989). Early limited success associated with single organisational initiatives led to the development of collaborative models such as the Breakthrough Series at IHI (Bataldan, 2003). Typically, a QIC includes groups of professionals coming together either from within an organisation or across multiple organisations, to learn from and motivate each other to improve the quality of health services (de Silvo, 2014). Although many variations have subsequently emerged, they infrequently report limited success (Garcia-Elorria et al., 2019; Nadeem et al., 2013; Solberg 2005).

Outside of the formal narrative of improvement collaboratives, there have been studies which have considered multisite improvements programmes which have not been referred to as collaboratives yet still offer some learning. For example, the Matching Michigan, a patient safety programme, which aimed to reduce central line infections in over 200 intensive care units in England (Bion et al., 2013). Dixon-Woods et al's (2013) review of this programme reported that its impact may have been limited due to how the programme was designed and executed in the UK. For example, it was not an exact replica (i.e., did not include all elements of the original Matching Michigan programme) of the original programme, and the outer and inner contexts strongly modified the programme's effects. The team concluded there was a need to distinguish between technical interventions from implementation strategies, and to ensure careful consideration is given to the contexts in which improvement programmes are introduced (Dixon-Woods, et al., 2013). Similarly, the evaluation studies of the Health Foundation's collaborative programme, the Safer Patients Initiative (SPI1), found there was improvement in one clinical process and one measure of staff perceptions of organisational climate. When comparing several clinical processes of the four SPI1 sites with the 18 control hospitals that were not specific SPI1 yet might have been expected to improvement, there was no significant difference between two groups over time (Benning et al., 2011).

This paper aims to report on the evaluation of a pilot Collaborative designed largely on the IHI model which was implemented within one organisation. The QIC ran during 2019 -2020 in Hywel Dda University Health Board located in South West Wales, UK. The decision to offer the programme emerged from the Health Board's Quality Improvement Framework and there appeared to have been a smooth shift from the strategy, through policy, to implementation. The QIC involved a substantial investment of time and people. The curriculum was based on a similar programme offered elsewhere in Wales and was deemed therefore to have been successfully trialled. A large investment in time, money and reputation was signed off by the Health Board in the summer of 2018, which provided the go ahead for the development and planning of the QIC programme which started in June 2019. A structured nine-month programme of team activities included the recruitment of eleven teams (8 members per team) from across the Health Board brought together so that each could undertake an improvement project within their clinical area. A formal selection process took place where teams were required to complete an application form to join the QIC.

The evaluation was underpinned by two frameworks. Firstly, the programme team were using Kirkpatrick and Kirkpatrick's (2006) model with particular emphasis on levels 1 and 2 to assess satisfaction and learning by participants, i.e., their focus was on the first of the focus areas described above. It therefore seemed appropriate for the evaluation team to utilise Kirkpatrick to focus on levels 3 and 4 (behaviour change and impact of the interventions). However, Kirkpatrick is vague about how behaviour and impact can be assessed, and therefore a second framework devised by Stake (2010) was employed to enable the context and mechanisms associated with the Collaborative to be judged. This paper focuses on the way this second framework (Stake 2010) guided and enabled the evaluation team to address questions such as:

- Can we form judgements about behaviour and impact change as an outcome of engaging in a Collaborative?
- What lessons could be learned for future Collaborative programmes?
- Is the Stake framework useful as a guide to evaluating a Collaborative?

The remainder of this paper is organised as follows. First, we review the limited literature on QICs, followed by an explanation of the theoretical underpinning of the evaluation models employed for the evaluation. We then provide details of the approach taken and the methods of data collection employed. The findings of the evaluation are presented followed by a discussion in relation to the learning garnered from the local intervention and the wider lessons learnt for the use and design of QI collaboratives. Finally, conclusions are drawn which include limitations of the study and areas of further research.

Quality Improvement Collaboratives

Improvement collaboratives, first developed in the US, have become an international phenomenon with models being implemented in countries such as Australia, France, the Netherlands, Norway, Sweden and the UK (Wilson et al 2003). As noted earlier, many of the models used are derived from the IHI's Breakthrough Series (Kilo, 1998; Kilo 1999), which typically involves eight steps (see table 1) or communities of practice (De Silva, 2014). In some instances, it is difficult to assess what model has been employed, as we have seen the original breakthrough model evolve to include additional days at the start of the collaborative, increasing emphasis on teamwork and multi-disciplinary/ multi-agency teams including users/patients and carers. De Silva (2014) provides a taxonomy of characteristics of collaboratives which she suggests may influence the extent to which they are

successful. Such characteristics include who is involved, span of the collaborative (e.g., national or within organisations) design of collaborative (e.g., duration, frequency and mode of contact) and subject of collaborative (e.g., disease specific). The primary rationale underpinning collaboratives is that by collaborating and comparing practice, professionals and teams are motivated to do things differently, which leads to an improvement in services and patient outcomes (de Silva, 2014).

Insert table 1 about here

Wilson et al., (2003) conducted a review of Collaborative projects across seven countries and identified seven components that were deemed important determinants of success: sponsorship; topic/area of focus; ideas for improvement, participants; senior leadership support; preliminary work and strategies for learning about and making improvements. The authors report considerable variation in how these components were implemented. A more recent study which reports on the effects of a quality improvement collaborative to reduce stillbirths and neonatal mortality in Indian hospitals also reports on the programme not being implemented as intended (Zamboni et al., 2021).

An evidence scan conducted by de Silva (2014) found 2 systematic reviews, five randomised trials and 167 other studies which examined the effects of collaboratives. Of these studies, 33% of the trials and reviews and 72% of the other studies found improved patient outcomes, and 100% of trials and 89% of the other studies identified improvements in service use and costs, although the author provides a cautionary note that the numbers reviewed were small. Other observations about QICs from the scan include the difficulties associated with attributing the effects to collaboratives due to other interventions often being implemented at the same time and the research methods not always being robust. The geographic and policy context surrounding the running of collaboratives may also impact on the level of success (de Silva, 2014).

