Mobilising Q methodology within a realist evaluation – lessons from an empirical study

Abstract

Realist evaluation and Q Methodology are established approaches in social science. However, integration of Q methodology within a realist evaluation is scarce. This paper attempts to illustrate (through a recent evaluation) how Q methodology can support a realist evaluation. The paper attempts to capture the philosophical compatibility of the two approaches creating an argument for Q's integration within realist evaluation. Through the case study selected (a realist evaluation of an evaluation capacity building framework) the iterative methodological process is presented, capturing a snapshot of the findings from the evaluation. This illuminates how Q met the philosophical aims of a realist evaluation to make sense of how and why and under what circumstances a programme or intervention works. It is argued that Q methodology is entirely suitable for capturing the three key stages in a realist evaluation of developing, testing, and refining programme theory. It is intended that this paper can contribute to inspire other realist evaluators and methodologists to make use of Q as a tool in their evaluation.

Key words: Realist evaluation, Q methodology, programme theory, logic of abduction

Introduction

Realist evaluation (Pawson and Tilley, 1997) is an established evaluation methodology within the field of social science, which seeks to derive a deeper understanding into how and why programmes may work, and for whom (Westhorp, 2014). With its generative approach to causality realist evaluation has become attractive for those developing, delivering and evaluating programmes to capture the role that context plays alongside resources and reasoning (Westhorp, 2014) to explain programme outcomes (Dalkin et al, 2015).

Since its emergence and mobilisation across multiple fields of social change and innovation, realist evaluation has reflected a method neutral growing evaluation approach that continues to be forward thinking, open to methodological innovation, and debate. On a methodological level Pawson and Tilley (1997) have always encouraged evaluators making use of this approach to explore various qualitative and quantitative approaches to serving the needs of a realist evaluation (be it questionnaires, surveys, interviews, content analysis or participant observation). In terms of debate, interesting exchanges have taken place in recent years between Porter (2015a, 2015b) and Pawson (2016). Much of this debate has focused on Porter problematizing the philosophical determinants of realist evaluation, and in particular how Pawson and Tilley (1997) conceptualise the term 'mechanism'. Furthermore, whilst Pawson 2016 has sought to defend his positon accordingly, these debates have ramifications for methods.

Nevertheless, despite this context of debate realist evaluation has not been short of methodological innovation and creativity with academics developing for example the realist interview (Manzallo-Santaella, 2016) and exploring its compatibility with randomised control trials (Hawkins, 2014). Dalkin et al 2018 have also made a very novel contribution with their integration of Soft Systems Methodology which enables a better understanding of complex situations through participant reflexivity as well as the dissaggregation of the programme

mechanism (Dalkin et al 2015). In addition, Shearn et al's (2017) depiction of initial rough programme theories to support realist evaluation has also been valuable. This is exciting because it demonstrates the reflexivity of the field to experiment with new approaches, providing that those approaches recognise and enable the ontological depth and the mechanisms of change at play within the programmes to be realised.

Within a context of methodological innovation and opportunity this paper aims to position the established approach of Q methodology (Stephenson, 1953; Brown, 1980) as a suitable method to support a realist evaluation across all stages of programme theory development, testing and refinement. According to Watts and Stenner (2012) Q methodology focuses on the subjective viewpoints of social actors, to decide what is meaningful and what does and does not have value and significance from their perspective (Watts and Stenner, 2005). Apart from Harris (2018) there have been limited published examples of Q methodology's integration with realist evaluation, although, there are many examples of its implementation in evaluation practice (Ramlo, 2017; Baptiste, 2011; Ramlo et al 2008).

This paper draws upon the methodological framework of a recent realist evaluation which mobilised Q alongside conventional methods of interviews to develop test and refine a programme theory. The key emphasis of the paper seeks to offer insight and guidance as to how realist evaluation can be combined with Q methodology. Whilst it is not the aim of the paper to focus entirely on the context and findings of the evaluation, the case study and a selection of the findings will be presented to bring the methodological design to life (see Harris (2018) for the full findings).

To provide a roadmap of the paper, it will begin with a set of key arguments for realist evaluation and Q's integration. This will be followed by a brief description of the case study and programme theory under test which then leads to the methodological design and illustration of selected empirical findings.

The Context of Realist Evaluation and Q

Realist evaluation

Realist evaluation, emanating from the philosophical foundations of realism offers crucial insight to make sense of what works for whom, why and in what circumstances within programmes. For causality, this is conceptualized as the behavioural triggers and mechanisms that a programme may bring about (Pawson and Tilley, 1997). Within a realist sense, causality is not something that is successionist (a=b); instead it is generative and places emphasis on how reasoning, attitudes and behaviours of participants interact with the resources of a programme. This is what Pawson and Tilley refer to as the programme mechanism. This helps to explain how resources are provided in social systems and reasoned against by individuals (Dalkin et al, 2015). To reiterate the terms of mechanism and context above, Marchal et al (2012) state that realist evaluators believe causality to be generative in the sense that actors have the potential for change by their very nature (agency). As for context, potential mechanisms leading to outcomes will only be activated if certain conditions are right. Thus, an understanding of context and its fluidity of change are central to realist evaluation to understand what it is about a programme or intervention that makes it work or not (Pawson and Tilley, 1997).

In concise terms, there is no universal defined and agreed approach for carrying out realist evaluation (Marchal et al, 2012; Evans and Killoran 2000; Pawson and Tilley, 1997). However, a clear consensus surrounds the importance of developing programme theory, usually via the formulation of the CMO (context, mechanism and outcome) configuration which draws upon data gathered about the programme in hand via key stakeholders, primary and secondary data. These CMOs (also referred to as conjectures) are then usually tested via various qualitative and quantitative mixed methods which should seek to identify the causal processes at play within the context of the intervention or programme under evaluation (Wong et al 2016). Key questions are framed around what worked for whom, in what circumstances and why? The final stage, having analysed the data and tested the CMOs then consists of refining the programme theory which then leads to further testing (Rycroft-Malone et al, 2010). In most simplistic terms a realist evaluation strategy answers what works for whom in what circumstances and why within the research strategy of developing programme theory, conjecturing CMOs then refining the programme theory through a range of mixed methods.

