Involving stakeholders in programme theory specification: discussion of a systematic, consensus-based approach

Felix C. van Urk, Sean Grant, Chris Bonell

Introduction

The importance of theory-based intervention design and evaluation has been recognised for decades (Chen & Rossi, 1983; Michie et al., 2005; Rogers, 2007; Weiss, 1997). Pawson and Tilley (1997) argue that all interventions are per definition 'theories incarnate' because they initiate a particular causal process leading to particular impacts. As a result, explicit use of programme theory greatly improves the informative power of evaluation research. This is particularly pertinent for complex social interventions, as clarifying how prescribed intervention components are supposed to facilitate mediating change processes and how the process is moderated by contextual factors prior to a full-scale evaluation can provide important information about the intervention for evaluation design (Craig et al., 2008). Recent guidance on developing complex interventions therefore suggests that explicit theories of change be used to inform their design (Campbell et al, 2007; Craig et al, 2008; Glanz & Bishop, 2010). Researchers furthermore stress the important of using explicit theory within evaluation because it facilitates the contribution of trials to the testing and refinement of generalisable programme theories across contexts and populations, instead of merely specific intervention programmes (Bonell et al., 2012; (Michie & Abraham, 2004; Michie & Prestwich, 2010) Further benefits include an improved understanding of why interventions are effective or ineffective, and potentially a greater influence of evaluation results on popular opinion and policy Weiss, 1995). There is increasing evidence that theory-based intervention may be more effective (Glanz & Bishop, 2010; Noar et al., 2008).

In spite of these recommendations, reviews have shown that studies often fail to explicitly apply and test theory, and in some cases researchers or evaluators fail even to refer to programme theory (Gardner et al., 2010; Noar & Zimmerman, 2005; Trifiletti et. al, 2005). This does not imply that such studies evaluate interventions or programmes that have no basis in theory: even when the theory of change and the structure of a programme is not specified explicitly, they are often present implicitly (Chen, 1990). Nevertheless, to systematically guide a process- or outcome evaluation, as well as programme implementation, an explicit theory of change is preferable (Connell & Kubisch, 1998; Eldridge et al., 2005). Given the intricacies of programme theory for some interventions, the specification of logic models (i.e., graphic depictions of a theory of change) has been recommended for complex interventions, as these models make programme theories easier to comprehend by a variety of stakeholders (Cooksy et al., 2001; Trevisan, 2007).

Elucidating implicit programme theory and building logic models

It is not uncommon for evaluators to specify the programme theory that underlies an intervention prior to evaluation (Connell & Kubisch, 1999). When developing programme theory and building logic models, Patton (2008) distinguishes between three overarching approaches: deductive, inductive, and user focused. *Deductive approaches* involve researchers using previous empirical research and dominant theories from different fields to inform programme theory development. This is typically how researchers develop programme theories before a programme is

implemented. If executed rigorously, these approaches hold the advantage of being rooted in systematic scientific enquiry. However, they do not allow for the integration of implementation experiences as well as contextual factors into their models. When using an *inductive approach*, researchers generate programme theory by observing the programme as it is being implemented and reviewing documentation (Oosthuizen & Louw, 2013; Savaya & Waysman, 2005). Although not always grounded on a large body of prior research and theoretical frameworks, inductive approaches allow researchers to base their programme theories on data from the world of actual practice. For example, Savaya and Waysman (2005) reconstructed the programme theory of a contact-resuming training programme for parents who had been separated from their children. Specifically, they assembled logic model components based on government documents describing programme expectations, constructed a coherent set of items, and finally examined how these could be linked based on the documentation.

Within *user-focused approaches,* evaluators collect information from programme staff through interviews or surveys and use the data to construct a programme theory. When employing such approaches, some evaluators actively involve a programme's practitioners, stakeholders, or purveyor organisations in elucidating an implicit theory of change through interviews and consultation processes (Christie et al., 2003; Julian, 1997; Oosthuizen & Louw, 2013; Sullivan et al., 2002). When Yampolskaya and colleagues (2004) elucidated the programme theory of a child mental health programme, for example, programme staff members and coordinators were involved in group brainstorming and a rating process to arrive at a logic model that was subsequently reviewed again by participants. Such participatory efforts may improve the accuracy of specified programme theories as well as increase stakeholder commitment to these theories and subsequent evaluations.