Systematic reviews of QICs have also suggested ways of strengthening the collaborative model (Zamboni et al., 2020; Garcia-Elorria et al., 2019; Nadeem et al., 2013; Hulscher et al., 2013; Solberg 2005). For example, Zamboni et al., (2020, pg 17) maintain that the role of context and mechanisms of change are underexplored and state "*adequacy and appropriateness of external support, functionality of QI teams, leadership characteristics and alignment with national systems and priorities may influence outcomes of QI Collaboratives but the strength and quality of the evidence is weak".* They maintain that facility size, voluntary or compulsory participation, baseline performance and factors related to health facility readiness may all be significant. Whereas Solberg (2005, pg 198) says that *"without involved leadership the organisational changes are likely to be less important and long lasting; a process-focused improvement effort will not occur so all that happens are some targeted efforts led by enthusiastic champions without much authority".*

Many of the collaborative studies are based on the IHI Breakthrough Series model with focus on a specified topic, facilitated by clinical experts and experts in QI, include a structured learning programme, involve multi professional teams and use rapid cycles of testing change. The QIC that is the subject of this evaluation departs from the need to have a specified topic. Participants came from across all four sites/three counties of the Health Board and represented a wide array of specialisms including multi-site areas such as communication with patients.

More broadly Dixon-woods et al (2012) remind us of the challenges associated with improving quality in healthcare. The authors draw key lessons from the Health Foundation's programme evaluations and other relevant studies. It would seem that all ten challenges are relevant to the design and implementation of QICs and helpfully the authors suggest how the challenges might be addressed. The challenges are categorised in three areas: designing and planning improvement interventions, organisational and institutional contexts, professions and leadership; and beyond the intervention – sustainability, spread and unintended consequences. These are all areas that should be continually reviewed as QIC teams and participants embark and continue with their improvement

efforts. Table 2 summarises the key attributes and possible enablers and challenges associated with QICs.

Insert table 2 about here

Theoretical Underpinning of the Evaluation approach

Very few collaboratives include a formal evaluation and even fewer use a designated model to guide the process. This paper shares the results of an evaluation study conducted alongside a QIC which took a hybrid approach using two evaluation models. There were two distinct but complementary foci for the evaluation. Firstly, it sought to establish the extent to which the QI programme was successful in meeting its objectives of providing support to the teams as they developed and implemented their understanding of QI tools and techniques. The evaluation explored what was intended when the programme was devised and whether those intentions were met. This is important because the degree to which the programme met (and continued to meet) its aims influenced the second aspect of the evaluation. An accompanying part of the evaluation examined the extent to which the programme has had and continues to have a discernible impact on the performance of individuals undertaking the learning intervention and whether the improvement projects delivered their desired outcomes.

In considering the challenges of evaluation of a complex initiative we reviewed the circumstances and the work that had been initiated by the QIC project team before deciding which of the several models available might be of greatest use. we undertook a review of the literature on evaluation techniques in order to establish if, and how many, evaluations were conducted using an established framework (Authors, 2021). Our conclusion was that Kirkpatrick and Kirkpatrick (2006), while universally acclaimed, fell short of providing an adequate procedure for comprehensive judgements to be made. Since this model is so widespread in use, we suggested that it could be retained as an ordering devise then complemented by an additional framework that focused on the context within which learning takes place. We cited Stufflebeam (2001) Donabedian (2005) and Stake (2010) as providing such foci. To evaluate the QIC which forms this case study we decided to use Stake's (2010) model as a complement to Kirkpatrick and Kirkpatrick (2006).

The evaluation model that has external recognition and dominance is that attributed to Kirkpatrick and Kirkpatrick (2006) and it could be argued that this 4 level framework (see figure 1) is a necessary component of a thorough evaluation although not of itself sufficient (Authors, 2021). The model is intuitively understandable and well regarded in practice for that reason, but it is sketchy on process and thus problematic in use.

Insert Figure 1 here

The first two Kirkpatrick levels can be addressed adequately; the third with minor challenge, but the fourth level of analysis presents significant problems for evaluators. To assess results or impact of a learning intervention requires isolating effects evidentially attributable to the programme while disregarding those resulting from parallel socio-political-financial changes happening alongside.

To evaluate levels 3 and 4, an additional framework was required, that could offer greater guidance on how to analyse data to judge behaviours and post-programme impact. Stake's (2010) model offers a lens to assess detailed pathways and provides three dimensional descriptions and judgements about the provenance, delivery and outcomes of a learning intervention. Figure 2 below gives a schema of the framework. On the extreme left is a box marked "Rationale" and it is here that the purpose and form of the intervention is defined. Thereafter the framework has two distinct

aspects to it – on the one hand it is about describing the intervention under scrutiny, on the other hand it is about judging the effectiveness of the intervention.

Insert Figure. 2 here

Data can be collected using conventional methods such as questionnaires, interviews, observations. Stake advocates a naturalistic approach designed to enable a holistic picture to evolve so would advise against putting too much structure around data gathering, preferring a conversational, informal approach with the testing and challenging of hypotheses.

The model does not look at individual performance, instead emphasis on individual differences among participants gives way to attention to the contingencies (strengths, gaps, relationship) among background conditions, environmental activities, and learning outcomes. The model looks at the congruence (links) between what was intended to happen and what actually happened, with a view to deciding whether there were any gaps that might prevent maximal outcome.

Every programme has its own rationale, though often this is implicit, indicating the philosophic background and basic purposes of the intervention and providing a basis for evaluating its intent. Questions can be asked about whether the plan developed by the designers / deliverers constitutes a logical step in the implementation of the basic purposes. The rationale also is of value in choosing the reference groups which later are to be used to formulate judgment on various aspects of the programme. Because often implicit, a statement of rationale can be difficult to obtain.

The bridge between the intentions of the programme and the actuality of its delivery, form the countenances of the framework. These are the gaps between what was intended to happen and what actually did happen. Data collection should focus on the antecedents, the transactions, and the outcomes and questioning should seek to shed light on the countenances of each from the perspective of the various stakeholders. Analysis should consist of comparing the varying accounts alongside the external comparator, be that similar interventions or published standards.

Whether goals are local or national, the measurement of excellence requires explicit rather than implicit standards. How standards are defined and used will frequently be a matter of serendipity; if publicly available standards are available then they would constitute best practice but using comparable interventions as comparators may be the only recourse. There are two principal ways of processing descriptive evaluation data: finding the contingencies among antecedents, transactions, and outcomes and finding the congruence between intents and observations. The data for an intervention is congruent if what was intended actually happens. To be fully congruent the intended antecedents, transactions, and outcomes would have to come to pass.

Methodology

This mixed methods evaluation of the QIC involved Kirkpatrick and Kirkpatrick's (2006) approach to gathering data on participant satisfaction and learning, and Stake's (2010) framework to gather data and form judgements about the impact of the intervention (see figure 3).