Q Methodology

Q emerged in 1935 as a methodology created by William Stephenson as an innovative adaptation of Charles Spearman's traditional method of factor analysis, which focuses on the revealing of patterns of association between variables within a given data matrix (Watts and Stenner, 2012). Put simply, Q focuses on the subjective viewpoints of its participants (Watts and Stenner, 2012; Brown, 1980). It asks its participants to decide what is 'meaningful' and hence what does (and what does not) have value and significance from their perspective (Watts and Stenner, 2005). As proposed by Kelly et al (2016) Q methodology attempts to make sense of the range of communicated ideas on a topic, "followed by exploration of the prevailing variations in it, and finally, to logically connect these variations in an orderly way to each other" (pg.3).

In practice, Q involves developing a set of statements that may represent viewpoints of certain individuals from the 'concourse' about a particular topic, programme or issue. These statements are then ranked relative to one another by those individuals into piles of agreement, disagreement and neutral feeling by use (usually) of a bell curve (Watts and Stenner, 2012). Factor analysis is then implemented to uncover shared viewpoints amongst those individuals about the issue in question which are then interpreted to produce holistic narratives in accordance with each factor. To summarise, Q usually involves the following 7 stage process as depicted in table 1:

Insert table 1 about here

Q's distinction and novelty lies within its ability to synthesise quantitative and qualitative methods. Stenner and Stainton-Rogers (1998 in Watts and Stenner, 2005) refer to Q as being 'qualiquantological' where it offers the potential to blend the two positions together. Barry & Proops (1999) highlight how Q seeks to establish patterns within and across people as opposed to patterns across individual traits. This is further supported by Pike et al (2015) who suggest that whilst a questionnaire may broadly establish the ways stakeholders may feel about a topic, it is "constrained by the limited prior understanding held by the researcher of the range of accounts constructed by the participants" (pg.5). Moreover, Q's rigorous

approach can deliver a semi quantitative approach for understanding shared viewpoints (Pike et al, 2015; Watts and Stenner, 2005). Watts and Stenner (2005) point out that instead of breaking up subject matter into a series of constituent themes like other phenomenological and interpretive approaches, Q can show the primary ways in which themes are interconnected or related by a group of participants.

What are the arguments for embedding Q with realist evaluation?

Having provided an overview of realist evaluation and Q in the preceding sections, this section intends to create a series of arguments concerning how they relate to one another and why Q may be a suitable method for supporting a realist evaluation.

The first argument to make for their alliance concerns the point that realist approaches to evaluation favour and support the combination of qualitative and quantitative methods (Pawson and Tilley, 1997; Pawson and Manzano, 2012). However, one could question how often the two are combined, and Manzano (2016) has highlighted the predominant use of interviews in realist evaluation. Making use of qualitative and quantitative approaches in realist evaluation can support claims for reliability and validity (Pawson and Tilley, 1997; Westhorp, 2014) if used appropriately. Q's qualiquantological approach helps to mitigate against the issue given the way it synthesises quantitative factor analysis with qualitative interpretation. However, it is recognised that the right methods depend on the programme under study and the field in which it is situated. Whilst Q may support the alliance of quantitative and qualitative approaches evaluators and researchers must consider whether the sample have the capacity to engage with Q (for example, demographic variables), and whether the context of where the programme is situated, and the programme itself would allow Q methodology to be undertaken.

The second argument for embedding Q within realist evaluation relates to their philosophical compatibility. It is acknowledged here that any such discussion could be a paper in itself, therefore below provides a concise overview of the arguments for their alignment.

The most natural place to begin in this argument focuses on the very starting point of any Q study; concourse. Put simply, the concourse is "a field of shared knowledge and meaning from which it is possible to extract an identifiable universe of statements about any situation or context" (Stephenson, 1986a, pg.44). This aligns neatly with the scientific realism foundations of realist evaluation given the similar comparisons to that of 'context' which is so crucial within realist methodology to capture the conditions needed to trigger mechanisms that lead to outcome patterns (Pawson and Tilley, 1997). As such the context represents multiple dimensions of a programme under evaluation such as environmental and individual circumstances and an appreciation of context is thus required for any insight to take shape. As Stephenson (1982) asserts, the concourse represents the individual's cultural heritage, and, methodologically in reference to Watts and Stenner (2012) the overall population of statements from which a Q-set is sampled. It can be drawn together based on understandings about individuals, cultural heritage and policy – which are many of the aspects that should be considered for context in realist approaches. In addition, this can be representative of mid range theories (Shearn et al, 2017) and provisional programme theories themselves for testing which coheres nicely with the realist position in recognising context to inform programme theory. Q does this as any Q study requires a clear picture of the concourse / context and provisional programme theories which drives its design and development of statements (Q-set) to be sorted. The same approach applies to realist evaluation where the quest to unearth

mechanisms is dependent on a rich understanding of the context. Thus, mobilising Q methodology with realist evaluation provides an excellent foundation whereby the concourse activities capture and recognise key elements of the context and provisional programme theory for testing (Harris, 2018).

Interestingly, another argument for compatibility focuses on the logic of abduction which is a central focus with Q methodology. Emanating from the work of Peirce (1839-1914) abduction can be simply defined as "studying facts and developing theory to explain them" (1931 / 1958, pg: 90 cited in Watts and Stenner, 2012, pg:39). Unlike induction's attempts to describe phenomenon, abduction and its synonomous comparison with retroduction (Peirce, 1955; Jagosh, 2015b) focuses on explanation and theory building by going back from, below or behind observed patterns (Lewis-Beck et al, 2003). This perhaps shows the strongest alignment to realist principles as within realist evaluation building theory to test and explain what works for whom and how and why is crucial. To support and strengthen this assertion further Watts and Stenner (2012) highlight the importance abduction places upon explaining why, which again is central to realist methods. In the same way that a Q methodologist searches for clues to form hypotheses and deeper levels of questioning to build a story, so too does a realist evaluator who is searching for clues and indications associated with mechanisms as part of their CMO building. At the heart of the realist evaluator's approach is the attempt to unearth what works for whom in what circumstances and why (Pawson and Tilley, 1997) where the identification of hidden mechanisms (Westhorp, 2014) is prioritised. Q holds significant potential for aiding this process through the abductive process of building theory / programme theories representative of certain shared viewpoints. Furthermore, it can be suggested that through the aid of factor analysis and in depth qualitative interpretation of such factors this can foster a deeper insight and testing of programme theory for conjecturing a more objective and reliable set of contexts and mechanisms where the individual interpretation of the researcher is not the only means.

