eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Critical Essay: What works for whom in which circumstances? On the need to move beyond the 'what works?' question in organizational intervention research

Karina Nielsen and Mariella Miraglia

In press Human Relations

Abstract

A debate has arisen out of the need to understand true intervention outcomes in the social sciences. Traditionally, the randomized, controlled trial (RCT) that answers the question of 'what works' has been considered the gold standard. Although RCTs have been favoured in organizational intervention research, there has been an increasing interest in understanding the influence of context and intervention processes on the outcomes of such interventions. In the present critical essay, we question the suitability of RCTs and meta-analyses to evaluate the effectiveness of organizational interventions and we suggest that realist evaluation that seeks to answer the questions of what works for whom in which circumstances may present a more suitable framework. We argue that examining the content and process mechanisms through which organizational interventions are effective, and the conditions under which these are triggered, will enable us to better understand how interventions achieve the desired outcomes of improved employee health and well-being. We suggest that organizational intervention content and process mechanisms may help bring about the desired outcomes of improved employee health and well-being and that contextual factors determine whether these mechanisms are triggered.

The randomized, controlled trial has been perceived as the gold standard for evaluating interventions (Guyatt et al., 1995). The RCT approach employs a successionist approach to causation, i.e. that randomization holds the context constant, there are no differences at baseline between the intervention and the control group, and outcomes can be inferred from comparing those exposed to the intervention to those not exposed to the intervention. When intended outcomes can be observed statistically above and beyond outcomes in the control group, it is assumed that these outcomes can be attributed to the intervention. Implicit in this line of thinking is that intervention outcomes can be aggregated in meta-analyses across diverse occupational and organizational settings and that intervention processes are uniform (Blamey and Mackenzie, 2007). Critics have argued that this line of enquiry is suboptimal in the social sciences (Pawson and Tilley, 1997) because it conceptualizes the context as confounding variables that need to be held constant and it views participants as passive recipients of interventions (Blamey and Mackenzie, 2007).

As a result, an opposing paradigm has gained impact in social science intervention evaluation. Contrary to the RCT successionist approach, realist evaluation employs a generative approach to causation (Blamey and Mackenzie, 2007). Realist evaluation seeks to answer the questions of what works for whom in which circumstances through studying what the Mechanisms of an intervention are (what makes an intervention work?) and the Context in which these may be triggered (what are the conditions in which an intervention is effective?) and bring about certain Outcomes (which improvements in working conditions and in employee health and well-being can be observed?), in what is also known as context-mechanism-outcome (CMO) configurations (Pawson and Tilley, 1997). Realist evaluators generate CMO-configurations that can be tested and, if necessary, revised. Realist evaluation has primarily been applied in healthcare (Marchal et al., 2012) and policy research (Pawson, 2013).

In the field of organizational interventions the RCT design has been met with criticism (Nielsen, 2013a). Organizational interventions can be defined as planned, behavioural, theory-based actions that aim to improve employee health and well-being through changing the way work is designed, organized and managed (e.g. Nielsen, 2013a; Richardson and Rothstein, 2008). In the present essay, we argue that realist evaluation (Pawson, 2013; Pawson and Tilley, 1997) may offer an opportunity to develop an integrated context, process and outcome evaluation framework that may advance our theoretical understanding of which elements of organizational interventions may be effective and in which conditions we can expect positive outcomes. Realist evaluation offers a way to conduct rigorous, theory-based analyses of what works for whom in which circumstances. Such analyses provide researchers and evaluators with insights into how to improve existing interventions and inform future interventions, while also ensuring internal and external validity.

In the following sections of the present essay, we first elaborate on the challenges of RCTs to discern the true effects of organizations interventions. Second, realist evaluation has been criticized for a lack of precision in the definitions of mechanisms and context (Marchal et al., 2012; Pawson and Manzano-Santaella, 2012) and we suggest operationalizations of both the mechanisms and context in organizational intervention research. Finally, we discuss the challenges of synthesizing organizational intervention research using meta-analysis and suggest realist synthesis as an alternative to meta-analysis.

Randomized, controlled trials – what are the challenges?

The central tenets of a study's quality concern its internal and external validity (generalizability). Internal validity is an indicator of the extent to which a cause-and-effect relationship between working conditions and its outcome is well-founded (Hoyle et al.,

2001), and it is crucial to evaluate whether an intervention produced any observed changes and to estimate the magnitude of these changes (Newcomer et al., 2010). To determine which designs are the most robust in terms of internal validity, the so-called 'levels of evidence' are used (Guyatt et al., 1995). These levels describe the hierarchical ordering of research design quality. A study has high internal validity when it fulfils the three conditions required for causal inference: covariation (that one thing is correlated with the other), time–order relationship (that one thing precedes the other), and elimination of plausible alternative causes (Shaugnessy et al., 2006). Drawing on the medical field, RCTs randomly allocate participants into either an intervention or a control group. The basic assumption is that any differences in effects in the two groups are attributable to the intervention rather than to any individual differences at baseline, e.g. gender, age or existing levels of resources. The goal of the RCT design is to assess whether the intervention had the desired effect or not, to answer the question of "what works?". RCTs are supposed to deal with the threats to internal validity by addressing the three aforementioned conditions for causal inference (Shaugnessy et al., 2006).

External validity refers to the extent to which an intervention will have the same effects in other settings or with other groups (Newcomer et al., 2010). In RCTs, one way of testing generalizability is to replicate the intervention across multiple sites in order to control for potential confounders such as organizational culture and occupational setting. Another way is to specify the settings in which an interventions is expected to work, for example, healthcare.

A study by Albertsen et al. (2014) exemplifies the challenges of employing a RCT design in organizational intervention research. In a study on the implementation of an IT rota system in eldercare, Albertsen et al. (2014) found very different results across three organizational settings. In organization A, no improvements in working conditions and well-

being were detected and interviews with employees and managers revealed that the organizational context had been problematic: downsizing had taken place at the same time and this had resulted in a temporary cancellation of the IT system and employees would be called in to work at short notice. Employees also found the system difficult to use. In organization B, improvements in work-life balance were observed and employees reported that the IT system supported existing roster planning procedures, the IT system had made the process fairer, and the system considered individual preferences. In the third organization, a deterioration in work-life balance was observed. Interviews revealed that although the IT system had been implemented, management had introduced a "buffer zone" that meant that they could delay or postpone working hours. This zone resulted in more evening work, variable working hours and unpredictability in when to start work. In other words, the same intervention had completely different outcomes based on the context and the implementation of the intervention.

Challenges to internal validity

Organizational interventions most often operate through a participatory, dynamic approach, where employees and managers in collaboration decide on the process and the content of intervention activities (Nielsen et al., 2010). Organizational interventions may best be classified as complex interventions because they work through an emergent and recursive causality (Rogers, 2008). This level of complexity creates challenges for internal validity in RCTs.

The RCT design only allows researchers to determine whether there was a change in the intended outcomes or not; it does not demonstrate whether this effect can be ascribed to the intervention itself or to other factors (Nielsen, 2013a). It has been argued that the intervention process and the behaviours of key players, such as line managers, have a

prominent role in supporting the intervention and shaping its outcomes (Nielsen and Abildgaard, 2013; Nielsen and Randall, 2009). Also the participatory process, i.e. the extent to which employees are involved in determining the intervention's processes and content, is essential in ensuring a successful intervention outcome (Nielsen and Randall, 2012). These outcomes are also strongly influenced by the multiple perceptions and interpretations of the intervention by the participants, as well as by their behaviours. According to realist evaluation, interventions work by giving participants the opportunity to make different choices about their agency (Greenhalgh et al., 2015). It is not the intervention itself that has an effect, but the choices made by participants on whether and how to change their behaviours (Pawson and Tilley, 1997). Translating this to the context of organizational interventions, an organizational intervention component, e.g. action planning, triggers certain behaviours of participants, namely their engagement in the participatory process of developing and implementing these action plans, and it is these behaviours that produce outcomes rather than the intervention itself (Nielsen and Abildgaard, 2013).