Insert figure 3 here

Participants

All participants that took part in this evaluation either participated in the QIC or were sponsors involved in the design and running of the QIC. Tables 3 and 4 show the number of participants that completed the survey and participated in the interviews. All members of the QIC were provided

with details of the evaluation at one of the QIC programme days and consent was sought before any interviews were conducted.

Insert tables 3 and 4 about here

Data collection
Data collection was achieved by:

- Scrutiny of related documents (e.g., quality and safety strategy, quality improvement framework, QIC team application forms) where available
- Online surveys to gather baseline from participants
- Semi-structured interviews with participants, sponsors and stakeholders
- Observation and opportunistic informal conversations with participants at programme days

Details of the QIC programme are shown in Figure 4.

Insert figure 4 about here

An initial online survey was designed for, and distributed to, 88 participants (11 teams) before the first workshop, to provide a broad understanding of their profile and experience, as well as their understanding and expectations of the programme. An additional online survey was planned for after the final celebration event, designed to provide evidence of whether expectations had been fulfilled, and what benefit had been gained from the intervention.

Semi-structured interviews were undertaken during the programme with sponsors (n=3) participants (n=17) and coaches (n=5). Observations (20 hours) of the programme were made by members of the evaluation team (LC, SW, AW). Discussions were held with stakeholders and members of the programme team (MD, SH and DB) in relation to the design of the programme and how this had evolved. With the consent of participants, all interviews were recorded and transcribed with permission of those involved.

Data analysis

Descriptive statistics were used to analyse the survey data. These are reported as part of the evaluation. Data from the interviews were coded in two ways. First, we analysed all the interviews using inductive thematic analysis (Braun and Clarke, 2006) to understand the generic feedback from all interviewees. Secondly, we coded the results according to the key elements of the evaluation framework using deductive thematic analysis. For both analyses items were coded independently by two researchers (LC and SW) and differences were discussed regularly to form an agreement. For accuracy themes were cross-checked by a third member of the evaluation team (AW).

Impact of Covid-19 on the QIC programme and evaluation design
As the QIC programme progressed the evaluation team attended and observed three workshops to test understanding and findings and gain valuable feedback on the programme. Part of the evaluation design was to
revisit the teams after a period of time to assess the impact and sustainable nature of the programme and projects. The taught element of the programme concluded in February 2020 and the intention of the programme team had been to run a celebration event in March 2020. At that point external factors intervened, culminating in the covid crisis and its necessary response. The evaluation team were fortunate in retaining access to the programme team and were able to interview a limited number of stakeholders and participants in September 2020, making it possible

to seek understanding of the impact of the pandemic and the extent to which the programme had helped participants to cope with the additional stresses experienced in the latter part of 2020.

The original intention post-programme was to explore the impact of the QIC programme, but by September 2020 circumstances were very different from those that had been envisaged during the design phase of the evaluation. We therefore altered the focus of our questioning to probe how the experience of the programme had made a difference to participants in terms of their capacity to cope with covid pressures. It was intended to interview at least one participant from each of the eleven projects that formed the core of the intervention, but in the event, this proved not to be possible. Second wave covid priorities meant that few participants were available for interview, but it was possible to speak with five, covering very different projects, and feel confident that this was a representative sample. Each interview was conducted over Zoom and lasted for 30-40 minutes. With the consent of participants, the interviews were recorded and then transcribed. An online celebration for the programme was held in November 2020, which provided an update on the improvement projects and reflections from participants on their learning during and post the taught programme.

Results

The results from this mixed-methods evaluation are presented as the data were collected: online survey, telephone interviews conducted during and post the QIC progamme. To understand the design and context of the collaborative and the eleven selected projects, the QIC team application forms were reviewed by the evaluation team, along with the design of QIC programme (see Figure 4).

Online Survey

An initial online survey was distributed to all 88 QIC participants. This was to provide a profile of the participants and their initial understanding of the QIC programme and QI experience. Although not mandated, 100% of participants completed the survey (see supplementary materials).
Questions probed the background of participants. When asked about their professional role, 55 (out of 88) answered this question; the largest group (34%) came from nursing / AHP professions; with the balance being made of up of medical staff (11%) and managers / admin staff (15%). There was a spread of seniority, with a large proportion of nurses (42%); 10% at consultant level. Length of service appeared evenly split between those in post for under 2 yrs. (47%) and those for 5+ yrs. (40.35%). Interestingly 66% had worked at the Health Board for more than 10 yrs. The majority (60%) did not have cross organisational responsibilities. Half had done some QI training in the past and gave positive responses when asked about the benefits of this.

Semi-structured interviews

Semi-structured interviews were conducted via telephone with key sponsors (n=3), coaches (n=5) and QIC participants (n=17) from the 11 project teams during the QIC programme.

Sponsors

Sponsors were asked about the rationale for the programme as this would dictate how success was judged. It was anticipated that differing perspectives would offer varied responses and that views about the reasons for the programme would not have been shared. When asked, there was a strong consensus towards the view that the progamme was designed to enable culture change amongst front line staff, with the QI project being seen as an integral part of this but not of primary significance. There was some indication that project outcomes were being monitored, and the hope was that outcomes would be spread. When asked what success might look like, one sponsor noted: *"For me it's if people talk about it – spread their knowledge – sustainability … there are some unintended successes – sustain and grow. Also, if people gain confidence and are able and willing to*

take on another project – feel that they have "permission" to do others. Of course, the Board want quantitative – performance measures too – and want to see evidence of culture change. They will be doing a CBA" (Int 2).

When asked about challenges the responses were almost entirely internal and at a logistical level rather than, for example, at strategic or managerial levels. Issues around finding appropriately qualified coaches figured strongly, and the programme team considered themselves fortunate in having been able to persuade a cadre of qualified and experienced coaches, some from the Health Board and some from external sources, to take part. The intention was to continue to offer the programme and perhaps increase the rate of delivery, and the programme team were conscious of this challenge.

The sponsors noted that the level of drop out was low and this was indicative of the regard with which the programme was held. Attendance at workshops was consistently high and feedback universally positive.

QIC Participants

For the QIC participants questions were asked about their experience of:

- The application process
- Team engagement
- The learning programme / workshops
- The support that the team had received
- The project

The overall response from the 17 participants interviewed was very positive towards the QIC programme, as noted by one participant: *"I have just been hugely impressed with this whole experience. I feel it's enriched me. And it's also, I feel, given me food for thought about my own career and what I'd like to do, so I can be nothing but grateful for being involved in it*" (Int 7). With another participant noting it had exceeded their expectations: "*It's exceeded my expectations. I feel that every study day has had high quality lectures and all pertinent. I have to say, it's been the highlight of my month!"* (Int 2).