Unfortunately, the specific methodological processes for elucidating programme theory are rarely explicitly reported. Consequently, it is often unclear exactly how information gained from deductive, inductive, and user-focused approaches resulted in a final product (e.g., a logic model). This lack of transparency makes it difficult for practitioners, evaluators, and other stakeholders to assess the validity of programme theories and prevents the accumulation of knowledge on methods to elucidate these theories for future research and evaluation.

The Delphi method

The Delphi method is a consensus development technique that involves facilitated, structured interaction amongst individual stakeholders informed on a given topic area. A series of questionnaires are presented in an iterative, multi-round process to systematically collect, aggregate, and present to a panel of stakeholders the group's opinions or judgments on specific questions and issues related to the topic area of interest (Brown, 1968; Hasson et al., 2000; McKenna, 1994). Three strengths of the Delphi technique are:
   i)   the absence of face-to-face contact, which allows interaction of a geographically sparse group of stakeholders while avoiding potential pitfalls of in-person discussion (e.g., dominance of discussion by overly assertive individuals, reticence by introverted individuals);
   ii)   the formal feedback of group views to participants after each questionnaire, which facilitates reconsideration of previously expressed views after comparison to others' responses;

iii)   the explicit aggregation of individual responses, which provides an auditable trail of how an end product is the result of stakeholder consultation (Murphy et al., 1998).

Its pragmatic utility has made the Delphi a popular method in the social and health sciences. The Delphi method has been used for a variety of purposes, including the determination of core outcomes to measure in randomised controlled trials (RCTs), setting research priorities in educational technology, and developing reporting guidelines for health research (Hopewell et al., 2008; Pollard & Pollard, 2004; Sinha et al., 2011). In the current study, we combined inductive and user-focused approaches to elucidate the programme theory for a novel model of secondary education: Studio Schools. Namely, we incorporated the Delphi method into an explicit and transparent methodological process of incorporating stakeholder in programme theory development. The objectives were to develop a logic model for Studio Schools that will guide a subsequent evaluation, and to assess if incorporating the Delphi method in this process could facilitate stakeholder participation in logic model development and support for use of the logic model in subsequent evaluations.

Method

A two-round Delphi process was integrated between into a systematic effort to involve stakeholders in elucidating the theory of change of Studio Schools that took place from December 2012 to August 2013. Participants were emailed a link to the electronic questionnaires at the beginning of each round. Please see the online supplementary materials for an overview of the questionnaires used and round-by-round report of results. This study was reviewed and subsequently approved by the Departmental Research Ethics Committee, Department of Social Policy and Intervention, University of Oxford (Ref: 2012-13_06).

Studio Schools

Studio Schools is a new model of state-funded secondary education in England for young people of all academic ability levels between the ages of 14 and 19 years old. The schools are designed to tackle youth disengagement with secondary education as well as a shortage of transferable employability skills among secondary school leavers. To do so, Studio Schools offer a contextualised approach to learning that is centred around enterprise projects, personal coaching and students gaining work experience (Studio Schools Trust, 2012). Students are prepared for public exams in core academic subjects and spend a significant portion of their time doing work placements. Studio Schools are the product of a partnership between The Young Foundation and Edge, two organisations devoted to challenging inequality and promoting practical and vocational learning in the UK. The schools are administratively categorised as Free Schools and primarily supported and promoted by the overarching Studio Schools Trust (SST), which functions as a linking point between schools and works closely with the English Department for Education and other nationally operating organisations. The first Studio Schools were opened in 2010, and there are currently 36 schools in operation.