Introducing the realist evaluation case study mobilising Q

Having provided a rationale for the integration of Q methodology within realist evaluation, the next section of the paper will now focus on the evaluation case study which operated as a context for the delivery of Q within the realist evaluation. The programme under evaluation is concisely described which leads to the methodological design adopted.

The context of the evaluation was an educational based sport and physical activity programme at a university in the south of England. The programme involved student practitioners as part of their curriculum researching a social issue in their surrounding community to devise a physical activity intervention to address the issue (see Harris, 2018). For the purpose of this programme each intervention was called a Coaching Innovation Project (CIP). The programme as a whole constituted at least 20 CIPs (made up of between 2 and 4 students per intervention) being delivered to address social issues through physical activity in the surrounding areas. These types of interventions (known as Sport for Development) currently are subjected to key questions concerning their impact (Coalter, 2013; Harris, 2018) and to what extent evidence is accurately collected.

On this back drop a participatory, capacity building evaluation framework (Harris, 2018) was then constructed to address some of these questions with the intention to enable practitioners delivering small scale physical activity and social change projects to be able to independently evaluate their projects to create a sound evidence base for how and why they worked. The

framework consisted of building capacity with the student practitioners in a participatory and co-productive environment. To clarify 5 diverse CIPs made up of between 2-4 student practitioners opted into the capacity building framework totalling 15 student practitioners. The CIPs opting in engaged with the following social issues spanning diverse areas:

- Providing physical activity for young black and minority ethnic group mums.
- Providing physical activity and social stimulation for older, lonely or isolated people.
- Providing physical activity movement based activities and resources within a primary school for year 3 children.
- Providing educational and physical activity for young people in college with special educational needs.
- Increasing the volume of volunteering in sport for young people in a secondary school environment.

Specifically, the framework attempted to train the practitioners to be able to evaluate their own small-scale interventions through the use of realist evaluation. There was a strong rationale to build capacity in realist evaluation in order to demystify the black box of social interventions concerning how and why they work (Harris and Adams, 2016). The participatory dimension of the framework involved providing workshops and one to one sessions to support evaluation competency in addition to the provision of additional resources and templates to aid the process. In essence, they were supported (within a realist theme) to develop a programme theory for their intervention, and then through its delivery test it via realist evaluation. The workshops and one to one sessions were constructed and delivered alongside the relevant stages of the evaluation such as programme theory development, choosing appropriate methods to answer evaluation questions, data analysis and presentation of evaluation findings.

Having delivered the capacity building framework, the next step involved evaluating it through a realist evaluation. At the heart of the programme theory underpinning the framework was the assumption that realist evaluation in a participatory and co-productive setting could enable practitioners to make a much clearer and reflexive sense of how and why their programmes were working. The key aim of testing the framework was thus to understand what approaches to evaluation worked best for practitioners working within such a context. More crucially, understanding the contexts and how and why the framework worked for those involved in mobilising it influenced the evaluation towards a realist design which will now be explained.

Methodological design

The implementation of Q within the evaluation attempted to make sense of the practitioners subjectivity and shared view-points which could foster an understanding of what worked for whom in what circumstances and why within the capacity building framework.

The evaluation was made up of a series of phases (see figure 1) which sought to depict a clear and comprehensive approach to develop, test and refine programme theory.

Insert figure 1 about here

Establishing a programme theory for testing

The starting point of the design consisted of embracing the underlying philosophy of a realist evaluation which is to understand what worked for whom in what circumstances and why in relation to the practitioners' experience of the capacity building framework. This was firstly addressed through the establishing of the concourse taking into consideration the contextual characteristics of the practitioners and the environment surrounding its delivery (Brown, 1980). This stage considered other key aspects informing the framework such as primary and secondary literature focusing on evaluation capacity building. Naturally, the concourse stage enabled an even deeper grasping of what the evaluation framework intended to achieve which supported the establishment of the provisional programme theory for testing (a set of tentative CMOs were conjectured) which is standard conventional practice within realist evaluation (Pawson and Tilley, 1997). This enabled a clear synergy between realist evaluation and Q and how the concourse supported this element.

As Watts and Stenner (2005) state a comprehensive approach to the concourse is absolutely crucial to inform the Q set statements which are indeed reflective of the concourse itself (Hensel, 2017). The Q set in this instance was made up of a total of 30 statements emanating from the concourse, and they were constructed in accordance with the programme theory under evaluation. In essence they were structured across three key areas; 1: competency in Monitoring and Evaluation, 2: competency in and attitudes towards realist evaluation and 3: feelings about the participatory dimension of the framework (see table 2). The P-Set for this study was made up of a 100% representation of all those who took part in the capacity building framework. In total 15 participants from all of the projects outlined earlier took part in the Q sorting exercise.

Insert table 2 about here

Testing the programme theory

The Q sorting exercise was mobilised over four different sessions where the participants would attend either separately or within their project groups totalling no more than four members at a time. This made the process more manageable and easier to conduct, and a clear set of instructions were provided to the P-set. Within each session the participants were given the 30 statements and were asked to place them in three piles of agree, disagree and neutral. The statements were then ranked accordingly into the bell curve (see figure 2) made up of 30 squares / sections. This was mobilised by the P-set placing the statements of agree on the right side, disagree on the left side and neutral in the middle of the curve.

Insert figure 2 about here

Following factor analysis, four factors were identified for interpretation, using conventional criteria. Analysis was conducted through the use of Schmolck and Atkinson's (2010) 'PQ Method' software which is commonly used by Q practitioners (Pike et al, 2015). Data linked to each sort was entered accordingly into the program and followed the procedure described by Pike et al (2015). This involved extracting principal components from the correlation matrix by using the Principal Components Analysis (PCA) function followed by VARIMAX rotation. From this a series of seven factors were extracted in line with Brown's (1980) default number for extraction. Moving away from the PQ method software each factor document was investigated to make a decision as to how many factors overall would be retained and selected for interpretation. This was done by adopting the "objective decision-

making criteria" as advocated by (Watts and Stenner, 2012 pg.106). This focuses on various criteria such as study variances associated with each factor (eg. how much percentage of the study variance each factor represents) which according to Watts and Stenner is a strong indication of a factor's strength.