Some initial ambivalence on the part of a few was quickly overcome and all interviewees confirmed the benefit and relevance of the programme. Their motivation for joining the programme varied; some had been encouraged to apply as a useful way of personal development; some had identified the programme as a way of progressing a long-standing issue within their workplace; some were co-opted by managers for their knowledge and experience.

Many of the teams were made up of members who worked together regularly and so were able to build on existing relationships, but many were not, and this was significant in team formation. When a participant was asked whether she was part of an existing team, she replied: "*No... how different that is for us than it is for a lot of people who are actually working together and have brought their project with them.*" (Int 12). When asked if the team was working well together the same participant responded: "*We are, we are. I think one of the most profound things for me is coming together in a team, you know, looking at other skills of people that you don't work with on a day-to-day basis from different arenas and actually learning from them … I think sometimes identifying your own strengths and weaknesses within that team*".

When asked about the workshops there was general approval, although this was not entirely non-critical. For example, one participant noted the need for more training on IT applications and data management skills: *"I've really enjoyed.. [but] I think what the health board perhaps needs to think*

about is that not many people in the room are actually trained to use Excel and … other sort of applications, software applications, that would make our lives a bit easier (Int 15).

The support offered by coaches received wholehearted approval. Although there was evidence that some teams received more time and access than others, and that the breadth of experience of coaches varied, with some gaps in knowledge, e.g. IT, nevertheless there was reluctance on the part of participants to articulate anything negative.

QIC Coaches

Five coaches were asked questions about:

- How they became involved in the programme
- Their relationship to the team
- Their view of the programme overall
- Their support role
- Progress of the project

A proportion of the coaches held posts related to QI within the Health Board, but some came from external organisations and were approached because of their reputation either as coaches or as QI facilitators. Alongside participant workshops, coaches were given five days of training at the outset and met regularly during the nine months that the programme ran. There was wide acknowledgement that this arrangement was significant to the success of the programme.

The matching of coach and team appeared to have been serendipitous. Time pressures, geography and opportunity costs meant that some teams met with coaches more frequently than others, but communication links were preserved using social media and IT as needed. Coaches were asked about their views of the make-up of teams and whether they perceived any gaps in experience or commitment. One coach felt there was a gap in representation: "*There are gaps, yeah, very good question, we are short of medical representation*" (Int 3). Like the QIC participants, an issue that was mentioned by two coaches was about sharing expertise in e.g., Excel and data collection. Asked if teams were working well together, coaches' responses were overwhelmingly positive.

Projects varied in the degree of progress being made. A significant element of many projects was their attempt to prove a concept with the intention of spreading changed practice across the Health Board and / or across Wales. Indications that this might prove challenging were suggested by two coaches, and they believed that the cascade model generally might prove harder to achieve than initially anticipated.

Post- programme interviews with QIC participants

Semi-structured interviews were undertaken with five participants six months after the QIC programme had ended. Interestingly the impact of the covid crisis was not seen by the interviewees as altogether negative. One respondent said that it meant that the team got more equipment and there was less red tape, enabling faster decision making and greater flexibility of action. She felt that the QIC programme had given her and team members greater confidence and this had resulted in a surer response to the ensuing pressures. Another interviewee said that the advent of virtual clinics meant that patient numbers could be kept up or exceeded and targets set by the project could be comfortably met.

Another respondent talked about the logistical changes that had been required to cope with covid and the alteration in the number and profile of patients, resulting in the need for creative management and further small changes being made. She felt that these changes were being sustained and were merely the start of a continual, ongoing process. Her experience of the programme had helped her to cope with these upheavals.

There was undoubtedly some disruption caused by the covid challenge, and in one case it seemed that due to frontline pressures there was a need to abandon their current project. However, the interviewee did note the value of attending the programme despite it being interrupted by the pandemic "*I felt valued as an individual by the [QIC] programme.. often under pressure that can get lost"* (Int 3). All five interviewees were asked whether it was likely that they would undertake further projects, and although this has not happened to date there certainly appeared to be enthusiasm for this to happen.

Discussion

It is unarguable that the rationale for most Collaboratives is about enabling collaboration and learning amongst professionals and teams, making comparisons across practice, and doing things differently, and thus leading to improvement (de Silva, 2014). But the boundaries of what constitutes a Collaborative are unclear. A literature scan reveals many definitions of QICs in a variety of settings making it difficult to compare and identify a useful profile. DeSilva's (2014) taxonomy includes who is involved; the span / range of influence; the design of the Collaborative and the subject of focus. It's not difficult to disagree with these points but often these are not reported. Here we have provided details of the design of the study and span of focus which is health board wide and non-specific in terms of subject focus.

The wider literature suggests there is consensus about the components that lead to success and that they include *inter alia* sponsorship, area of focus, ideas for improvement, enthusiasm on the part of participants, senior leadership support, preliminary work, and strategies for learning about improvement (e.g. de Silva, 2014). But arguably these elements are necessary for any initiative, whether it is identified as a Collaborative or not; so questions remain about what is distinctive and what added value does the collaborative model offer. A detailed evaluation of the context and structure of an initiative could throw light on what is more or less significant to ensuring success.

The heterogeneity of QICs in the literature has meant that simple evaluation questions (do QICs work? What are QICs good at?) are mainly facile while their complexity has meant that they are ill suited to conventional evaluation methods. To be useful, evaluation approaches must capture the range of possible effects of QICs and be sensitive to unique contexts (Dixon-Woods et al., 2013).

This evaluation was framed within two complementary models; that of Kirkpatrick and Kirkpatrick (2006), which advocates analysis of outcomes in terms of four levels and the Stake (2010) model which focuses on the third and fourth Kirkpatrick level, with particular attention to nuances of the context and mechanisms that accompany the experience of delivering a learning intervention (See figure 3). Together the two models provide a rounded interpretation of the extent to which the QIC intervention was fit-for-purpose.

Kirkpatrick and Kirkpatrick (2006) suggest that an intervention should be judged in terms of level 1 (satisfaction); level 2 (learning); level 3 (behaviour change) and level 4 (outcomes /impact). Evidence from interviews suggested that participants, coaches, and sponsors were all very satisfied with the way in which the programme evolved and ran. This was reinforced by the manner in which attendance at events and retention generally was maintained throughout the nine months of the programme as well as feedback following workshop days. When asked, most participants claimed that they had learned a lot from the programme and from each other. There were some gaps reported, particularly for example, in understanding of measurement and IT. Woodcock et al., (2021) note measurement is a critical element of QI projects, but note it is a highly technical task that requires a degree of expertise and likely to require a separate capacity building programme.