Participants

To be considered as stakeholders eligible for the current study, participants had to be i) originators of the Studio Schools approach, ii) responsible for its promotion and facilitating its implementation, or iii) implementors of the model in schools (i.e., principals of Studio Schools). Ten individuals (four female, six male)

were approached to take part in the study, based on their nomination during preliminary interviews with both the originators of the Studio Schools approach as well as members of the Studio Schools Trust. Three of these had shared some responsibility for developing the original Studio Schools concept. A further three participants were current members of the Studio Schools Trust and responsible for the model's dissemination across England. Finally, four participants were current principals of Studio Schools. All approached individuals agreed to participate and provided written consent.

Procedure
Preliminary interviews
        Prior to commencing the Delphi process, the first author held semi-structured interviews with the three participating originators and developers of the Studio Schools approach and two members of the Studio Schools Trust. These interviews provided an understanding of the background and context of the Studio Schools development process, and helped identify potential Delphi participants. The interview data were also used to inform the development and analysis of the Delphi questionnaires.

Delphi process
-Round One
        The Delphi process started with an 'open' first round. That is, participants were sent an open-ended questionnaire asking them to nominate, in their own words, the most important features of the Studio Schools model (see Supplement A, available online). With no explicit logic model to draw on, an open first round was deemed important so that the theory of change could be elucidated by stakeholders of the Studio Schools model rather than being developed and later imposed by the research team. Participants had four weeks to complete Round One.
        At the start of the questionnaire, participants were provided an introduction explaining the concept of a logic model and the purpose of the study. The rest of the questionnaire was structured by different elements that form part of a logic model ('objectives', 'inputs/resources', 'activities', 'outputs', 'proximal outcomes' and 'distal outcomes' of the programme). Participants were asked to list all components of the Studio Schools model, providing a short explanation (100 words) of what is meant by each feature and about the importance of each feature. As it was assumed that not all respondents would be familiar with the elements of a logic model, one hypothetical example of a Studio Schools feature was given per logic model element (e.g., project-based learning for 'activities'). There was a specified limit (n=8) of items respondents could suggest per logic model element, informed by initial interviews which suggested this was an appropriate number.
        The first two authors independently analysed and integrated responses into a list of non-redundant items, structured by the elements of a logic model. Disagreements between the researchers were resolved following discussion, after which a final list of items was constructed. This list was sent to all participants prior to their rating of items in Round Two. It was accompanied by each participant's original responses and a description of how each was integrated into the list, so that participants could understand the process leading to the development of the Round Two questionnaire and add or rephrase items if deemed necessary. Participants were not provided with the individual responses of any other panel member: they only saw

others' responses in aggregate form. The process of integrating participants' responses and preparing the feedback reports took four weeks to complete.

- Round Two

The second round consisted of respondents filling out a closed-response questionnaire (see Supplement C, available online), and took nine weeks for all participants to complete. This questionnaire contained the consolidated list of items that were identified in Round One, and it was structured by the same logic model elements as the Round One questionnaire. Respondents were asked to rate the relevance of each nominated item to the Studio Schools model on a 1-10 Likert-scale, with a rating of '1' representing a judgment of an item as 'not at all relevant', and a rating of '10' representing a judgment of an item as 'highly relevant'. In addition, participants were given the opportunity to provide written clarification for their rating of each item.

Ratings were summarised as simple descriptive statistics (medians, modes, lowest and highest values). Ratings were then analysed using adapted criteria from the RAND/UCLA appropriateness method manual, which classifies items in accordance with i) their median panel rating and ii) some measure of the dispersion of panel ratings, which is taken as an indicator of the level of agreement with which the ratings were made (Fitch et al., 2001). Upon receipt of participants' responses, the decision was made to use the 'A7S' procedure for classification of items in order to preserve a strict definition of consensus while minimising the influence of extreme ratings by outliers. According to this definition of agreement (Fitch et al., 2001), an item was considered to have consensus if, after discarding the most extreme high and the most extreme low rating, the remaining ratings all fell within the same 'range category' of 1-4, 5-7, or 8-10. Adapting this procedure for our panel size of 10, initially items could be classified as included, excluded, or uncertain (see Table 1 for a description).