Of the four factors 6 participants loaded into factor one, 4 into factor two, 3 into factor three and 2 into factor four. This full qualitative interpretation employed the logic of abduction to qualitatively capture the significant viewpoints within each factor. This logic of abduction started with identifying where certain factors had ranked certain statements as well as distinguishing statement rankings compared to other factors (as well as taking into consideration demographic variables). In this case Watts and Stenner's (2012) crib sheet was used to build a picture of how and why the evaluation framework worked for the practitioners. To be clear this enabled the theorising of key contextual, mechanistic and outcome related insight coming from those shared viewpoints. This led to the provisional draft of holistic narratives which were written to tell a story about the shared viewpoints. At this stage, the holistic narratives began to create a picture which then enabled the drawing up of a series of CMOs following Dalkin et al's (2015) approach towards disaggregating the programme mechanism.

Using interviews to explore narratives and CMOs

Interviews were then employed to expand upon the Q findings and facilitate deeper insight around the narratives. This was important for a number of reasons. Interviews are commonly used in Q studies to add further depth to interpretations and to validate the factor interpretations produced by the researchers (Pike et al 2015; Gallagher and Porrock, 2010). They were implemented here because within a realist philosophy there is never an end to how far a researcher or evaluator can go to explore mechanisms of change. Gallagher and Porrock (2010) advocate interviewing as it can generate new theory to support the qualitative interpretation. Stergiou and Airey (2011) state that follow up interviews with participants following the Q sort can provide verification of the accuracy of the interpretation. Thus, in this sense, "the use of interviews can increase the validity of Q-studies by illuminating the interpretation of respondents' views through additional qualitative analysis" (Gallagher & Porock, 2010, pg.317).

Gallagher and Porrock also point out that "it is important to highlight that it is not the point of the interview to explore what people thought of the phenomena, but rather to reveal the underlying beliefs and values that lead to a particular stance on the issue" (pg.297). This is in line with realist evaluation, which does not just seek one's view about a phenomenon, but attempts to make sense of the mechanisms that lead to these views.

Four realist interview schedules were created for each factor and a sample of participants from each of the four factors were selected. The interview schedules were independently designed around the provisional narratives and CMOs conjectured from the factor interpretation. This enabled further exploration into the contexts and mechanisms employing realist style questioning as advocated by Manzano (2016). Manzano's (2016, pg.12) "topic guide for qualitative interviewing" helped with the construction of specific questions. Some of these questions explored context, whereas others focused more on the mechanism and outcome characteristics. Some examples are included below:

- "How competent were you in conducting evaluation prior to taking part in the training framework?" (Context)
- "It seems to be challenging at present in the industry to carry out evaluation, and lots of practitioners are being criticised for focusing too much on numbers. How do you feel about that?" (Context)
- "How and in what ways has realist evaluation influenced you"? (Mechanisms)
- "Are there any changes that you are seeing in your professional practice since being involved in the framework?" (Mechanisms and outcomes)

The purpose of the interviews was not simply to verify the accuracy of the interpretation of the factor, but was to allow more focused exploration of the CMOs in relation to each factor.

Having carried out this process and analysed the interviews the provisional holistic narratives and CMOs were revisited and refined accordingly as the final stage in refining the evaluation framework under evaluation.

Snapshot of findings

The iterative methodological process above uncovered four sub groups of practitioners. Each of the four subgroups (depicted by the factors) demonstrated how the framework worked for different practitioners in variable ways, and supported the refining stage of the programme theory under test. It is not possible to display all four factors, therefore table 3 below provides a succinct summary of distinctions across each factor. Factor one has been chosen to demonstrate how the holistic narratives and CMO configurations were displayed.

Insert table 3 about here

Factor 1 Holistic narrative: "From zero evaluators to accomplished"

(The narrative below is supported with specific reference to the Q set statements and where they were ranked, eg. 1+3 means that those within this subgroup rated statement 1 as +3)

Factor 1 explained 27% of the study variance. 6 participants are significantly associated with this factor. 4 are female and 2 are male. Apart from 1, all share an average age of 20. They all had limited or no experience of eliciting monitoring and evaluation work and came into the evaluation framework with limited knowledge.

The viewpoints from this factor indicate that the monitoring and evaluation competency of the practitioners has increased significantly from being involved in the framework (1+3). It supports the importance of capacity building through the workshops and support to fulfil monitoring and evaluation tasks. This also indicates the value placed around practitioners having more responsibility in monitoring and evaluation but relative to other factors those loading here scored lower (30+1) compared to those in factors 2 and 3. This is an interesting finding given that there are questions within the sport for development literature regarding how involved practitioners should be in monitoring and evaluation. This supports the view that they should be involved yet there is not a strong view that monitoring and evaluation should be a key part of a practitioner's role. Leading on from this, it is viewed as necessary but not excessive to have monitoring and evaluation continual professional development embedded in the support and development of a practitioner role (15+1). Thus it could be

concluded that despite the increased competency these practitioners are cautious about how much monitoring and evaluation should be embedded perhaps due to the time constraints and additional workloads (29-2) yet they do feel considerably more confident about doing monitoring and evaluation in the future (2+4). In terms of distance travelled they have clearly progressed significantly from a limited competency in monitoring and evaluation partly due to the independence and responsibility gained from running their own project, applying monitoring and evaluation in practice, and appreciating how gaining a more in-depth knowledge of a project can benefit learning. Furthermore, their enthusiasm and interest in sport for development are key factors in this.

Realist evaluation competency has improved (3+3) which is further justified by practitioners indicating their improved understanding of mechanisms within their project (5+4). This is not only a key tenet of realist evaluation but also one of the most problematic wherein the realist community many find it hard to uncover mechanisms. Practitioners have shown clear engagement with this in this factor (13+2) despite their neutral views of the importance of realist evaluation as an evaluation methodology (11: 0) and intent to use it in the future (14-1). This is further supported with other key conceptual elements that support the realist evaluation process such as the importance of developing programme theories of change (6+2), seeing the value in developing CMO configurations (4+2) and simply understanding more about how and why their intervention worked (12+2).