Furthermore, because of risks of contamination, i.e. the introduction of changes to procedure are likely to influence all employees in the department, organizational interventions employ cluster randomization. Therefore organizational interventions most often target an entire organizational unit; however, not all of participants in an organizational unit, e.g. a department, have a shared understanding of the changes that need to be made. For example, line managers, senior managers and Human Resource managers may not share their understanding of the process (Hasson et al., 2014) and these diverging interpretations impact on the outcomes of the intervention (Hasson et al., 2016). Realist evaluators argue that the impact of interventions cannot be determined without understanding the impact of individual perceptions and behaviours (Blamey and Mackenzie, 2007).

Finally, it could be argued that the RCT design fails to test "what works" because it does not identify which elements of complex interventions worked, e.g. was it the development and implementation of action plans or was it the fact that participants engaged in joint decision making? As such RCTs only allows researchers to draw conclusions about "whether the intervention worked or not".

Challenges to external validity

A further challenge of employing RCTs in organizational interventions relates to external validity or the generalizability of findings. RCTs are generally low on external validity as they rely on a highly controlled environment where the context is held constant (Guyatt et al., 1995). In organizational interventions it is difficult to keep the context constant and the context may determine the extent to which the intervention succeeds or fails (Nielsen et al., 2010).

RCTs advocate the use of control groups to infer the outcomes of interventions, however, complex contextual factors impair this assumption. The basic tenet of the RCT design is that randomization controls for baseline levels and accounts for potential confounders. However, due to aforementioned risks of contamination, cluster randomization is most often employed and as it is not practicable to recruit hundreds of organizations, N is often too small to rule out confounders.

Also other challenges are associated with the use of control groups in organizational interventions. It is notoriously difficult to find equivalent control groups that enable reliable comparisons between the intervention and the control groups (Pawson, 2013). If an entire department is participating, there may not be another department in the organization that provides an adequate match. For example, in hospitals there is only one Accidents and Emergency department, and an Accidents and Emergency department at another hospital may

have different procedures and policies. Different production units in a manufacturing plant may differ in terms of size and production methods. Furthermore, organizational interventions may be a response to national legislation (e.g. in Denmark it is a legal requirement to assess the working environment at least every third year and take appropriate measures to address any issues identified, Nielsen, 2013b) and thus the intervention must be implemented in the entire organization. In addition, it may be considered unethical to withhold the intervention from a group that has been identified as "at risk" and a wait-list control design may result in aggravation of problems in the wait-list group (Nielsen, 2013a). Moreover, research has indicated that the sheer act of randomization increases readiness for change in the intervention group due to feelings of "having won the lottery" (Nabe-Nielsen et al., 2015). As a result, merely comparing control and intervention groups does not allow for understanding any changes in the intervention group above and beyond the control group as we cannot draw conclusions as to why any such difference has occurred.

Furthermore, the existing conditions of the participants may influence the intervention's success. For example, is an intervention that relies on high levels of technology expected to be generalizable beyond a group of "tech savvy" engineers? RCTs are implemented in a specific context and discrete period of time, making the generalization of the findings questionable across contexts (Karlan et al., 2009) and offering limited information on the transferability of complex interventions. Subsequently, it is not possible to determine whether any changes in working conditions and employee health and well-being have been caused by the intervention or other factors such as concurrent events, e.g. downsizing or additional resource allocation.

In summary, RCTs assume that the intervention is the causal effect, but exclude the many process components that may explain an intervention's outcomes, potentially disregarding the real causes of the intervention's success or failure (Pawson, 2013). It is

essential to recognize that workplace interventions are "active" rather than passive programmes (Pawson, 2006) that continuously respond to contextual factors and emerging processes. Addressing the challenges related to internal and external validity, Newcomer et al. (2010) recommended that evaluators should assess the extent to which an intervention was implemented (a manipulation check), to examine whether any other events may account for the observed changes, consider whether sufficient time has lapsed for any changes to be attributed to the intervention (e.g. an intervention aimed at improving trust between management and employees may take some time to take effect), and evaluate whether similar interventions have produced similar results. These recommendations are in line with realist evaluation (Pawson, 2013). We therefore propose that realist evaluation may offer a way forward for evaluating organizational interventions.

The basic principles of realist evaluation

According to realist evaluation, we can build and test models that explain why an intervention succeeds in or fails to bring about intended outcomes (Pawson, 2013). The focus of realist evaluation is to answer the questions of "what works for whom in which circumstances?" in an attempt to open the black box of how and why interventions may or may not work. The unit of analysis in realist evaluation is an intervention's programme theory, i.e. the theory of what causes change, not whether an intervention worked or not (Greenhalgh et al., 2015). A programme theory can be seen as an implicit set of assumptions that steers the choice and design of an intervention (Pawson, 2013) and spells out the coordinated sequence of activities (e.g., preparation, screening, action planning, implementation and evaluation in organizational intervention, Nielsen et al., 2010) that are thought to be necessary to achieve the intervention's intended outcomes (Manzano and Pawson, 2014).

A central tenet of realist evaluation is to answer these questions through theoretically developing and testing Context + Mechanism = Outcome (CMO)-configurations (Pawson and Tilley, 1997; Pawson, 2013). A CMO-configuration can pertain to the entire, or parts of, an intervention and one CMO-configuration can be embedded in another (Pawson and Tilley, 1997). The realist evaluation strategy focuses on three themes: Understanding the Mechanisms through which an intervention achieves its Outcomes, understanding the Contextual conditions necessary for triggering these Mechanisms, and understanding Outcome patterns (Pawson and Tilley, 1997). Some factors in the context may enable certain mechanisms to trigger intended outcomes (Greenhalgh et al., 2015) and therefore interventions cannot simply be transferred from one context to the other; there is always an interaction between context and mechanisms and it is this interaction that creates the intervention's outcomes (Greenhalgh et al., 2015). The interplay between participants in the intervention and the structures in which the intervention is embedded determines the outcomes of the intervention and research should thus focus on how these agent-structure interactions produce outcomes (Greenhalgh et al., 2015).

Realist evaluators argue that it is crucial to understand that, because interventions work differently in different contexts and through different mechanisms, they cannot easily be transferred from one setting to another; however, the in-depth understanding of what works for whom in which circumstances is transferable (Goodridge et al., 2015).

Realist evaluation, however, is not without its challenges. Challenges have been identified with regards to the definition and operationalization of mechanisms and context and the empirical testing of CMO-configurations, together with the confusion that some view specific elements as mechanisms while others define them as context and vice versa (Lacouture et al., 2012; Marchal et al., 2012; Pawson and Manzano-Santaello, 2012). In the

following we discuss how mechanisms, context and CMO-configurations may be operationalized in organizational interventions.

Mechanisms: What makes organizational interventions work?