Table 1. *A7S Criteria for the Classification of Items after Delphi Round Two Ratings*

by Participants (adapted from Fitch et al., 2001)

Item rating after discarding highest and lowest rating	Decision
Value of 8 or higher by all panel members	Included
*Median value ranges from 5-7 *Values do not all fall into a single range category (1-4, 5-7, 8-10)	Uncertain
Value of 4 or lower by all panel members	Excluded

Using Delphi Results to Create a Logic Model

Items classified as 'uncertain' from the quantitative decision rules above were then either included or excluded from the logic model by the first author, based on data obtained from: i) numerical ratings and qualitative responses provided by participants in the first- and second rounds of the Delphi process, ii) semi-structured

interviews that were conducted with some panel members prior to the first round, and iii) documentation provided by the Studio Schools Trust and the Studio Schools online network. Items were included if explicit support for their classification as a necessary feature of the Studio Schools model could be found in at least one of the sources above. Items were excluded if such support was absent from all of the sources above.

Following this selection process, the first author created a logic model linking the included items. Links (relationships) between items were again based on explicit support for such relationships in the same sources used for the inclusion or exclusion of 'uncertain' items described above: i.e., Delphi participant responses, semi-structured interviews, and official Studio Schools documents. The first author then wrote a report describing the Delphi process and its rationale, and depicting the proposed Studio Schools logic model. All participants and the second coder (SG) were sent the report, given forms to provide feedback on the Delphi process and the proposed logic model, and asked to suggest changes if they felt these were needed. Feedback led to a revised draft logic model.

Results

Nine participants completed the Delphi Round 1 questionnaire and nominated items for inclusion in the Studio Schools logic model. Ten participants completed the Round 2 questionnaire, rating and commenting on proposed logic model items (see Figure 1).

Figure 1. Flow of Participants and Items through the Study
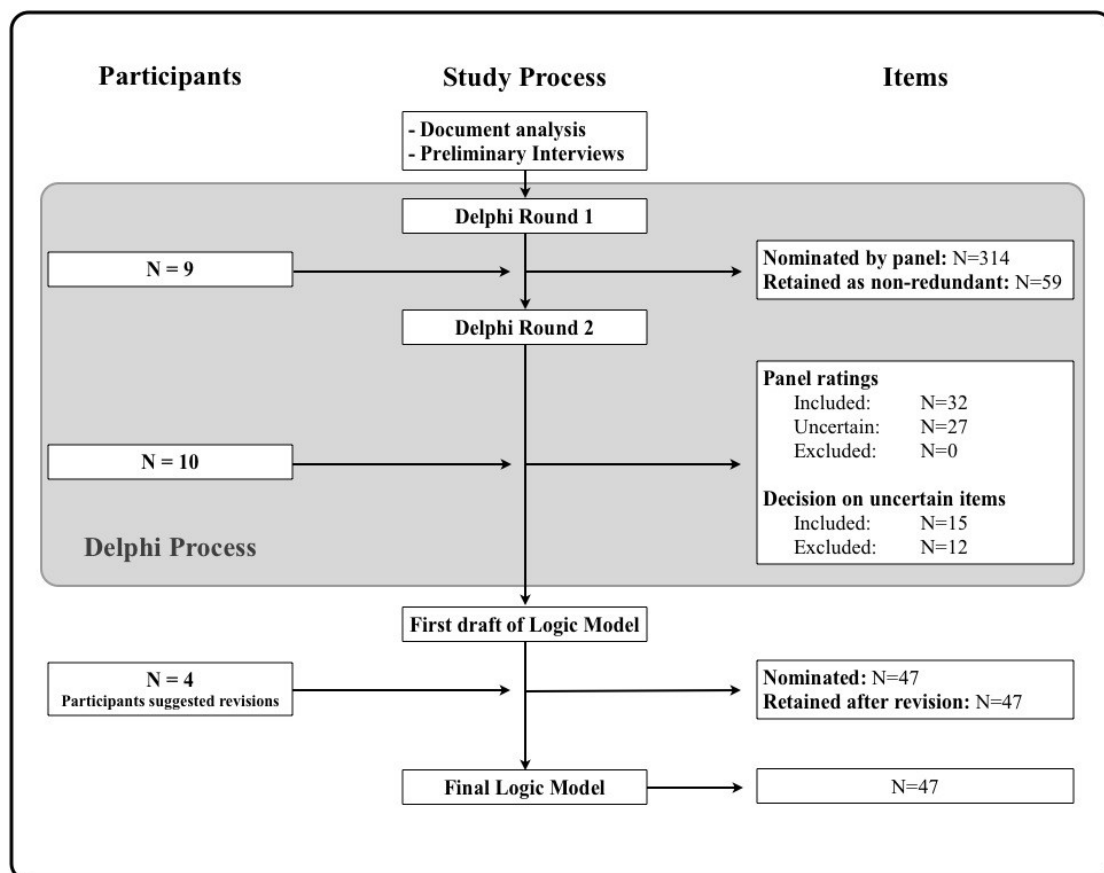
Delphi process