Despite being able to mobilise realist evaluation and see its benefit, practitioners within this factor are less reflexive and critically engaged with the monitoring and evaluation discourse compared to those particularly in factors 2 and 4. They have identified knowledge regarding the monitoring and evaluation concepts they were using but they do not appear to demonstrate any critical issues or strong feeling about realist evaluation which is reflected in their neutral view of realist evaluation (11:0) compared to other factors. This may suggest that although heavily involved, they have relied to some extent on others to overcome issues and specific barriers imposed by the evaluation approach because they lack that deeper engagement. Linked to this, there is a lack of confidence that the findings of their evaluation will have any real impact on the views of external partners (24-4). This is partly because the practitioners are not as critical or aware of the limitations of that landscape and what constitutes evaluation practice within it. They are less likely to enhance the use of the evaluation on any transformational level concerning change.

As far as the collaborative dimension of the framework goes, those within this factor indicate (18-2) that the structure of the framework did not necessarily provide a clear grasp of monitoring and evaluation. Yet, they rank higher (relative to other factors) the statements associated with the two models (7+1; 8:0). This suggests a positive engagement with the structure and the need for support, and guidance based on the contextual needs. Statements 6+2 (developing programme theories of change), 5+4 (improved understanding of mechanisms) and 4+2 (valuing CMO configurations) would support this view, as would the demographic information available which suggests that most of these participants were not the overall driving forces behind their intervention. Those in this factor required more control, support and special attention to facilitate the process and understand how to conduct the monitoring and evaluation. Like other factors practitioners loading into this one struggled to control (27-4) the mobilisation of the evaluation in practice. They felt in control to a degree but became anxious when confronted with practical issues and conceptual barriers.

Based on the Q factor interpretation and interviews a selection of refined CMO configurations are displayed below in accordance with the evaluation capacity building framework programme theory for factor 1. These CMOs are displayed in table 4 in line with Dalkin et al's (2015) dissagregation of the programme mechanism.

Insert table 4 about here

Refining programme theory

The iterative methodological process employed within the evaluation offered significant insight for refining the programme theory underpinning the evaluation capacity building framework moving forward. As Pawson and Tilley (1997) highlight, any testing of a programme theory should lead to the refining of that programme theory for further refining. In the case of the evaluation illustrated in this paper, the testing phase of Q (Q sort, factor analysis and qualitative interpretation) reliably informed the refinement stages. What was useful about Q was the way in which the factors were able to provide the 'how' and 'why' insight into the different typologies of practitioners which had their own underlying contexts mechanisms and outcomes. Whilst only one of the factors has been described in this paper it is clear that the framework worked for the practitioners in different ways and this was reflected by the four factors (subgroups) of practitioners who shared varying viewpoints. This was so valuable for moving the framework forward for further testing in knowing that within a similar context it should be delivered nuanced and in recognition of the varying viewpoints. For example, it should be delivered in a specific way in line with the subjective viewpoints of the types of practitioners in factor one and so on for the other factors.

In essence, the realist evaluation supported by Q provided a typology of four different types of practitioner reasoning and reaction to the evaluation capacity building framework, each with their own comprehensive compilation of CMO configurations representing the differing programme theories at play (Harris, 2018). These nuanced and specific CMOs represented a nuanced and contextual refinement to the framework moving forward for further delivery and testing. These refinements captured the changes, learning and reflexivity needed to inform its future delivery. Given the cyclical nature of a realist evaluation, naturally these CMOs and subsequent testing of this programme theory would inform the concourse of a Q informed realist evaluation if it were to be mobilised in the same way again.

Reflections on the embedding of Q within a realist evaluation

The preceding sections have sought to justify the case for adopting Q as a suitable methodology within a realist evaluation. As highlighted, at the heart of a realist evaluation lies the intention to make sense of how and why a programme may work for whom in certain circumstances (Pawson and Tilley, 1997). Here lies the intention in any realist evaluation to dig beneath the surface to make sense of the generative mechanisms of change at play within a programme or intervention. On reflection, the mobilisation of Q as a method within the realist evaluation case study of this paper was able to support the three key stages of developing programme theory, testing it and then making refinements.

Regarding programme theory development, the concourse of Q methodology complemented this process of taking into consideration all of the circumstantial and contextual variables to

verify what constituted the provisional programme theory for testing. Given the importance of context within any realist evaluation, and specifically in the conjecturing of a provisional programme theory, the attention given to the concourse stage enhanced this process significantly. The context rich characteristics involved in the development of the concourse enabled a deeper illumination of the provisional programme theory for testing through this 'concourse driven' stage. In the case of this evaluation the capacity building framework had already been delivered, and as such a candidate programme theory was already elucidated. However, the concourse stage provided great insight to capture and enhance an appreciation of other crucial contextual aspects for consideration in the programme theory. This was able to drive the next stage of theory testing via the formation of the Q set statements which reflected the programme theory to be put under test. This is a powerful synergy between realist evaluation and Q.

More specifically, using Q methodology to test the programme theory instilled confidence in the evaluator in that all of the relevant aspects of the candidate programme theory were being tested via the Q sorting exercise. This process led nicely into the logic of abduction where the shared viewpoints generated from the factor analysis were able to facilitate the building of theory to explain how and why these shared viewpoints took shape, and were supported by the interviews. Q's ability to capture the shared viewpoints was also a powerful way to capture specific contexts and illustrate (in the case of this evaluation) different typologies of people. In essence, the case study shown here was able to tell a story of how and why, under what circumstances different groupings (factors) of practitioners engage with evaluation practice (Harris, 2018).

Leading quite nicely on to programme refinement, using the conventional CMO configurations to support the generation of the holistic narratives was a powerful way to use the abductive theory building of Q's qualitative interpretation to reliably inform this CMO construction based on each factor. Naturally, this iterative process was able to verify and refine the initial programme theory for testing, and any improved, nuanced delivery of the framework in the future.