Following guidance from Stake (2010) questions were framed around the rationale for the programme; in particular whether there was consensus on why the programme was run. Amongst key sponsors the firmly held view was that the purpose of the programme was to bring about culture change across the Health Board, that this was merely the first of an ongoing programme of encouraging major change. It was not clear whether all Board members subscribed to this view as time did not allow for wider questioning, nor whether the intervention is a long-term commitment and how success will be judged. Other interviewees expressed the conviction that the intervention was about enabling change and spread of good practice in systems and standards. This points to a common feature of collaborations of this kind (de Silvo, 2014); differing reasons for undertaking an intervention can lead to a range of views about what makes for success (Stake, 2010).

Stake (2010) reminds us to look at the intention of the intervention; what was planned to happen and what did happen. There was some delay in starting the programme and it may be that the time required for launch was not optimally achieved. Participants reported some confusion about the focus and direction of the programme initially. There were ongoing logistical issues around paperwork and venues which added to confusion.

Team makeup varied; some were existing work teams; some teams were created. The former tended towards hierarchical working replicating existing patterns rather than breaking the mould. Cohesion amongst team members seemed high but it was noticeable that when asked, teams reported gaps in knowledge and skills and some representational gaps amongst members. There was a not uncommon and specific problem in getting junior doctors to engage (Jamil et al., 2012); medical involvement included 8 senior doctors across six of the teams.

Project selection was intended to be done early as part of the application process. This presented obvious challenges as teams may not have been formed and topics could have been underdeveloped and/ or unmanageable within the time frame. Some regular improvement challenges were reported, about re-organising work, and about team engagement. There are obvious benefits in creating a team from a wide constituency as they would bring different perspectives and networks to the team, but this can delay the norming, storming and performing elements. It appeared that use of improvement techniques varied, and some projects were closer to the conventional improvement model than others. But if the purpose was to encourage culture change this was perhaps not crucial, particularly as team members are encouraged to continue with improvement work post-programme. Although Board priorities are foremost in project selection, it is likely that any improvement project would fit somewhere in the Board's direction of travel.

An enabling factor for the delivery of the programme was having attracted a sufficient number of experienced and well qualified coaches to mentor and guide teams through their projects (Zamboni et al., 2020). They demonstrated availability and willingness of appropriate levels of expertise in both the soft skills of coaching others and the knowledge base of improvement science and clinical issues that underpin the relationship between teams and coaches. Some coaches already worked for the Health Board, and some came from external sources. It is likely that access varied because of this, given that all coaches had tight schedules. Collaboration between coaches to plug weaknesses in knowledge base happened at the monthly network meetings where coaches were able to share updates on their project teams and request additional support where needed. Cross -fertilisation of QI knowledge and expertise was important for the development of the coaches and the teams. Previous research has noted the importance of having an internal quality improvement team where the membership is stable, multi-disciplinary and includes opinion leaders with previous experience in quality improvement (Zamboni et al., 2020).

Stake's (2010) model invites us to look at the outcomes of the programme; what was intended to happen afterwards and what did happen. It is evident that many projects were designed to be "proof of concepts" and that outcomes if positive would be spread and sustained across the Health Board. Spread and sustainability is generally found to be hard to achieve (Parand et al., 2012) and can be difficult to understand its impact (Lau et al., 2021). Hence, several interviewees expressed doubts about the possibility of this occurring to any great degree. Some pilot outcomes looked doubtful for acceptability at other sites, with a potential dichotomy between the need for standardisation and that of local autonomy. To support further spread and scale of improvements and innovations, leaders both clinical and managerial have a crucial part in continuing to create the conditions of collaborative learning and risk taking outside of the QIC programme (Greenhalgh and Papoutsi, 2019).

The impact of an intervention such as this one is assumed to be felt after the lapse of some time, and even in "normal" times, it is difficult to differentiate change due to an intervention and change due to external factors. The timing of this QIC intervention bears consideration; on the one hand it uniquely prepared participants to address Covid-19 pressures; but on the other hand, Covid-19 severely interrupted the implementation of new skills and systems. Zamboni et al (2020) remind us of the impact of contextual factors, which is this case was completely out of the control of the QIC team. As a result, unfortunately, it was only possible to interview a minority of graduates of the programme, but their responses were probably typical of what others would have said. Enthusiasm for the programme remained, and as a support for the development of leadership skills and as preparation for unprecedented work pressures, it appeared to serve its purpose. If the aim of the programme was to create opportunities for capacity building and peer recognition (Zamboni et al., 2020), then that appears to have been achieved.

It is undoubtedly the case that Covid-19 interfered with implementation of outcomes at a critical point and better could have been achieved in other circumstances (Zamboni et al., 2020; de Silva., 2014) . An adverse result of Covid-19 was the lack of evidence of active engagement in subsequent improvement projects by interviewees.

Conclusions

The aim of this paper was to report on the evaluation of a QIC implemented by one Health Board in Wales. At the outset of this evaluation, it was believed that the focus would be on process and outcomes, and indeed there were many lessons to be learned from following the intentions and actuality of the programme's design and implementation. But external events forcefully highlighted the significance of timing and context for all interventions of this kind.

The evaluation was designed to examine the extent to which the QIC programme was fit-for-purpose in meeting its intended aims of supporting front-line staff as they acquired and used their knowledge and skills of QI techniques. Although the focus of the programme was on encouraging culture change there was also a widespread desire to implement new working practices and encourage specific alterations in methodologies for patient care.

In examining the preparation and roll-out of the programme the evaluation team adopted two models which in complementary ways enabled the gathering and analysis of evidence of activities and outcomes. Using Kirkpatrick and Kirkpatrick's (2006) model of four levels of analysis the team concluded that the programme provided universal satisfaction, and that a significant body of knowledge and skill was imparted to participants, the majority of which was applied to concomitant projects. Using the Stake (2010) model the evaluation team looked at what was intended to happen, what actually happened, and whether the outcomes were optimal. The conclusion was that,

although overall the programme received positive feedback, there were some areas that could be addressed when planning and delivering the intervention to future cohorts.