Interventions work by giving participants the opportunity to make different choices about their agency (Greenhalgh et al., 2015). Making and sustaining such changes in behaviour requires a change in people's reasoning and/or the resources they have available to them (Greenhalgh et al, 2015). Mechanisms can be expressed through interpretations, considerations, decisions and behaviours of participants, and outcomes are the result of their actions and interpretations of the intervention (Pawson, 2013). Despite the definition of mechanisms put forward by Pawson (2013), there is confusion as to what the term includes. In a review of the mechanism concept in realist evaluation, Lacouture et al. (2015) identified 49 public health intervention studies based on realist evaluation. They concluded in their analyses of the characteristics of mechanisms that 1) mechanisms can be defined as the reasoning and reactions of participants, 2) mechanisms may change over the duration of the intervention and that they are latent and only reveal themselves during the implementation of the intervention, 3) although mechanisms cannot be observed, they are real and exist regardless of whether they are activated or not, and 4) mechanisms interact with each other and may be linked in negative or positive feedback loops, i.e. one mechanism informs the other. In a review, Marchal et al. (2012) identified mechanisms at the individual (e.g. readiness for change), group (e.g. social capital building) and organizational (e.g. management behaviours) level.

Although there has been little explicit use of realist evaluation in the organizational intervention domain, a few studies on organizational practices and policies have employed a realist evaluation approach to identify mechanisms that may also be applicable to

organizational interventions. In an intervention aimed at facilitating return to work among employees on long-term sick leave, Higgins et al. (2015) suggested three mechanisms at the managerial level: line managers' communication with employees on sick leave, line managers' regular contact to employees at an early stage, and senior managers providing clear guidance on how to manage employees on sick leave.

Furthermore, mechanisms of Lean in healthcare have been suggested to be: 1) authorization of resources allocated to Lean activities resulting in employee participation, 2) a common set of Lean tools will ensure consistent improvement process and practices thus generating improved decision making which will lead to better quality of care, 3) leaders at the local level promoting the use of Lean tools will increase participation in quality improvement processes, and 4) increased levels of staff and patient participation in quality improvement processes and visibility of leaders will contribute to increased transparency which will in turn ensure leader accountability for implementing Lean (Goodridge et al., 2015).

Although rarely explicitly examined in organizational intervention research, a number of studies have implicitly suggested mechanisms. We argue that two main categories of mechanisms may be at play in organizational interventions. First, the content of the intervention, i.e. the content of action plans, may realize the intended outcomes. Second, the process of the intervention may also be an important mechanism that brings about improved employee health and well-being.

The content of contextualized action plans has been found to improve job control and employee well-being (Holman and Axtell, 2016). In a call centre study, an action plan specifying that employees rather than line managers would complete administrative tasks and deal with minor customer queries and complains was assumed to result in employees experiencing higher levels of job control and in turn well-being. Second, an action plan aimed

at clarifying the performance criteria and holding regular team briefings was assumed to lead to improved feedback and well-being among staff. The intervention outcomes were found to be improved job control, feedback and employee well-being (Holman and Axtell, 2016).

The intervention participatory process may be another mechanism (Bond et al., 2008). Here, it was assumed that enabling employees jointly decide on how to make changes to the way work is organized, designed and managed would to lead to improvements in job control and employee well-being. Bond et al. (2008) did find support that the intervention improved job control and employee well-being.

A limitation of both these studies is that, although they articulated potential mechanisms, these were not tested empirically. For example, in Holman and Axtell's study (2016), the extent to which employees took over administrative roles (the mechanism) was not measured and not linked to outcomes. Interestingly, the two studies both included job control as an outcome and argued that increased job control was either the result of the content or process of the intervention. This lack of clarity calls for studies that empirically examine and measure both types of mechanism to determine whether job control is triggered by the intervention's content or process.

In a qualitative study among Danish postal service workers, Nielsen et al. (2014) explored the use of a tailored questionnaire measuring these workers' local work environment as an intervention tool. The mechanisms underlying the tailored questionnaire were proposed to be 1) collective sensemaking of the results of the tailored questionnaire and the problems experienced by the group, 2) facilitation of prioritization because feedback on the results allowed participants to see how many employees reported an aspect of the work environment to be a problem and how strongly this aspect was linked to outcomes, 3) ownership among participants because they felt management were interested in their problems and their perspectives, and 4) facilitation of action planning because the questionnaire pointed to issues

specific to postal service workers, e.g. employees' opportunities to be heard in connection with changes in the postal routes.

Although it has been argued that mechanisms cannot be observed (Lacouture et al., 2015), the inability of the Bond et al. (2008) and the Holman and Axtell (2016) studies to determine whether the mechanism that triggered job control and well-being was the participatory process or the content of action plans calls for the operationalization of mechanisms. Current research on realist evaluation has been criticized for not linking mechanisms to outcomes (Pawson and Manzano-Santaello, 2012). Pawson and Manzano-Santaella (2012) argued that identifying mechanisms requires qualitative evidence but observing how mechanisms are linked to outcomes require quantitative analyses. Although qualitative methods can be used to examine who did what and what actually happened, qualitative information must be quantified (Picciotto, 2014).

Possible operationalizations of the participatory process could include, among others, measures of joint decision making, felt ownership over the process and perceived use of expertise. Operationalizations of the content could include manipulation checks of whether changes in administrative procedures had in fact been implemented. In Nielsen et al. (2014) mechanisms could have been operationalized as ownership, sensemaking and organizational learning. Testing these mechanisms and finding support for them could provide valuable information on how to design and implement future interventions: Organizational interventions should use tools that create a sense of ownership and joint decision making (Bond et al., 2008, Nielsen et al., 2014). Tools should also allow for the development of context-specific action plans that address the issues found to be important by participants (Holman and Axtell, 2016, Nielsen et al., 2014).

Understanding context in organizational interventions

Realist evaluation suggests that interventions work in different ways for different people depending on the context within which they find themselves, and thus whether and how mechanisms trigger outcomes depend on the context (Greenhalgh et al., 2015).

Pawson and Tilley (1997) defined context as the conditions that interventions find themselves in. Macfarlane et al. (2011) analysed the context at four levels: The individual (e.g. values, roles, and knowledge), the interpersonal (e.g. communication, collaboration and networks), the institutional (e.g. informal rules, organizational culture, leadership and regulations), and the infrastructural (e.g. political support).

The two aforementioned studies by Higgins et al. (2015) and Goodridge et al. (2015) offer some useful examples on context that may also apply to organizational interventions. The contextual factors identified by Higgins et al. (2015) included an organizational climate where senior management had good relations with line managers and the trade unions, communication between line managers and employees was respectful, line managers had prior adequate training to deal with employees on sick leave and where financial pressures motivated line managers to develop flexible return to work practices, such a flexitime working.

In the Goodridge et al. (2015) study, a number of contextual factors may have hindered mechanisms being triggered. Goodridge et al. (2015) suggested that the mechanisms of the intervention would not be triggered and Lean would not be implemented in regions: 1) where Lean was poorly integrated and aligned with other initiatives, 2) where leaders who were responsible for Lean implementation perceived a lack of centralized support to implement Lean, 3) where central authorities did not support adjustment to local context, 4) where measurement data on which to base quality improvement processes were not readily

available or of poor quality, and 5) where leaders feared they could not make mistakes

without repercussions.