However, despite the many positives of embedding Q within a realist evaluation, one must acknowledge some potential limitations in their alliance. The first point concerns the generic process of the Q sorting exercise which requires the participants to place and sort their items accordingly into the piles of agreement and then within the Q grid. Participants need to be substantially involved in this process to understand how something like Q works (which is new to many of them). This also requires time to explain and enable the process to take shape and allow the participants to really think about the statements they are sorting. This of course impacts of the reliability of the sorting exercise to mitigate participants placing items anywhere on the grid because they are in a rush, or confused. Whilst in the case of Harris (2018) this was feasible through the workshop environment with limited numbers, in different contexts with a larger sample it may be more challenging. Therefore, within the realms of practicalities it is important for future researchers and evaluators to consider how much time needs affording to this process. In addition, certain demographics may engage with the process more than others so this should also be considered. The second point to consider is that Q may not always be entirely appropriate for all realist evaluations. Based on the practicality point, a time sensitive realist evaluation with a limited time frame may not allow for a Q (given its extensive process) and any researcher should also consider the potential of other innovative methods (Dalkin et al 2018) to serve as appropriate ways to capture generative causality. The final limitation to consider concerns the point that Q is a

methodology in itself with a rich heritage and philosophical depth. This isn't a weakness of Q per se but more so a recommendation for realist researchers and evaluators to entirely understand the key processes of Q that underpin the concourse, q-set, p-set, factor analysis and qualitative interpretation. This really does require comprehensive understanding which again takes time. If not actioned or understood properly, a Q centred realist evaluation could be problematic.

Conclusions

Overall, this paper has attempted to showcase (using a portion of data from another study) how Q methodology can support and mobilise a realist evaluation. Limited research has been published to date illustrating the compatibility of realist evaluation and Q. This paper has attempted to capture the philosophical compatibility of the two and brought this to life through the iterative methodological process that was followed. The case study chosen for this paper has attempted to showcase how the methodological process followed was able to meet the philosophical aims of a realist evaluation which is to make sense of how and why a programme / intervention works.

It is intended that this paper can add to and support some of the recent innovations in realist evaluation (Dalkin et al, 2018; Manzano, 2016; Shearn et al, 2017). Whilst it is acknowledged that Q may not always be suitable for a realist evaluation, it is another option for a realist researcher / evaluator to consider.

In moving this forward it is recommended that more capacity building and promotion is carried out with the realist academic community to learn about Q and how it can be integrated. It is also important to present the integration of the two approaches carefully and diplomatically within their bespoke areas to ensure there is no misinterpretation of their philosophical compatibility.

References

Baptiste Lennise (2011), "What educators learn when they evaluate students "*, Operant Subjectivity: The International journal of Q Methodology,* Vol. 34, No. 2, pp. 104-123.

Barry, J. and Proops, J. (1999), *Seeking sustainability discourses with Q methodology*.

Brown, S.R. (1980), *Political subjectivity*, 1. printing ed., translated by Anonymous Yale Univ. Press, New Haven [u.a.].

Coalter, F. (2013), "'There is loads of relationships here': Developing a programme theory for sport-for-change programmes"*, International Review for the Sociology of Sport,* Vol. 48, No. 5, pp. 594-612. available at: http://journals.sagepub.com/doi/full/10.1177/1012690212446143.

Dalkin, S.M., Greenhalgh, J., Jones, D., Cunningham, B. and Lhussier, M. (2015), "What's in a mechanism? Development of a key concept in realist evaluation"*, Implementation science : IS,* Vol. 10, No. 49, pp. 49. available at: http://www.ncbi.nlm.nih.gov/pubmed/25885787.

Evans, D. and Killoran, A. (2000), "Tackling health inequalities through partnership working: Learning from a realistic evaluation"*, Critical Public Health,* Vol. 10, No. 2, pp. 125-140. available at: http://www.tandfonline.com/doi/abs/10.1080/09581590050075899.

Gallagher, K. and Porock, D. (2010), "The Use of Interviews in Q Methodology: Card Content Analysis"*, Nursing Research,* Vol. 59, No. 4, pp. 295-300. available

at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&NEWS=n&CSC=Y&PAGE=fulltext&D=ovft&A
N=00006199-201007000-00009.

Wong. G, Westhorp, G, Manzano, A, Greenhalgh, T, Jagosh, J, and Greenhalgh, T. (2016),
"RAMESES II reporting standards for realist evaluations", *BMC Medicine,* Vol. 14, No. 1, pp.
96. available at: https://www.ncbi.nlm.nih.gov/pubmed/27342217.

Harris, K. (2018), "Building Sport for development practitioners capacity for undertaking
monitoring and evaluation - reflections on a training programme building capacity in realist
evaluation", *International Journal of Sport Policy and Politics,* pp. 1-20.

Harris, K. and Adams, A. (2016), "Power and discourse in the politics of evidence in sport for
development", *Sport Management Review,* Vol. 19, No. 2, pp. 97-106.

Hawkins, A. (2014), "The case for experimental design in realist evaluation", *Learning
Communities: International Journal of Learning in Social Contexts,* Vol. 14, pp. 46-59.

Hensel, D. (2017), "Using Q Methodology to Assess Learning Outcomes Following the
Implementation of a Concept-Based Curriculum", *Nurse Educator,* Vol. 42, No. 5, pp. 250-
254. available at: https://www.ncbi.nlm.nih.gov/pubmed/28045739.

Jagosh, J., Bush, P.L., Salsberg, J., Macaulay, A.C., Greenhalgh, T., Wong, G., Cargo, M.,
Green, L.W., Herbert, C.P. and Pluye, P. (2015), "A realist evaluation of community-based
participatory research: partnership synergy, trust building and related ripple effects", *BMC
public health,* Vol. 15, No. 1, pp. 725. available
at: http://www.ncbi.nlm.nih.gov/pubmed/26223523.

Kelly, S.E., Moher, D. and Clifford, T.J. (2016), "Expediting evidence synthesis for healthcare
decision-making: exploring attitudes and perceptions towards rapid reviews using Q
methodology", *PeerJ,* Vol. 4, pp. e2522. available
at: https://www.ncbi.nlm.nih.gov/pubmed/27761324.

Lewis-Beck, M.S., Bryman, A. and Liao, T.F.(. (2003), *SAGE Encyclopedia of Social Science
Research Methods* translated by Anonymous SAGE Publications, Thousand Oaks.

Manzano, A. (2016), "The craft of interviewing in realist evaluation", *Evaluation,* Vol. 22, No. 3,
pp. 342-360. available at: http://journals.sagepub.com/doi/full/10.1177/1356389016638615.

Marchal, B., Van Belle, S., Van Olmen, J., Hoerée, T. and and Kegels, G. (2012), "Is realistic
evaluation keeping it's promise? A review of published empirical studies in the field of health
systems research", *Evaluation,* Vol. 2, No. 18, pp. 192-212.