Stake's (2010) model enabled a rounded picture to emerge with particular focus on the contingencies – i.e., what needed to be in place at varying stages of the programme. For example, the preparation stage needed to be adequate to enable full comprehension of what is intended by all parties involved, including participants. In an ideal world this can be achieved but in reality, there are frequently obstacles to this, which have resonance later in the programme. Stake (2010) also encourages us to look at the congruence between what was intended to happen and what really happened. Here some gaps are apparent between the way the intervention was designed and subsequent needs at the Health Board.

In Stake's (2010) terms, intentions and realities collided at the close of the programme and this seeks to remind of the significance of context (Zamboni et al., 2020). On the one hand it can be considered that the programme was timely in preparing staff to cope with the covid crisis; but on the other hand, it was unfortunate that the full fruits of the intervention were not able to be retained. In summary, Stake's (2010) model was useful in guiding the broad approach to evaluating the intervention, by helping to highlight the necessity for careful consideration of the contingent dependence of each stage of the rollout. It proved equally helpful to look at the congruence between what was planned and what actually happened in particular given the unprecedented circumstances that followed the close of the taught element of the intervention.

Limitations of this Research

The study is limited to one health organisation which operates across four sites. The timing and context of this evaluation study were significant and arguably led to less confidence in the findings than might have been the case. The learning programme and project experience that formed the core of the QIC took place during 2019 culminating in an intended celebration in early 2020. This coincided with the outbreak of the covid pandemic and therefore with the event not going ahead presented outcomes of the improvements being shared. However, this was not all negative. Participants found themselves more prepared to embrace change than they might otherwise have been, but it did curtail many plans for sustainability and spread, and it did hinder plans for the second run of the programme. A condensed version was run online. This interruption to the programme did make data collection for the evaluation more of a challenge than it might otherwise have been.

The Stake (2010) model, although useful, is very much a case study approach and caution should be exercised when trying to generalise from findings. The particular circumstances of this evaluation make it problematic to even reach a definitive conclusion of what might have been given from other timeframes. And this does not detract from the opportunity offered by this model to focus on the possibility of measuring the gap between the rationale for a programme and its impact.

Further Research

Evaluations of this type are essential if the Collaboratives model is to be refined; further research can help establish those elements that are necessary and those less so. This collaborative involved a large number of participants and a significant investment of time and money on the part of the Health Board making it necessary to ensure that future iterations benefit from lessons learned at this pilot stage, not just for the Health Board itself but for the wider professional community.

The combination of frameworks employed, using Kirkpatrick and Kirkpatrick (2006) to judge key individual outcomes, with Stake's (2010) approach to judge the significance of context and systems, makes for a powerful combination that would benefit from further testing, with particular emphasis

on other designs of collaboratives and the longer-term impact of such programmes. Attention to the synergy between rationale and outcomes can provide indicators of impact.

References

Batalden, P.B. (2003). *The Breakthrough Series: IHI's collaborative model for achieving breakthrough improvement.* Retrieved from https;//www.ihi.org/resources/pages/IHI White Papers / The Breakthrough Series IHIs Collaborative Model for Achieving Breakthrough Improvement.aspx

Benning, A. Ghaleb, M. Suokas, A. Dixon-Woods, M. Dawson, J. Barber, N. et al. (2011). Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation BMJ, Vol. 342, d195 doi:10.1136/bmj.d195

Berwick, D.M. (1989). *Continuous Improvement as an Ideal in Health Care. New England Journal of Medicine*, Vol. 320, No. 1, pp. 53–56. doi:10.1056/NEJM198901053200110.

Bion, J. Richardson, A. Hibbert, P. et al., (2013). 'Matching Michigan': a 2-year stepped interventional programme to minimise central venous catheter blood stream infections in intensive care units in England. *BMJ Quality & Safety*, Vol. 22, pp. 110-123.

Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, Vol. 3, No. 2, pp. 77-101.

Caley, L., Williams, S., Spernaes, I., Thomas, D., Behrens, D., and Willson, A. (2021). Frameworks for evaluating education programmes and work-related learning: a scoping review. *Journal of Workplace Learni*ng, Vol. 33, No. 6, pp. 486-501.

Cooke, J. Langley, J. Wolstenholme, D. and Hampshaw, S. (2017). "Seeing" the difference: The importance of visibility and action as a mark of "authenticity" in co-production. *International Journal of Health Policy and Management*, Vol. 6, No. 6, pp. 345-348.

Deming, W.E. (1986). *Out of the Crisis*. Cambridge, MA: MIT Press

De Silva, D. (2014). *Improvement collaboratives in health care: evidence scan*, The Health Foundation, London.

Dixon-Woods, M. Leslie, M. Tarrant, C. and Bion, J. (2013). Explaining Matching Michigan: an ethnographic study of a patient safety program, *Implementation Science*, Vol. 8, No. 70. https://doi.org/10.1186/1748-5908-8-70.

Dixon-Woods, M. McNicol, S. and Martin, G. (2011). Ten challenges in improving quality in healthcare: lessons from the Health Foundation's programme evaluations and relevant literature, *BMJ Quality & Safety*, Vol. 21, No. 10, pp. 876-884.

Donabedian A (2005) Evaluating the Quality of Medical Care. *Milbank Quarterly*, Vol. 83, No. 4, pp 28-42.

Engestrom, Y. (1987) *Learning By Expanding An Activity- Theoretical Approach to Developmental Research.* Helsinki: Orienta-Konsultit.

Garcia-Elorria, E. Rowe, S. Teijiro, M. Clapponi, A. and Rowe, A. (2019). The Effectiveness of the Quality Improvement Collaborative Strategy in Low- and Middle-income Countries; A Systematic Review and Meta-Analysis. *Plus/One*. Retrieved from https://doi/10.1371/journal.pone. 0221919

Greenhalgh, T. and Papoutsi, C. (2019). Spreading and scaling up innovation and improvement. *British Medical Journal*, May 10, Vol. 365. doi: 10.1136/bmj.l2068

Hulscher, M. Schouten, L. Grol, R. and Buchan, H. (2013). Determinants of success of quality improvement collaboratives: what does the literature show? *BMJ Quality & Safety*, Vol. 22, No. 1, 19-31.

IHI (Institute for Healthcare Improvement). (2003*). The Breakthrough Series: IHI's Collaborative Model for Achieving Breakthrough Improvement.* IHI Innovation Series white paper. Boston.

Ishikawa, K. (1985). *What Is Total Quality Control? The Japanese Way*. Englewood Cliffs, NJ: Prentice-Hall.