Exploring context in organizational interventions

In recent models of organizational intervention evaluation, a distinction between

omnibus and discrete context has been suggested (Nielsen and Abildgaard, 2013; Nielsen and

Randall, 2013). Essentially, the omnibus context concerns the stable characteristics of the

organization, these include the "maturity" of the organization in terms of the organizational

culture, readiness for change, existing working conditions, and health and well-being of

employees. These aspects were embedded in the organization prior to the intervention and are

not manipulated or targeted as part of the intervention. The discrete context concerns the

concurrent changes taking place during the intervention period. This distinction is useful in

realist evaluation. Traditionally, realist evaluation has understood the context as the setting in

which the intervention takes place, e.g. a postal service where existing levels of employee

health and well-being point to the need for intervention. This approach has been criticized for

being too limited in that contexts are dynamic and thus the focus on a pre-existing and stable

context does not capture its complexity (Dahler-Larsen, 2001). It could be argued that the

omnibus context captures the stable, cultural aspects whereas the discrete context captures the

dynamic elements. These dynamic elements interact with the mechanisms as and when they

occur, i.e. downsizing during the action planning phase may divert attention from developing

viable action plans.

Omnibus context can be examined at the individual, group and institutional level.

Examples of omnibus context, i.e. the contextual factors that were present prior to the

intervention, at the individual level have been reported in a number of organizational

interventions. Bond et al. (2008) found that an organizational intervention had positive effects

on mental health and absence rates for employees high in psychological flexibility. Readiness for change or openness to the intervention has also been found to be important: Participants need to acknowledge problems in the current situation, see the need for change, believe the intervention will have the desired effect and be motivated to follow the requirements for behavioural change made by the intervention (Nytrø et al., 2000). The degree to which employees welcome change, and feel they can effectively handle the change, predict the degree to which they participate in change activities and feel that they themselves have made a significant contribution to the change (Cunningham et al., 2002). Employees' resources are also an important contextual factor; Nielsen et al. (2006) showed that employees with little formal education found it challenging to engage in the participatory process. In their review, Lyubomirsky et al. (2005) concluded that individuals high in affective well-being are more proactive and creative, and engage in more activities and problem solving and thus high well-being levels may facilitate the participatory process.

At the group level, in groups where employees are not used to communicate or solve problems at work, participation in the intervention process is reported as difficult and insights about problems do not lead to willingness or ability to make things happen (Mikkelsen et al., 2000).

At the institutional level, poor pre-intervention working conditions and well-being have been shown to limit implementation processes (Taris et al., 2003), however, another body of research suggests a ceiling effect. Nielsen et al. (2006) found that an intervention group that had, prior to the intervention, worked with organizational development and thus reported having good working conditions prior to the intervention, reported few improvements. Nielsen and Randall (2012) found that high levels of pre-intervention social support were related to few changes in existing working procedures in an intervention aiming to improve social support. Together, these studies suggest the importance of intervention fit

(Nielsen and Randall, 2015), i.e. that the intervention activities need to be fitted to the organizational context. If the organization is already in good shape, the intervention may not be needed but on the other hand, a support structure may need to be in place for the intervention to work. If leaders and employees have no shared understanding and limited experience in problem solving, engaging in the participatory process may be an unsurmountable challenge.

In organizational intervention research, concurrent changes, part of the discrete context, have often been described as "disturbing" the intervention and as "unnecessary noise" (Mikkelsen and Saksvik, 1999; Nielsen et al., 2006). Discrete events both at the intra-organizational level, such as the introduction of conflicting initiatives (Nielsen et al., 2006) or of new technology (Saksvik, et al., 2002), and at the extra-organizational level, for example economic recession leading to layoffs within the organization (Landsbergis and Vivona-Vaughan, 1995; Mikkelsen and Saksvik, 1999), have been found to impact intervention outcomes negatively. In today's globalized economy, organizations constantly reorganize and restructure in order to adapt to the demands of their environment (Grant, 2007) and realistically evaluating the discrete context would enable the analysis of the impact of changing organizations on the outcomes of organizational interventions. We need to understand whether, and if so how, the dynamic context acts as a barrier or a facilitator for certain mechanisms (Noblet and LaMontagne, 2009). For example, if a new line manager is appointed during the action planning phase, employees may not be familiar with him or her and therefore may not feel confident sharing their views and ideas openly during the participatory process; hence, the mechanisms of the participatory process may not be triggered. Knowledge about this threat may lead to the development of supportive activities, i.e. team building activities to build trust before progressing to develop action plans.

In summary, we argue that in the context of organizational interventions, we need to understand, measure and build upon knowledge about the omnibus and the discrete context when we plan future interventions.

Analyzing organizational interventions using CMO-configurations

As mentioned, realist evaluation has rarely been used in organizational intervention research and many studies of organizational interventions examine either only the context or the mechanisms without explicitly using realist evaluation as their framework. A limitation of current realist evaluations is the failure to test CMO-configurations (Pawson and Manzano-Santaella, 2012). In this section, we present a few cases of existing research where CMO-configurations can be applied to understand the intervention's outcomes and discuss their implications. To illustrate the interplay between context and mechanisms, we explore contextual factors as both moderators (third factors that enable or condition the relationship between intervention components) and antecedents of mechanisms.

Based on the suggestion of Higgins et al. (2015), one mechanism in organizational interventions may be training of leaders to equip them with the skills to implement changes. In a teamwork intervention, Nielsen and Daniels (2012) examined the effects of training newly appointed team leaders in how to implement and manage teams during a time of team implementation. The rationale for team leader training was that, through training, leaders would come to see team implementation as a positive challenge to develop in their jobs rather than a threat to their status (Lazarus and Folkman, 1984). Nielsen and Daniels (2012) found that training leaders (mechanism) leads to improved leader well-being (outcome), but only in circumstances where employees reported being ready to work in teams (omnibus context). Where employees did not report being ready to work in teams, trained leaders experienced poorer well-being compared to leaders who had not received training. This multi-level

intervention (team implementation and leader training) provides important insights in what works for whom in which circumstances. Nielsen and Daniels (2012) thus operationalized mechanisms as mediators (a component of the programme) that intervene between baseline and follow-up to influence the outcome, and context as a moderator.

Two CMO-configurations were implicitly tested by Nielsen and Randall (2012). First, a CMO-configuration in which the participatory process was supposed to be the mechanism assumed to trigger job control (as in the Bond et al., 2008 study) and social support and subsequently well-being. It was proposed that this mechanism would only be triggered among employees who had a certain level of resources, e.g. job control, social support, job satisfaction and affective well-being. Second, the mechanisms of changes in work procedures to reflect teamwork would increase employees' opportunities to make independent decisions on how to do their job in a collaborative fashion and thus trigger the outcomes of autonomy and social support and therefore also job satisfaction and well-being. As with the participation mechanism, it was assumed that only among individuals with good existing resources would the mechanism be realized. In a structural equation model, Nielsen and Randall (2012) found that pre-existing levels of autonomy and job satisfaction (omnibus context) enabled the participatory process (mechanism), which led to improvements in social support and job satisfaction. They also found that pre-existing levels of affective well-being and social support (omnibus context) triggered the changes in procedures (mechanism) which led to improvements in autonomy, affective well-being and job satisfaction.