Pawson, R. and Manzano-Santaella, A. (2012), "A realist diagnostic workshop", *Evaluation,* Vol.
18, No. 2, pp. 176-191. available
at: http://journals.sagepub.com/doi/full/10.1177/1356389012440912.

Pawson, R. and Tilley, N. (1997), *Realistic evaluation*, 1. publ. ed., translated by Anonymous
SAGE Publ, London [u.a.].

Pawson, R. (2016) 'The ersatz realism of critical realism: a reply to Porter', *Evaluation*, **22**, 1, 49–
57

Peirce, C.S. and Buchler, J. (1955), *Philosophical writings of Peirce*, 1. publ. ed., translated by
Anonymous Dover Publ, New York, NY.

Pierce, C.S. (1958), *Collected Papers (vols 1-8).* translated by Anonymous Harvard University
Press. • Peirce, C., Cambridge, MA.

Pike, K., Wright, P., Wink, B. and Fletcher, S. (2015), "The assessment of cultural ecosystem
services in the marine environment using Q methodology", *Journal of Coastal*

Conservation, Vol. 19, No. 5, pp. 667-675. available at: https://search.proquest.com/docview/1719537634.

Porter, S., (2015a). The uncritical realism of realist evaluation. *Evaluation*, 21 (1), 65-82.

Porter, S., (2015b). Realist evaluation: an immanent critique. *Nursing Philosophy*, 16 (4), 239-251.

Ramlo Susan (2017), "Improving Student Evaluation of Teaching: Determining Multiple Perspectives within a Course for Future Math Educators"*, Journal of Research in Education,* Vol. 27, No. 1, pp. 49-78.

Rycroft-Malone, J., Fontenla, M., Bick, D. and Seers, K. (2010), "A realistic evaluation: the case of protocol-based care"*, Implementation science : IS,* Vol. 5, No. 1, pp. 38. available at: http://www.ncbi.nlm.nih.gov/pubmed/20504293.

Schmolck, P. and Atkinson, J. (2010), *, PQ method software and manual* [2015, 10th May)].

Shearn, K., Allmark, P., Piercy, H. and Hirst, J. (2017), "Building Realist Program Theory for Large Complex and Messy Interventions"*, International Journal of Qualitative Methods,* Vol. 16, No. 1 available at: http://journals.sagepub.com/doi/full/10.1177/1609406917741796.

Smith, J.A., Harre, R. and Langenhove, L.V. (Eds.) (1995), *Rethinking Methods in Psychology*, Sage Publications Ltd, GB.

Stenner, P. and Rogers, R.S. (1998), "Jealousy as a manifold of divergent understandings: a Q methodological investigation"*, European Journal of Social Psychology,* Vol. 28, No. 1, pp. 71-94. available at: https://search.proquest.com/docview/1877069922.

Stephenson, W. (1953), *The study of behavior: Q-technique and its methodology* translated by Anonymous University of Chicago Press, Chicago.

Stephenson, W. (1982), "Q-Methodology, Interbehavioral Psychology, and Quantum Theory"*, The Psychological Record,* Vol. 32, No. 2, pp. 235. available at: https://search.proquest.com/docview/1301194158.

Stephenson, W. (1986), "Protoconcursus: The Concourse Theory of Communication"*, Operant Subjectivity: The International journal of Q Methodology,* Vol. 2, No. 9, pp. 96. available at: http://www.dx.doi.org/10.15133/j.os.1985.002.

Stergiou, D. and Airey, D. (2011), "Q-methodology and tourism research"*, Current Issues in Tourism,* Vol. 14, No. 4, pp. 311-322. available at: http://www.tandfonline.com/doi/abs/10.1080/13683500.2010.537316.

Susan E. Ramlo, David McConnell, Zhong-Hui Duan and Francisco B. Moore (2008), "Evaluating an Inquiry-based Bioinformatics Course Using Q Methodology"*, Journal of Science Education and Technology,* Vol. 17, No. 3, pp. 219-225. available at: https://www.jstor.org/stable/41219415.

Watts, S. and Stenner, P. (2012), *Doing Q methodological research*, 1. publ. ed., translated by Anonymous Sage, Los Angeles [u.a.].

Watts, S. and Stenner, P. (2005), "Doing Q methodology: theory, method and interpretation"*, Qualitative Research in Psychology,* Vol. 2, No. 1, pp. 67. available at: http://www.tandfonline.com/doi/abs/10.1191/1478088705qp022oa.

Westhorp, G. (2014), "Realist Impact Evaluation - An Introduction"*, A method lab publication,* .

FIGURES AND TABLES

Figure 1: Methodological design

Figure 2: Example of Q grid

Table 1: Q stages

Stage	Summary
1. Review the concourse	According to Brown (1980) this is the most crucial stage of the Q process because it is here where knowledge is gained and accrued about the field or context. Using primary, and or secondary research, increased knowledge regarding the context drives the derivation of statements (Watts and Stenner, 2012). The process for creating a Q sample begins by gathering communications that represent the wide range of opinions, attitudes, or beliefs of a group with a shared understanding of a phenomenon of interest. While these communications are most often written statements, they can also be non-linguistic in nature including aromas, photographs, objects, and music (Hensel, 2017, pg.251).
2. Select Q-Set sample of items / statements	To enable data collection the concourse is reduced to a manageable number of items. The Q sample is the selected subset of the concourse presented to participants for sorting (Hensel, 2017, pg.251). This stage involves the generation of a series of (Q-set) statements / items. A Q-Set in practice is usually made up of 30-80 statements (Watts and Stenner, 2005; Stainton Rogers, 1995).
3. Select P-Set / participants	Having developed the Q-Set this stage involves the selection of a P-Set which is a representative sample of participants within the research.
4. P-Set sort items into a subjectively meaningful pattern	This stage requires the P-Set to rank the statements presented to them based on their subjective viewpoints. This is initially done by sorting the statements into piles of agree, disagree and neutral. These are then plotted onto a bell curve / forced distribution which has become the standard approach for Q methodologists because it enables a pragmatic and convenient means of facilitating the subjective evaluations (Watts and Stenner, 2012).