Jamil, F. Foley, T. Bate, A. Tse, Y. and Moghal, N. (2012). Barriers and enablers to the engagement of junior doctors in quality improvement, *Archives of Disease in Childhood*, Vol. 97, No. 1, pp. A157-A158.

Juran, J.M. (1951). *Quality Control Handbook.* New York: McGraw-Hill.

Juran, J.M. (1964). *Managerial Breakthrough.* New York: McGraw-Hill.

Kilo, C. (1998). A framework for collaborative improvement: Lessons learned from the Institute for Healthcare Improvement's Breakthrough Series, *Quality Management Healthcare*, Vol. 6, No. 4, pp. 1-13.

Kilo, C. (1999). Improving care through collaboration, *Pediatrics*, Vol. 103, No. 1 Suppl E, pp. 384-393.

Kirkpatrick D. and Kirkpatrick J. (2006) *Evaluating Training Programs: Third Edition*. Berrett-Koehler San Francisco.

Laur, C. Corrado, A. Grimshaw, J. and Ivers, N. (2021). Trialists perspectives on sustaining, spreading, and scaling-up of quality improvement interventions. *Implementation Science Communications*, Vol. 2, No. 35. https://doi.org/10.1186/s43058-021-00137-6

Nadeem, E. Olin, S. Hill, L. Hoagwood, K. and Horwitz, S. (2013). Understanding the Components of Quality Improvement Collaboratives; A Systematic Literature Review. *The Milbank Quarterly,* Vol. 9, No. 2, pp. 354–394.

Normann, R. (1991). *Service management: strategy and leadership in service business*, Chichester: John Wiley and Sons.

Osborne, S. and Strokosch, K. (2013). It takes two to tango? Understanding the co-production of public services by integrating the services management and public administration perspectives, *British Journal of Management*, Vol. 24, pp. S31-S47.

Osborne, S. Radnor, Z. and Strokosch, K. (2016). Co-Production and the CoCreation of Value in Public Services: A suitable case for treatment?. *Public Management Review*, Vol. 18, No. 5, pp. 639-653. doi: 10.1080/14719037.2015.1111927.

Ostrom, E. (1972). Metropolitan Reform: Propositions derived from two traditions, *Social Science Quarterly*, Vol. 53, pp. 474-493.

Øvretveit, J. Bate, P. Cleary, P. Cretin, S. Gustafson, D. McInnes, K. McLeod, H. Molfenter, T. Plsek, P. Robert, G. Shortell, S. and Wilson, T. (2002). Quality Collaboratives: Lessons from Research. *Quality & Safety in Health Care*, Vol. 11, No. 4, pp. 345–51.

Parand, A. Benn, J. Burnett, S. Pinto, A. and Vincent, C. (2012). Strategies for sustaining a quality improvement collaborative and its patient safety gains. *International Journal for Quality in Health Care.* Vol. 24, No. 4, pp. 380–390.

Rycroft-Malone, J., Burton, C.R., Bucknall, T., Graham, I.D., Hutchinson, A.M., and Stacey, D. (2016) Collaboration and Co-Production of Knowledge in Healthcare. *International Journal of Health Policy and Management*. Vol 5, No. 4, pp. 221-223.

Solberg, L.I. (2005). If You've Seen One Quality Improvement Collaborative. *Annals of Family Medicine*, Vol. 3, No. 3, pp. 98–99. doi:10.1370/afm.304

Stake, R. (2010) Qualitative Research: *Studying how things work.* The Guilford Press New York.

Stufflebeam D. (2001). The Metaevaluation Imperative. *American Journal of Evaluation*, Vol. 22, No. 2, pp. 46-62.

Vygotsky, L. (1978). *Mind in Society*. Massachusetts, Harvard Business Press.

Wenger E (1998) *Communities of Practice: Learning as a social system.* Retrieved from http://www.co-i-l.com/coil/knowledge-garden/cop/lss.shtml (Accessed 4th August 2021).

Wilson, T. Berwick, D. and Clearly, P. (2003). What do collaborative improvement projects do? Experience from seven countries, *Joint Commission Journal on Quality and Safety*, Vol. 29, No. 2, pp. 85-93.

Woodcock, T. Liberati, E. and Dixon-woods, M. (2021). A mixed-methods study of challenges experienced by clinical teams in measuring improvement, *BMJ Quality & Safety*, Vol. 30, No. 2, pp. 106-115.

Zamboni, K. Baker, U. Tyagi, M. Schellenberg, J. Hill, Z. and Hanson, C. (2020). How and Under what Circumstances do Quality Improvement Collaboratives Lead to Better Outcomes? A Systematic Review. *Implementation Science*, Vol. 15, No. 27. https://doi.org/10.1186/s13012-020-0978-z

Zamboni, K. Singh, S. Tyagi, M. Hill, Z. and Hanson, C. (2021). Effect of collaborative quality improvement on stillbirths, neonatal mortality and newborn care practices in hospitals of Telangana and Andhra Pradesh, India: evidence from a quasi-experimental mixed-methods study, *Implementation Science*, Vol. 16, No. 4. https://doi.org/10.1186/s13012-020-01058-z

Table 1. Typical steps in the Breakthrough Series Collaborative Model

1. Lead organisation identifies topic areas where improvements are required to close the gap between best and typical practice
2. The collaborative lead organisation assembles an expert panel
3. Panel proposes a set of ideas for closing the gap
4. Teams are recruited to take part in the collaborative
5. Participants undertake pre-work including the formation of local teams, development of goals and measures and understanding of current practice
6. Over a 6-12 month period participating teams attend three learning sessions on improving practice and improvement techniques which are implemented between sessions.
7. Between the teaching sessions, teams share learning and experiences via conferences calls and emails and submit progress reports.
8. The lessons learned are disseminated via national meetings, conferences and reports.