Participants nominated a total of 314 unique items in Round One. The majority of qualitative comments explained why and how nominated items fit into the overall Studio Schools model (e.g., how an 'objective', 'input', or 'activity' is important to the development of students or facilitating other components of the model). In some cases, participants clarified or elaborated on their meaning of a nominated item; for example, what they specifically meant by an item in the context of Studio Schools (for a complete list of all nominated items including explanations and elaborations by participants, see Supplement B). After independent integration of items into a non-redundant list by the first two authors, a total of 59 items were rated by participants in Round 2. Upon receiving the integrated item list and their original responses prior to the start of Round 2, participants did not express any concerns regarding the validity of the list.

Participants rated 59 items in Round Two. Median ratings for each item ranged from 6 to 10, though individual participant ratings spanned the full range of 1 to 10. When commenting on ratings, participants mostly elaborated on relatively low ratings. In some cases, participants explained why an item was not essential to the Studio Schools model, or was only relevant to some schools. Some participants were also concerned that, although relevant, an item might not be unique to the Studio Schools model. Finally, participants occasionally explained that their relatively low rating for an item was the result of the item being phrased either slightly off-point or too exclusively for some students at Studio Schools.

Table 3 presents the classification of items after Round Two. Overall, 32 items were classified as included based on participants' ratings, according to the adapted A7S procedure (Fitch et al., 2001). A further 27 items were classified as uncertain. No items were classified as excluded at the end of the second round.

Table 3. *Classification of Items Based On Ratings in Delphi Round Two*

Item category	Rated as included (all ratings 8-10)	Rated as uncertain (4 < median < 7 OR ratings do not fall in same category)	Rated as excluded (all ratings 1-4)
Objectives	3/7	4/7	0/7
Inputs	10/13	3/13	0/13
Activities	5/15	10/15	0/15
Outputs	2/6	4/6	0/6
Proximal Outcomes	8/10	2/10	0/10
Distal Outcomes	4/8	4/8	0/8

Total	**32/59**	**27/59**	**0/59**

Creation of Studio Schools Logic Model

Of the 27 items classified as uncertain at the end of Round Two, 15 items were included by the first author based on support found in Delphi participant responses, semi-structured interviews, and official Studio Schools documentation. The remaining 12 items were excluded because no explicit support for these was found in these sources. An example of the process of final classification of 'uncertain' items nominated under 'outputs' is presented in Table 4.

Once a list of included items was defined, and the first author developed a first version of a logic model, some items were integrated under common themes to preserve a logically coherent model. For example, some core activities such as 'project-based learning' and 'personal coaching' of students were integrated under the theme 'school-based curriculum'. After the report describing the Delphi process and the first draft of the logic model was distributed to all 10 participants, 4 of these requested some items to be disintegrated from common themes, as these were considered distinct operating functions of the model. These participants represented key stakeholders for the subsequent evaluation phase. Revisions were made accordingly and then the logic model was re-circulated. Participants expressed no concerns regarding the final version of the logic model (see Figure 2).