Although not discussed by Nielsen and Randall (2012), we can identify two separate CMO-configurations: First, C2 (pre-existing levels of autonomy and job satisfaction) triggers M2 (participatory intervention processes), which leads to the Outcome of job satisfaction O2. Second, C1 (among employees who have high levels of affective well-being and social support) triggers M1 (changes in procedures to implement teamwork are introduced) that

leads to O1 (leading to increased autonomy and in turn affective well-being and job satisfaction). Transferring the understanding obtained through these two CMO-configurations to the design of future organizational interventions would suggest that the mechanism of a participatory process is triggered when employees experience high levels of autonomy, i.e. have experience engaging in independent decision making, and are satisfied with their jobs prior to the intervention. The participatory process cannot be expected to bring about increased levels of social support and well-being, but autonomy and job satisfaction are likely outcomes of the participatory process. Other activities that target social support such as the establishment of mentors or coaches may help increase employee wellbeing. The mechanism of changes in work procedures (a content mechanism) can only be triggered if employees have high levels of affective well-being, i.e. the energy to change working procedures and where they feel supported by their colleagues prior to the intervention (omnibus context). In these circumstances the mechanism of changes in procedures may be triggered and can be expected to result in both affective well-being and job satisfaction. Realist evaluation seeks to establish CMO-configurations that can be tested across a range of interventions and organizational settings and on the basis of their support or lack of refined CMO-configurations can be identified (Pawson, 2013). The Nielsen and Randall (2012) study provides a good example of how CMO-configurations can be tested and subsequently revised.

Synthesizing the effects of organizational interventions

In organizational intervention evaluation, meta-analyses are considered the standard when aiming to estimate the overall effectiveness of interventions and to draw evidence-based conclusions and recommendations (Rousseau et al., 2008). Meta-analyses are based on the hierarchy of evidence (Guyatt et al., 2001), suggesting that the RCT offers the "best

evidence" and so some meta-analyses only include RCTs while others also consider studies with non-randomization, or no control groups.

By quantitatively combining existing studies, meta-analyses statistically synthesize the results of several RCTs across contexts, to obtain a single weighted average measure of the effect of a certain intervention (i.e., its effect size). Meta-analyses have the merit to attain statistical precision, by weighting the estimated effect size on sample size and study artifacts (Schmidt and Hunter, 2015) and by assessing across-study variability around the effect size (Borenstein et al., 2009). Moreover, they strive to achieve methodological rigor and replicability, by transparently reporting the numerous procedural judgements and choices a meta-analyst faces. This strategy could give the illusion that meta-analyses may overcome the contextual specificity of RCTs.

In organizational intervention research, meta-analyses most often compare organizational interventions with individual-level interventions, i.e. interventions aimed at developing the knowledge, skills and abilities of employees and enable them to cope with the demands of the job (Richardson and Rothstein, 2008; Routsalainen et al., 2008; van der Klink et al., 2001). These reviews have demonstrated the effectiveness of individual-level interventions, especially cognitive-behavioural types, compared to organizational-level interventions in reducing work-related stress (Richardson and Rothstein, 2008; Routsalainen et al., 2008; van der Klink et al., 2001). Van der Klink and colleagues (2001) distinguished four categories of interventions (i.e., cognitive-behavioural, relaxation techniques, multimodal, organizational) and found no effect of organizational interventions targeting the way work was organized, designed and managed (K = 5; d = .08, 95% CI = -.03, .19), concluding that interventions that target individuals are most effective. The ineffectiveness of organizational interventions has been supported by a further meta-analysis (Richardson and Rothstein, 2008; K = 5; d = .14, 95% CI = -.12, .41), which extended the previous study

(van der Klink et al., 2001) by including the so-called "grey literature" (unpublished studies) and only methodologically strong designs (i.e. RCTs). Moreover, in a synthesis of systematic reviews and meta-analyses, Bhui et al. (2012) concluded that organizational interventions show limited and/or mixed evidence of benefits. Based on realist evaluation principles, we question whether such meta-analyses and reviews are appropriate when synthesizing organizational intervention studies, and we argue that they suffer from a number of pitfalls that may invalidate their findings and any related evidence-based recommendations.

Meta-analytic challenges to organizational intervention evaluation

As mentioned above, organizational interventions have been contrasted with individual-level interventions in order to assess their efficacy; however, we argue that these reviews compare apples and oranges. Individual-level interventions follow a pre-planned and pre-defined protocol and can therefore be defined as simple or complicated (Rogers, 2008), while, as described above, organizational interventions are complex and based on emerging processes and causation. The main challenge for meta-analyses of organizational intervention effectiveness is to disentangle this complexity and detect the real effects of the intervention components; both process and content.

Meta-analytic analysis adopts an over-simplistic approach, which fails to realistically review interventions in the social science domain (Pawson, 2006). In order to achieve rigour and net effect-size estimates, simplifications are used at multiple stages of the meta-analytic process, from formulating the review question to disseminating the findings (Pawson, 2006).

First, when first sifting the literature, clear criteria are set to decide which studies should be included or excluded from the analyses, and this may force to eliminate most of the contextual information and details of the interventions. Qualitative studies are excluded for obvious reasons, but they could be useful to understand the contextual factors and

mechanisms underlying the effectiveness of organizational interventions. Since one poorly conducted study could erroneously alter the meta-analytic effect size, only those studies that meet certain standards of methodological and statistical quality are included. This assures the accuracy of the estimation, but at the same time exclude a whole set of valuable information about processes and mechanisms. Indeed, poorly conducted or unsuccessful studies can reveal a lot about how and why organizational interventions worked or did not, offering useful insights for planning future interventions (e.g. Albertsen et al., 2014; Nielsen et al., 2006). An example of the selectivity of meta-analyses is offered by the work of Richardson and Rothstein (2008), who, as aforementioned, only included RCTs, resulting in total number of 5 meta-analytic samples. If, on the one hand, this reflects the sifting process that meta-analyses have to perform to achieve rigour and precision, on the other hand it speaks for the lack of evidence and paucity of studies on organizational interventions, due to the difficulties of conducting complex interventions (Bhui et al., 2012; Richardson and Rothstein, 2008). The limited number of organizational intervention studies calls for further research on the topic, but also for cautious interpretation of the meta-analytic results as they are inevitably based on limited information (Richardson and Rothstein, 2008; van der Klink et al., 2001).

Second, and directly linked to the issue of included/excluded research, meta-analyses are vulnerable to publication bias, which occurs when the retrieved studies are systematically unrepresentative of the study population (Rothstein et al., 2005). Studies may be excluded from the literature (i.e., not published) because they did not confirm the expected hypotheses or did not achieve statistically significant results. Although recently more advanced methodologies have been used to assess the effect of publication bias (e.g., Duval and Tweedie, 2000; Sterne and Egger, 2005; Sterne et al., 2005), the failure to include unpublished studies – that report unsuccessful results from an intervention point of view – not

only threatens the validity of the meta-analysis (Dickersin, 2005), but limits our understanding of the key factors underlying intervention effectiveness.

Third, these expected outcomes do not always provide useful insights into how the intervention operates, into the processes through which organizations may target resources or maximize the impact of their efforts to improve employee health and well-being. In other words, it is difficult to understand the mechanisms through which the intervention works, mainly because these mechanisms are rarely measured and therefore cannot be included in meta-analyses. For example, the van der Klink et al. (2001) meta-analysis found only one study that tested intermediate outcomes, representative of mechanisms, to illustrate how the organizational intervention exerted its influence on the expected individual-level outcomes. Specifically, the study reported that the intervention reduced individual stress (i.e., the individual-level expected outcome) via its impact on "psychological responses and resources", a measure of self-esteem, self-efficacy and coping strategies (i.e., the intermediate outcome).

Fourth, it can take time to see the effect of organizational interventions on individual employees (van der Klink et al., 2001). The variable "time" is never taken into account by meta-analyses of organizational interventions. Sensitivity analyses, which can differentiate among studies according to the time span occurring between the intervention and the measurement of its outcomes, could help to gauge the effect of time. Existent meta-analyses have not considered this possibility, probably also due to the limited number of included studies (K = 5, Richardson and Rothstein, 2008; van der Klink et al., 2001), which may have prevented such analyses.