5. Factor analysis	Following sorting by the P-Set, ==by-person== factor analysis is carried out through the use of statistical programmes (such as PQ Method) where the sorts are entered accordingly. The programme looks for groups of persons who have rank ordered the statements in a similar fashion and share similar perspectives (Watts and Stenner, 2012). Factors are then produced for interpretation.
6. Interpretation	Factors are selected and interpreted to a high level of qualitative detail ==(full qualitative interpretation)==. This is usually mobilised via the identification of statements ranked the highest, lowest within each factor and in relation to other factors. Through the logic of abduction (Watts and Stenner, 2012) a story begins to emerge based on this interpretation. In addition, interviews may also be conducted (Pike et al, 2015) to clarify the emerging story surrounding those viewpoints within a specific factor.
7. Production of holistic narrative	Factor analysis and interpretation enables the production of a holistic narrative which portrays a story about the shared viewpoints within each factor.

Table 2: Q set statements for sorting

1. My competency for carrying out M and E has improved	2. I feel more competent about undertaking M and E in the future
3. I feel more competent in carrying out Realist Evaluation	4. I can see the value in developing context – mechanism – outcome configurations to aid my evaluation
5. I feel that the framework has made it possible for me to identify mechanisms of change in my project	6. I think that it is important to develop a programme theory before an evaluation is carried out
7. Phase 1 enabled me to explain my programme theory effectively	8. Phase 2 provided me with a clear set of steps to follow for conducting a realistic evaluation
9. I am more reflexive now about the processes I take to M and E programmes / projects	10. Working with my fellow practitioners in my project enabled me to understand RE
11. Realist evaluation is an appropriate evaluation methodology for my project	12. I now understand more about how and why my project worked
13. I can now see the need to understand how and why SFD programmes work	14. I would implement realist evaluation in future roles
15. I feel that M and E training should be a key part of a practitioners professional development	16. The resources (eg literature) surrounding RE enabled my understanding of the evaluation methodology
17. I felt that my relationship with	18. I feel that the structure to the M and

the trainer enabled me to progress to understand how to M and E my project	E process through the workshops enabled me to carry out my M and E
19. I felt engaged in the M and E	20. I felt that my views within the collaborative process were respected
21. I felt that throughout the M and E training programme I was able to highlight any issues or tensions	22. My role within the M and E was entirely clear
23. The trainer and I shared the understanding of the evaluation goals we had	24. I feel that the findings from our M and E will be valued by our partners
25. The training framework was suitable to meet the needs of my project	26. I was able to be apply the knowledge of my project within the M and E framework
27. Throughout the training programme I felt in control of the evaluation	28. Without the structure of the training programme I do not feel I would have been able to do realistic evaluation
29. The training programme has been possible to complete given my additional workload	30. M and E should be embedded within a practitioner's role

Table 3: Differences amongst factors

(*M and E refers to monitoring and evaluation; RE refers to realist evaluation; PT refers to programme theory)

Factor 1 "From zero evaluators to accomplished"	Factor 2 "Polished Problem solvers"	Factor 3 "Passive passengers"	Factor 4 "Proficient and competent and cautiously optimistic"
Differences between factors			

-Relied on support more than others to get through -The capacity building and support enabled them to get the job done -M and E competency and RE competency increased significantly (more so than other factors) -Despite competency still lack confidence to go alone -Less reflexive towards M and E -Less transformational on all levels -Increased understanding of mechanisms	-Use of the M and E a prevailing characteristic -Critical of RE but able to solve problems -Conceptually engaged in RE -Likely to implement RE in future based on conceptual awareness and grasping -Leaders / problem solvers -Cynical and negative about their findings having and impact -Critical of industry -Utilised semi autonomous control and support to aid use -As a result less enthusiastic about M and E role in the future	-Less critical and aware of issues in M and E -More attuned to coaching -Surface level approach to engaging in M and E -Looked to other members -Find it hard to synthesise project learning with M and E findings -Struggle with RE jargon -No significant distance travelled	-Value relationship with the trainer more so than other factors -More cautious of RE and may not favour it as an evaluation approach because of its compatibility with the context -Less likely to carry out RE in future -Apprehensive about developing PT -Strong belief in findings

Similarities across factors

-Practitioners should be focused on monitoring and evaluation (factors 3,2)
-Practitioners struggle to control the evaluation in practice (factors 1,2,4)
-Critical engagement with monitoring and evaluation (factors 4,2)
-Limited confidence in monitoring and evaluation findings (factors 3,2,1)
-Practitioners can apply monitoring and evaluation through the kinaesthetic act of running their CIP (factors 4,3,2,1)
-Practicality is more pertinent than empowerment (factors 4,2,1)
-Practitioners appreciate and have deeper realisation for how and why projects work (factors 4,3,2,1)
-Monitoring and evaluation competency is enhanced (4,2,1)
-Knowledge and understanding of realist evaluation has improved (4,2,1)
-Practitioners are keen to do well within the curriculum (4,3,2,1)

Table 4: Factor 1 CMO configuration (capturing refined programme theory)

Context	MResource	MReasoning	Outcome
Within a given Context	**A resource is introduced (MRES)**	**Enhances a change in reasoning / response (MREAS)**	**Producing Outcome**

Limited knowledge and understanding of monitoring and evaluation and realist concepts	Workshops (focus on mechanisms) and CIP	Contextualising real subjects within CIP created a light bulb moment of realisation that implementation resources are not solely mechanisms and participants reason to resources	Deeper understanding of generative mechanisms of change
Challenging environment for practitioners to do monitoring and evaluation	The CIP project / workshops	Created a sense of responsibility and independence concerning the importance of evidence	Appreciating the importance of monitoring and evaluation
University based project led by in-experienced student practitioners	Presentations (particularly those with activities focusing on partner collaboration)	Weak confidence in findings because of a negative perception of students as practitioners. This evoked little thought provocation or awareness of involving partners more within the monitoring and evaluation	Poor take up of findings by stakeholders / partners
Students who find lots of information and lectures challenging	Broken down smaller chunks workshops offered throughout life cycle of their CIP	Fostered the feeling that they could steadily clarify understanding with the trainer before progression to the next stage / pathway	Evaluated their project effectively
Inexperienced practitioners delivering complex interventions	Framework and appropriate steps to follow combined with repetition of workshops	Felt in control of doing the evaluation but needed and could call on support over time	Increased competency in monitoring and evaluation and more confident about doing it in the future