Table 4. *Example of Final Decision Process on Inclusion and Exclusion of Items.*

Example Items Represent the Six Outputs Rated by the Panel in Delphi Round Two

Item	Description	Median	Range	Classific.	Decision	Reason
1	Work experience	9	8-10	Included	Included	Ratings
2	Positive social networks	9	8-10	Included	Included	Ratings
3	Personalised curriculum	10	7-10	Uncertain	Included	*'One of 7 essential elements of a Studio School'* (SST, 2013)
4	Appropriate academic qualifications	10	7-10	Uncertain	Included	Low panel ratings were due to concerns related to item description, which was revised

5	Portfolio	8.5	7-10	Uncertain	Excluded	No support found in consulted sources
6	Extra-curricular achievements	7.5	5-9	Uncertain	Excluded	No support found in consulted sources

<<INSERT FIGURE 2 (ACTUAL LOGIC MODEL) HERE>>

Discussion

Summary of results

This study aimed to explore and demonstrate the possibility of a transparent, systematic, and user-focused approach for developing an explicit programme theory for a complex social programme based on the input of its stakeholders. Specifically, the present study provides an example of elucidating a logic model for a novel educational approach—Studio Schools—via active and systematic involvement of those involved in creating, promoting, and implementing it. We incorporated an iterative consultation method in the form of a modified Delphi process, integrating participants' quantitative ratings and qualitative responses on components of the Studio Schools model into a coherent conceptual framework. This initial logic model is underpinning a subsequent formative evaluation of the Studio Schools approach; as such, it is not intended to be a definitive logic model to which each individual school must strictly adhere. Nevertheless, it specifies the core theoretical assumptions of the current Studio Schools model, which can be used to inform theory-driven evaluation research.

Strengths of this study compared to other approaches

A particular strength of this study is the use of a participatory, systematic, and auditable approach to clarifying programme theory that is founded on explicit consultation with stakeholders. Including a modified Delphi process as part of the development of programme theory in combination with other commonly used inductive approaches—such as face-to-face interviews and document analysis—has several advantages. First, although the item inclusion process was to some extent influenced by the researchers, items nominated by participants in an open first round and subsequent rating process formed the basis of the model. Moreover, participants received explicit descriptions of all parts of the process and were given the opportunity to provide feedback at all stages. The resulting logic model therefore likely describes Studio Schools' programme theory as viewed by its main stakeholders. This decreases the risk of incorrect interpretation by the researchers and underpins the intention of increasing buy-in by stakeholders in the conceptual framework underpinning future evaluations. Had we relied solely on our own judgement based on analysis of programme documents or common social science theories, as is usually done within inductive and deductive approaches, this would have likely not been the case. Even user-focused approaches that involve stakeholders extensively in the programme theory development process often involve them after a preliminary model, based on researcher judgement, has been drafted and circulated (e.g. Christie & Alkin, 2003; Savaya & Waysman, 2005; Renger, 2011),

carrying a risk that participants' perceptions are influenced by those of the researchers.

Second, the absence of face-to-face interaction during an online Delphi process, in combination with the structured and anonymous feedback of participants' responses, aims to decrease the risk that dominant stakeholders impose their views on other participants. This increases the likelihood that the resulting logic model more accurately represents the views of all involved stakeholders. Addressing this risk is particularly important when some stakeholders hold positions of professional power over other participants, and when the results of subsequent evaluations could impact operations of intervention implementation. Previously user-focused approaches to programme theory development, such as structured concept mapping (Yampolskaya et al., 2004) and the Nominal Group Technique (Julian, 1997) all included interaction between stakeholders, leaving it unclear how this may have influenced their results.