Finally, a common criticism of meta-analysis is the superficial treatment of context. This is particularly critical for intervention studies, which are highly dependable on contextual factors. It could be argued that meta-analyses can perform moderator (and

sensitivity) analyses in order to investigate those differences among studies attributable to such contextual factors or to intervention types. Moderator analyses are considered one of the strengths of the meta-analytic approach because they recognize variation in the distribution of the estimated effect size and they aim to explain it (Borenstein et al., 2009). However, this variability may not easily be disentangled: its source may not be disclosed from the studies included in the meta-analysis as authors may have not measured the discrete contextual variables that account for differences in study outcomes, e.g. downsizing or the appointment of a new manager (or studies may not report net statistical indicators of these). Moreover, moderator analyses are often performed through meta-regressions, which require to select an even more strict number of studies (i.e., all studies that performed the same intervention in relation to the same outcomes, and reported useful statistics for the same moderators), further limiting the number of studies included in the subgroup analysis, and thus the power of the analyses (Borenstein et al., 2009).

An example of the limitation of meta-analyses is Nielsen et al. (2006), which compared two intervention groups and two non-randomized control groups. Improvements in working conditions and employee health and well-being were found in one intervention and one control group whereas deteriorations or no effects could be observed in the other intervention and control groups. A meta-analytic comparison of these groups would conclude that, overall, the intervention was ineffective. However, the study's process evaluation revealed that contextual differences and the fact that the "successful" control group implemented similar activities as the "successful" intervention group could explain the unexpected outcomes. The results indicate that the intervention had an effective programme theory (e.g. increasing employee health and well-being through activities to improve social relations) and that the mechanism of team building had a positive effect in the right circumstances (a context where baseline levels of social support were low). This example

illustrates the difficulty for meta-analyses of dealing with inconsistent findings, which can result in misevaluating intervention studies and in losing important information about their effectiveness.

In summary, meta-analyses fail to address the complexity of organizational interventions, thus potentially underestimating their effectiveness. Meta-analyses aim to quantitatively summarize the specific relationships between certain causes and a set of expected outcomes. In so doing, they fail to consider the complex patterns of contextual elements, emerging mechanisms, participative and recursive processes, participants' expectations, perceptions, interpretations and behaviours that ultimately determine the success (or failure) of organizational interventions. In other words, they may fail to identify the true relationships of interest and they may not (meta-)analyze those elements that account for the intervention's outcomes, missing complex and precious information.

Synthesizing the CMOs of organizational interventions: Realist synthesis

The limitations of RCTs and meta-analyses, along with our advocation for realist evaluation, have implications for the synthesis of organizational intervention effectiveness. Extending realist evaluation to the level of synthesis, we argue that realist synthesis may be a suitable avenue for understanding what works for whom in which circumstances. We argue that meta-analyses and RCTs do not meet good standards for internal and external validity in complex interventions such as organizational interventions. Internal and external validity may be better captured by studying which specific elements of the intervention work, how and why, and in which circumstances. Synthesizing the literature reviews that explore what works for whom in which circumstances in healthcare, Straus et al. (2016) found that alternative review methods have gained popularity. In 2000, ten alternative reviews had been published, but in 2013 the number has risen to 300 alternative reviews. In organizational intervention

research, however, the systematic reviews and meta-analyses still dominate (Bhui et al., 2012).

The question remains: How can we synthesize learning and offer recommendations on the design and implementation of future organizational interventions? We argue that realist synthesis, which builds on the principles of realist evaluation, can successfully be transferred to organizational intervention research by focusing on synthesizing CMO-configurations.

Realist syntheses in organizational intervention research would synthesize the mechanisms of the content and the process of the intervention and identify how participants' reactions (both cognitive and behavioural) shape the interventions' outcomes and in which contexts these mechanisms may be triggered. For example, is the intervention only likely to be successful in large organizations where existing health and safety management structures are well established or can elements of the intervention be transferred to interventions in Small and Medium-sized Enterprises? Rather than focusing on the outcomes or effectiveness of organizational interventions, prior to their analyses researchers conducting a realist synthesis would develop a series of CMO-configurations that could be tested. In Table 1, we outline the potential steps of a realist synthesis on organizational interventions.

Insert Table 1 about here

For example, a CMO-configuration or programme theory could be: If the aim is to improve job satisfaction among a group of workers (outcome), one way to achieve this outcome is to engage employees in a participatory process where managers and employees jointly identify and prioritize which issues they wish to focus on and jointly develop activities that address

these issues (mechanism). An important pre-condition for the participatory process is that managers and employees have a shared understanding of the issues at hand, they trust and respect each other's opinions and they have expertise in problem solving (context). Realist evaluators would thus search the literature for studies examining the participatory processes. They would then examine whether these studies lead to increasing job satisfaction and whether there were any descriptions of the context that triggered the mechanism.

The scarcity of published evaluations of organizational interventions is clear from the meta-analyses (Richardson and Rothstein, 2008; van der Klink et al., 2001. However, there is also a scarcity of studies documenting the outcomes of process and context (Nielsen et al. 2010, Nielsen and Abildgaard, 2013). The studies by Higgins et al. (2015) and Goodridge et al. (2015) provide good suggestions of the mechanisms that may also be important in organizational interventions, although neither study tested these mechanisms empirically. Through adopting a CMO approach, findings relating to the mechanisms can be incorporated thereby increasing the knowledge base concerning what works for whom under which circumstance. Furthermore, as realist syntheses focus on the mechanisms of interventions not interventions themselves, this enables the inclusion of studies from other areas of research and qualitative studies that may have studied a mechanism (Blamey and Mackenzie, 2007). For example, rather than limiting the selection of studies to organizational interventions we might turn to studies in healthcare to learn about mechanisms about the participatory process. Broadening our inclusion criteria in this way can help to synthesize learning across disciplines.

The potential research questions in a realist synthesis of organizational interventions are diverse and have implications for advancing our understanding of how and why organizational interventions. They could include: What are the programme theories for making changes to the way work is organized, designed and managed and the processes by

which change is brought about? More specific research questions could include: Which theories can be used to explain process mechanisms? Which theories can be used to explain content mechanisms? Which discrete context factors trigger a) process and b) content mechanisms? Which omnibus context factors trigger a) process and b) content mechanisms? What are the most important Outcomes brought about by these CM configurations? What areas for further research can be identified to inform and develop sustainable organizational interventions? These are important questions that cannot be answered by employing a meta-analytic approach.

Conclusion

In the present critical essay, we have argued that realist evaluation offers a way to open the black box of organizational interventions. After illustrating the limitations of RCTs and meta-analyses when evaluating the effectiveness of organizational interventions, we have put forward realist evaluation and realist synthesis as more suitable approaches to understanding the outcomes of an intervention and the explanations (i.e., context and mechanisms) for such outcomes. Acknowledging the challenges of current realist evaluation in determining which factors can be defined as context and mechanisms and testing complete CMO-configurations, we propose an operationalization of context into omnibus and discrete context and mechanisms concerning the content and the process of interventions.