Adapted from: Wilson et al., (2003, p86)

Table 2 Key attributes, enablers and challenges of implementing Quality Improvement Collaboratives

Key attributes	Possible enablers	Possible challenges
Multi-professional teams (Kilo, 1998; Nadeem et al, 2013) including patients and carers.	Collaborative to be facilitated by clinical experts and experts in QI.	Difficulties linking effects of the collaborative to outcomes due to other interventions (De Silva, 2014).
Scheduled **learning/education** sessions (Kilo, 1998; Nadeem et al, 2013; De Silvo, 2014)	Clear and relevant measurement strategies (De Silva, 2014)	Understanding the role of the context – both internal and external to the organisation (Zamboni et al., 2020; Dixon-woods et al., 2012).
Span of collaborative can extend from one organisation to a national event of many organisations (De Silva 2004)	Building sustainability in the collaborative programme (De Silvo, 2014).	Sustaining improvements often challenging once the collaborative has completed (Dixon-woods et al., 2012).
Design of collaborative – Time limited (De Silvo, 2014) typically between 6 and 18 months (Kilo, 1998).	Providing additional support to achieve spread and scale where appropriate.	Spread and scale of improvements beyond the collaborative (Dixon-woods et al., 2012).
Topic of collaborative – based on improving outcomes for patients and/or systems (De Silvo, 2014). National collaboratives tend to focus on one topic (disease) whereas collaborative run by an organisation tend to encompass various topics and services.	Joint learning approach that is flexible to the needs of the participants (De Silvo, 2014).	Managing unintended consequences of any of the collaborative's improvement interventions (Dixon-woods et al., 2012).
Measurement – each organisation/team set measurable targets and collect their own data (De Silvo, 2014).	Multi-professional teams that have volunteered to take part in the Collaborative	
Resources include building of improvement capability and application of improvement models and techniques e.g. Plan Do Study Act (PDSA cycles)		

Table 4 Details of interview participants

Number of interviewees	Role in QIC
17 QIC participants	From the 11 QIC teams
5 coaches	From 11 QIC coaches
3 stakeholders	From programme designers/senior managers

Table 3 Details of survey participants

Role of survey participants	Number of survey participants (n=53)
Clinical	75% (n=40)
Admin/management	25% (n=13)
Time in current post	Number of survey participants (n=88)
0-2 years	47% (n=40)
3-4 years	12% (n=11)
5 + years	41% (n=36)

Figure 1 Kirkpatrick and Kirkpatrick's 4 level evaluation framework

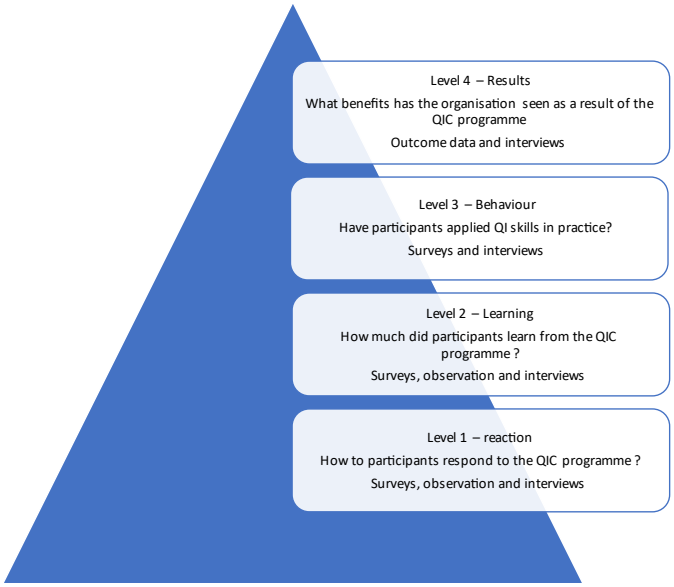

Figure 2 Stake's evaluation framework

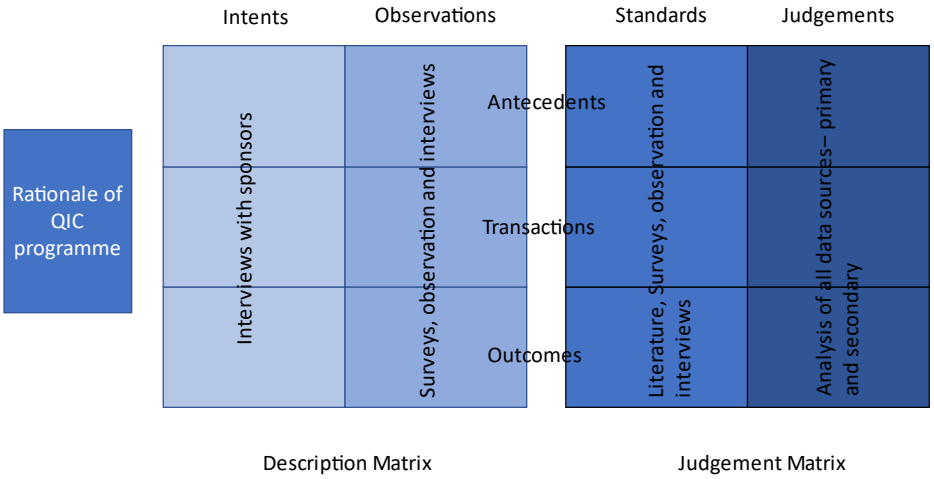

Figure 3 QIC evaluation framework

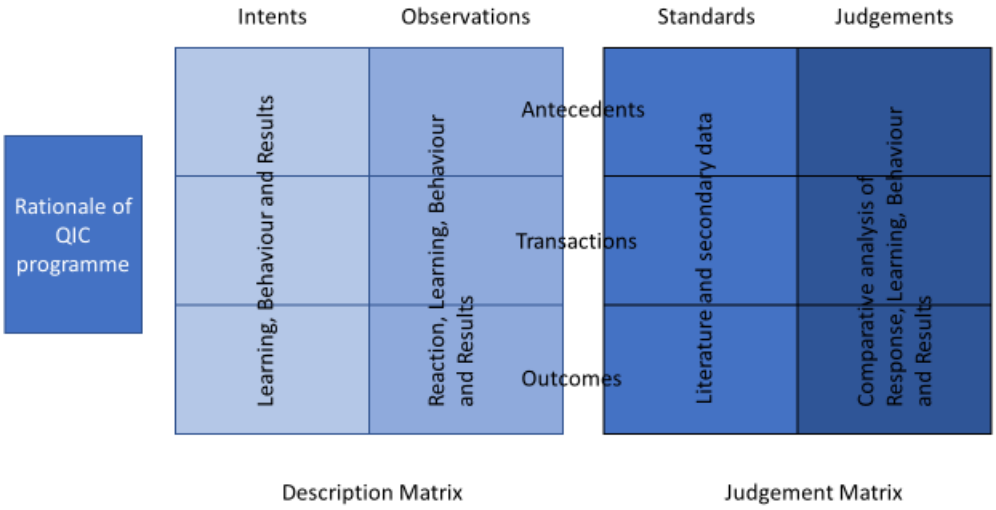

Figure 4 Design of Quality Improvement Collaborative