Limitations of this approach compared to other approaches

The decision to integrate a Delphi process into a larger, systematic user-focused approach to logic model development also carries some limitations. First, similar to previous user-focused efforts, our strong reliance on stakeholders' input and lack of deductive approaches based on formal social science theories resulted in a model that was heavily driven by non-evaluators. The limited experience of participants with evaluative science may have resulted in the inclusion of items that were not essential to actual programme success or in the exclusion of items that were. This creates a risk that evaluation resources could be allocated sub-optimally as a result of focusing on the wrong programme components. Second, although the absence of interaction between participants resulted in the previously discussed strengths, it also carries the limitation that it prevents the synergy between participants' accounts of a model that in-person interaction can offer as well as a good understanding of discrepancies in opinion and the reasons therefore. (Murphy et al., 1998). Furthermore, while face-to-face interaction was avoided due to aforementioned concerns, interaction between the researchers and participants could have actually increased participants' understanding of evaluative science. Finally, participants were unable to discuss discrepancies in their opinions related to specific items as well as the overall model directly with one another, there is a chance that optimal integration of their ideas was prevented by our approach.

Study limitations

Our study is also subject to some additional limitations. Primarily, only a select group of stakeholders (N=10) were invited to participate in the Delphi process. Participants were purposefully sampled based on nominations by the SST—the main promoter of the Studio Schools model, which indicated that no further individuals were eligible for nomination. Participants included stakeholders who were responsible for the original concept, expert staff of the SST, and a sample of school principals. It is unclear how a larger sample size would have affected the logic model and its components, and whether those nominated to participate are more likely to share views compared to other potentially eligible members of the stakeholder population. For example, it is unclear if the model we constructed would have looked differently had we included Studio School teachers and students in addition to those engaged in policy and school management. It is also worth noting that, in analysing Round Two results, the adapted A7S procedure of excluding extreme ratings resulted in a larger number of items classified as 'included' by the end of Delphi Round Two than would

have been the case without excluding extreme ratings. Thus, there are potentially more items in the logic model than would have been the case had we not discarded extreme ratings. Another potential limitation is posed by our decision not to extend the Delphi process to a third round. Instead of asking stakeholders to rate items that were classified as 'uncertain' for a second time and to have them link logic model components, a member of the research team was responsible for the decision to include or exclude uncertain items and then link items in the logic model. To compensate, participants were asked for their feedback on the integrated logic model that resulted from this decision-making process. The possibility exists that participants' 'passive' approval of the constructed model yielded a different set of included items than re-rating of items in a third round. Researchers or evaluators considering the use of the Delphi method to elucidate programme theory should carefully consider the time and resources needed to follow-up with participants for multiple questionnaire rounds.

Finally, we should stress that the approach described here aims to render consensual and explicit what might be contested and implicit about a programme theory of change. It does not aim to enhance this theory. It might be that in some cases, theories of change are logically deficient or are contradicted by existing empirical evidence. They also do not always draw on existing As realist evaluators have highlighted, programme theories may also inadequately theorise how intervention mechanisms interact with context to determine intervention outcomes (Bonell et. al, 2012). In such cases, the process that we have set out might need to be followed by further work with programme developers to enhance programme theory and modify the programme itself.

Implications for further research

This study poses some interesting possibilities and implications for the field of evaluation research. First, it confirms the feasibility of specifying and reaching agreement on a social programme's underlying theoretical assumptions after intervention design but prior to evaluation. Making such assumptions explicit is key to conducting useful evaluations—such as process evaluations within effectiveness trials—and this explication of programme theory should be done regardless of stage of implementation and evaluation. Second, our work demonstrates that a systematic, inclusive methodological process can be employed to elucidate implicit programme theories. We believe that this is likely to increase the credibility of results and buy-in from stakeholders involved in the process, as well as other key partners in subsequent programme evaluations.

Future studies undertaken to specify programme theories prior to evaluation are likely to benefit from use of methods like the Delphi process that promote systematic techniques and transparent reporting. These may include other explicit methods for building consensus (e.g., the nominal group technique, which includes more face to face interaction), or a series of well-reported interviews with developers that investigate how specific social and behavioural theories informed their programme. Studies comparing the results of different methods and approaches–for example a direct comparison of a systematic, user-focused approach with an inductive or deductive approach–can also be useful to advance methods for programme theory development.

Whichever methods future researchers or other evaluators use, this study demonstrates that non-academic developers of a complex intervention can arrive at consensus about programme theory after the programme has been developed and can

actively participate in the development of a logic model to be used as a framework for subsequent evaluation.