Realist evaluations can elicit knowledge and learning of interest to a wide range of stakeholders. First, this approach can be used by researchers to develop and test theories of how and why organizational interventions work, for whom and in which circumstances: Developing and testing theories that focus on the mechanisms that relate to the content or the process of the intervention. Second, realist evaluation can be used by organizations, HR and occupational health professionals who wish to improve employee health and well-being

through changing the way work is organized, designed and managed: Gaining an understanding of which specific components of an interventions' content and process helps organizational members better plan interventions and set realistic expectations of what can be achieved.

Finally, through developing an understanding of the discrete and complex content and process mechanisms that underpin effective interventions, realist evaluation can help inform policy at the national and intervention levels and can lead to the development of methods, tools and guidelines that policy bodies can recommend organizations use in their efforts to improve employee health and well-being.

References

Albertsen K, Garde AH, Nabe-Nielsen K, Hansen ÅM, Lund H and Hvid H (2014) Work-life balance among shift workers: results from an intervention study about self-rostering. International Archives of Occupational and Environmental Health 87: 265-274.

Bhui K, Dinos S, Stansfeld S and White P. (2012) A synthesis of the evidence for managing stress at work: a review of the reviews reporting on anxiety, depression, and absenteeism. Journal of Environmental and Public Health 2012: 515-874.

Blamey A and Mackenzie M (2007) Theories of change and realistic evaluation: peas in a pod or apples and oranges? Evaluation 13(4): 439–55.

Bond FW, Flaxman PE and Bunce D (2008) The influence of psychological flexibility on work redesign: Mediated moderation of a work reorganization intervention. Journal of Applied Psychology 93(3): 645-54.

Borenstein M, Hedges LV, Higgins JPT and Rothstein HR (2009) Introduction to meta-analysis. Chichester: Wiley.

Dahler-Larsen P (2001) From Programme Theory to Constructivism On Tragic, Magic and Competing Programmes. Evaluation 7(3): 331-49.

Cunningham CE, Woodward CA, Shannon HS, MacIntosh J, Lendrum B, Rosenbloom D et al. (2002) Readiness for organizational change: A longitudinal study of workplace, psychological and behavioural correlates. Journal of Occupational and Organizational Psychology 75, 377-392.

Dickersin K. (2005) Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein HR, Sutton AJ, Borenstein M (eds), Publication bias in meta-analysis: Prevention, assessment, and adjustments. West Sussex, England: John Wiley, 11-33.

Duval SJ and Tweedie RL (2000) Trim and fill: A simple funnel plot–based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56: 455-463.

Goodridge D, Westhorp G, Rotter T, Dobson R and Bath B (2015) Lean and leadership practices: development of an initial realist program theory. BMC Health Services Research 15(362): 1-15.

Grant AM (2007) Relational job design and the motivation to make a prosocial difference. Academy of Management Review 32(2): 393-417.

Greenhalgh T, Wong G, Jagosh J, Greenhalgh J, Manzano A, Westhorp G and Pawson R (2015) Protocol-the RAMESES II study: Developing guidance and reporting standards for realist evaluation, BMJ Open, 5.8

Guyatt G, Sackett D, Sinclair J, Hayward R, Cook D and Cook R (1995) User's guides to the medical literature. IX. A method for grading health care recommendations. Journal of the American Medical Association 274: 1800–1804.

Hasson H Villaume K, von Thiele Schwarz U and Palm K (2014) Managing implementation: roles of line managers, senior managers, and human resource professionals in an occupational health intervention. Journal of Occupational and Environmental Medicine 56: 58–65.

Hasson H, von Thiele Schwarz U, Nielsen K and Tafvelin S (2016) Are we all in the same boat? The role of perceptual distance in organizational health interventions. Stress & Health. DOI: 10.1002/smi.2703.

Higgins A, O'Halloran P and Porter S (2012) Management of long term sickness absence: a systematic realist review. Journal of Occupational Rehabilitation 22(3):322–332.

Holman D and Axtell C (2016) Can job redesign interventions influence a broad range of employee outcomes by changing multiple job characteristics? A quasi-experimental study. Journal of Occupational Health Psychology 21: 284-295.

Hoyle RH, Harris MJ and Judd CM (2001) Research methods in social relations (7th ed.). Belmont, CA: Wadsworth.

Jones JW, Barge BN, Steffy BD, Fay LM, Kunz LK and Wuebker LJ (1988) Stress and medical malpractice: Organizational risk assessment and intervention. Journal of Applied Psychology 73: 727–735.

Karlan D, Goldberg N and Copestake J (2009) Randomized control trials are the best way to measure impact of microfinance programmes and improve microfinance product designs. Enterprise Development and Microfinance 20(3): 167-176.

Lacouture A, Breton E, Guichard A and Ridde V (2015) The concept of mechanism from a realist approach: a scoping review to facilitate its operationalization in public health program evaluation. Implementation Science 10: 153-163.

Landsbergis P and Vivona-Vaughan E (1995) Evaluation of an occupational stress intervention in a public agency. Journal of Organizational Behavior 16: 29-48.

Lazarus R and Folkman S (1992) Stress, Appraisal and Coping. New York: Springer Publications.

Lyubomirsky S, King L and Diener E (2005) The benefits of frequent positive affect: Does happiness lead to success? Psychological Bulletin 131(6): 803-855.

Macfarlane F, Greenhalgh T, Humphrey C, Hughes J, Butler C and Pawson R. (2011) A new workforce in the making? A case study of strategic human resource management in a whole-system change effort in healthcare. Journal of Health Organization Management 25(1):55–72.

Manzano A and Pawson R (2014) Evaluating consent to deceased organ donation: The need for programme theory. Journal of Healthcare Organization & Management 28(3): 366-385.

Marchal B, van Belle S, van Olmen J, Hoerée T and Kegels G (2012) Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. Evaluation 18(2):192–212.

Mikkelsen A, Saksvik PØ and Landsbergis P (2000) The impact of a participatory organizational intervention on job stress in community health care institutions. Work & Stress 14(2): 156-70.

Mikkelsen A and Saksvik PØ (1999) Impact of a participatory organizational intervention on job characteristics and job stress. International Journal of Health Services 29(4): 871-93.

Nabe-Nielsen K, Persson R, Nielsen K, Olsen O, Carneiro IG and Garde AH (2015) Perspective on randomization and readiness for change in a workplace intervention study. In Karanika-Murray, M., and Biron, C. Derailed organizational stress and well-being interventions: Confessions of failure and solutions for success (pp.201-208). Dordrecht: Springer.

Newcomer KE, Hatry HP and Wholey JS (2010) Planning and Designing Useful Evaluations. In JS Wholey, HP Hatry, and KE Newcomer. (3rd ed.), Handbook of practical program evaluation (pp. 5-29). San Francisco: Jossey-Bass.

Nielsen K (2013a) How can we make organizational interventions work? Employees and line managers as actively crafting interventions. Human Relations 66: 1029-1050.

Nielsen K (2013b) Managing psychosocial risks in Denmark: research and policy initiatives. In J.M. Peiro and C. Molina (eds). International yearbook of psychosocial risk prevention and quality of life at work. Evaluation development of psychosocial risks in Europe. The state of scientific research and institutional experiences. Spain, Jaen: Secretary of Labour Health and Environment UGT-CEC.

Nielsen K and Abildgaard JS (2013) Organizational interventions: A research-based framework for the evaluation of both process and effects. Work & Stress 27: 278-297.

Nielsen K, Abildgaard JS and Daniels K (2014) Putting context into organizational intervention design: Using tailored questionnaires to measure initiatives for worker well-being. Human Relations 67(2): 1537-1560.

Nielsen K, Fredslund H, Christensen KB and Albertsen K (2006) Success or failure? Interpreting and understanding the impact of interventions in four similar worksites. Work & Stress 20(3): 272-87.

Nielsen K and Randall R (2009) Managers' active support when implementing teams: The impact on employee well-being. Applied Psychology: Health and Well-being 1(3): 374-390.

Nielsen K and Randall R (2012) The importance of employee participation and perception of changes in procedures in a teamworking intervention. Work & Stress 29(1): 111.

Nielsen K and Randall R (2013) Opening the black box: A framework for evaluating organizational-level occupational health interventions. European Journal of Work and Organizational Psychology 22(5): 601-617.

Nielsen K and Randall R (2015) Addressing the fit of planned interventions to the organizational context. In: Karanika-Murray M, Biron C (eds.) Derailed organizational stress and well-being interventions: Confessions of failure and solutions for success (pp. 107-118). Dordrecht: Springer.

Nielsen K, Randall R, Holten AL and Rial González E (2010) Conducting Organizational-level Occupational Health Interventions: What Works? Work & Stress 24(3): 234-59.

Noblet A and LaMontagne A (2009) The challenges of developing, implementing, and evaluating interventions. In: Cartwright S, Cooper CL (eds) The Oxford Handbook of Organizational Well-being. Oxford: Oxford University Press, pp. 466-496.

Nytrø K, Saksvik PØ, Mikkelsen A, Bohle P and Quinlan M (2000) An appraisal of key factors in the implementation of occupational stress interventions. Work & Stress 14(3): 213-225.

Pawson R (2013) The science of evaluation: a realist manifesto. Los Angeles, CA and London: Sage.

Pawson R (2006) Evidence-based policy: A realistic perspective. Los Angeles, CA and London: Sage.

Pawson RD and Manzano-Santaella A (2012) A Realist Diagnostic Workshop. Evaluation: international journal of theory, research and practice 18(2): 176-191.

Pawson R and Tilley N (1997) Realistic evaluation. Los Angeles, CA and London: Sage.

Picciotto R (2014) Have Development Evaluators Been Fighting the Last War… And If So, What is to be Done? IDS Bulletin 45(6): 6–16.

Richardson KM and Rothstein HR (2008) Effects of occupational stress management intervention programs: A meta-analysis. Journal of Occupational Health Psychology 13: 69–93.

Rogers PJ (2008) Using programme theory to evaluate complicated and complex aspects of Interventions. Evaluations 14: 29-48.

Rothstein HR, Sutton AJ and Borenstein M (2005) Publication bias in meta-analysis. In: Rothstein HR, Sutton AJ, Borenstein M (eds) Publication bias in meta-analysis: Prevention, assessment, and adjustments. West Sussex, England: John Wiley, pp. 1-7.

Rousseau DM, Manning J and Denyer D (2008) Evidence in management and organizational science: Assembling the full weight of scientific knowledge through synthesis. The Academy of Management Annals 2: 475-515.

Routsalainen J, Serra C, Marine A and Verbeek J (2008) Systematic review of interventions for reducing occupational stress in health care workers. Scandinavian Journal of Work, Environmental and Health 34(3): 169–178.

Saksvik PØ, Nytrø K, Dahl-Jørgensen C and Mikkelsen A (2002) A process evaluation of individual and organizational occupational stress and health interventions. Work & Stress 16(1): 37-57.

Shaugnessy JJ, Zechmeister EB and Zechmeister JS (2006) Research methods in psychology (7th ed.) New York, NY: McGraw-Hill.

Schmidt FL and Hunter JE (2015) Methods of meta-analysis: Correcting error and bias in research synthesis (3rd Ed.). Thousand Oaks, CA: Sage.

Sterne JA and Egger M (2005) Regression methods to detect publication bias and other bias in meta-analysis. In: Rothstein HR, Sutton AJ, Borenstein M (eds), Publication bias in meta-analysis: Prevention, assessment, and adjustments. West Sussex, England: John Wiley, pp. 99-110.

Sterne JA, Gavaghan D and Egger M (2005) The funnel plot. In: Rothstein HR, Sutton AJ, Borenstein M (eds) Publication bias in meta-analysis: Prevention, assessment, and adjustments. West Sussex, England: John Wiley, pp. 75-98.

Straus SE, Kastner M, Soobiah C, Antony J and Tricco AC (2016). Engaging researchers on developing, using, and improving knowledge synthesis methods: introduction to a series of articles describing the results of a scoping review on emerging knowledge synthesis methods. Journal of Clinical Epidemiology. doi:10.1016/j.jclinepi.2016.01.031

Taris TW, Kompier M, Geurts S, Schreurs P, Shaufeli P, de Boer E, Sepmeijer K and Wattez C (2003) Stress management interventions in the Dutch domiciliary care sector:

Findings from 81 organisations. International Journal of Stress Management 10(4): 297-325.

Van der Klink, J. Blonk,R., Schene A and van Dijk F (2001) The benefits of interventions for work-related stress. American Journal of Public Health 91: 270–276.

Table 1: Steps of a realist synthesis on organizational interventions

Stage	Activity	Key questions
Define scope of synthesis	Define research questions	What are the key elements of the intervention process? What are the key elements of the intervention content (i.e. content of action plans)? What are the contextual elements? What outcomes are expected?
	Clarify the objective of the synthesis	Does the intervention achieve the intended outcomes? Which theories seem to be supported? How does the intervention work in different (cultural and occupational) settings?
	Identify programme theories	What are the key theories used? How can theories be grouped (e.g. according to content or process)? What are the most likely programme theories?
Search for evidence	Search for evidence	What sampling strategy is to be used? What are the search terms, the sources, and the methods to identify studies? When have we reached saturation?
Appraise the evidence	Test relevance of identified studies	Does the study test the theory in question? Do the results support the conclusions drawn by the authors? Do the results support or contradict our programme theories? Do we need to revise our programme theories?

Karina Nielsen is the Chair of Work and Organisational Psychology at the Institute of Work Psychology, Sheffield University Management School at the University of Sheffield, UK. She is a research affiliate at the CPH-NEW, USA and Karolinska Institutet, Sweden, Her research interests lie within the changing organizations. She is particularly interested in the evaluation of organizational interventions and ways to develop methods to understand how and why such interventions succeed or fail. She is currently on the editorial boards of Human Relations, The Leadership Quarterly and Journal of Business and Psychology and is an

associate editor of Work & Stress. She has published her work in journals such as Human Relations, Work & Stress, The Leadership Quarterly, and Journal of Occupational Health Psychology. [Email: k.m.nielsen@sheffield.ac.uk]

Dr Mariella Miraglia is a lecturer in Organisational Behaviour at Norwich Business School, University of East Anglia, UK. Her main research interests lie in the field of organisational behaviours, human resource management, and occupational health psychology with a special focus on absenteeism and presenteeism, performance appraisal and management, coaching, job crafting, psychological capital and self-efficacy. She has published in journals like Journal of Vocational Behavior, Journal of Occupational Health Psychology, and Human Resource Management. [m.miraglia@uea.ac.uk]

Corresponding Author:

Karina Nielsen

Professor of Work and Organizational Psychology

Institute of Work Psychology

Sheffield University Management School

Conduit Road

Sheffield

S10 1FL

UK

k.m.nielsen@sheffeld.ac.uk

Other Author(s):

Mariella Miraglia

Norwich Business School

University of East Anglia

Norwich

NR4 7TJ

UK

m.miraglia@uea.ac.uk