Reconsidering context: six underlying features of context to improve learning from evaluation

Abstract

This article considers the role of context in 'theory-based' evaluations, particularly those that use chain-type path or logic models. Reflecting on the use of causal models in the school professional development field, a set of underlying features of context is developed: the article proposes that context can be dynamic, agentic, relational, historically located, immanent and complex. The article applies these features to a consideration of a commonly observed contextual factor: senior leader support for an intervention. The article argues that actively considering these underlying features can allow for a more sophisticated approach to context, and concludes with a set of related interrogatory questions for evaluators, aiming to improve learning in future evaluation.

Key words: logic models; context; theory-based evaluation

Introduction

From the early days of evaluation design, evaluators have recognised that the success or otherwise of a programme, intervention, project or approach depends on the specifics of the context within which it plays out. This article is borne from a sense of the shortcomings of my own application of evaluation models and the role of context within them. From my perspective as an evaluator of educational programmes working within the theory-based evaluation tradition, such evaluations often uncover a very similar set of contextual factors in each case, yet the learning that accrues is limited. So, for example, it is usually the case that well-motivated participants in programmes benefit more than others; yet on its own this is a pretty trivial point that helps move forward neither the evaluation field, in relation to education or more broadly, nor the programme in question.

This article considers why this might be the case and how we might improve matters, examining the role of context in evaluation, especially the family of 'theory-based' evaluations and in particular path or logic models. It presents an argument that even where evaluators carefully consider context, such approaches can lead to it being treated in relatively simplistic ways.

This article addresses this issue by drawing on work in the field of educational evaluation and broader theory-based evaluation to propose a set of inter-relating underlying features of context that can allow a more sophisticated consideration of context in a way that is novel in the field.

Underlying features of context and observed contextual factors

To develop the argument in the paper, I first need to distinguish between observed contextual factors and the underlying features of the context within which interventions take place. Observed contextual factors are aspects of the context that are found empirically to influence the implementation and outcomes of a programme or initiative. Observed contextual factors necessarily vary from initiative to initiative, but they are recurrent and there have been attempts to classify them previously. Most recently, Vanderkruik and McPherson (2017) synthesised earlier frameworks from implementation science and improvement science to present a classification of contextual factors at two levels. The primary level in their framework (ibid, p.351) consist of external environment, the organisation, the initiative and the 'site/local team' involved in implementation, each of which has secondary levels (for example organisational support and capacity; organisational relationship to the initiative; and organisational culture and engagement are secondary levels below the primary organisational level). Each of these secondary levels is associated with example components such as political climate (secondary level component of external environment) and 'clear team roles and responsibilities' (secondary level component within site/local team). These 'secondary level components' represent what I refer to as observed contextual factors.

This paper argues that such observed contextual factors have underlying features, which can help explain how they act to influence the programme or initiative that is subject to evaluation. A set of these features is drawn out in the body of the article.

Whilst developing this set of underlying features, the paper provides evidence of how their systematic consideration can support stronger, more insightful interpretation of the presence of observed contextual factors by demonstrating the tangible benefits in my own work and that of others in the educational evaluation field. The paper then aims to show how they can be more broadly beneficial by applying them to an observed contextual factor that is common to a number of fields: senior leader support, using empirical evidence from a range of studies from my own research and that of others. It is important to note at this point that whilst these features have been recognized previously, and some are widely used in evaluation practice, they have not previously brought together as features of context that can help explain how observed contextual factors can operate in relation to a programme or intervention; and they have not been systematically considered in a way that can enable their use together to support evaluation.

This is the core aim of the remainder of this paper: to abstract and carefully consider a set of underlying features of context which can be applied more widely to evaluation designs across social policy fields to help improve evaluators' interpretations of how the context influences the working of the intervention and its outcomes, and thereby improve future evaluation design and analysis. To enable evaluators to make practical use of these features in evaluation, the paper suggests a related initial set of interrogatory questions that could be used across social policy fields.

The remainder of this paper focuses on a set of six underlying features of context, but I begin by noting that in theory-based evaluation context is always understood as having two additional features. Firstly, the context for an initiative is always situated in relation to its spatial location in and around the places where it is plays out. Secondly, it is temporally located in the present; the period during which - and shortly before and sometimes after - the initiative takes place. I do not address these contextual features in further depth as they are so intrinsic to the usual meaning of context in evaluation (as defined in the next section, for example by Greene, 2005 and Pawson, 2013) that they are always considered and therefore do not require further discussion, except in relation to their intersection with the six features discussed in the article.

The main focus of the paper is on the 'theory-based' group of evaluation models that all involve "*some attempt to 'unpack' the black box so that the inner components or logic of a program can be inspected*" (Astbury and Leeuw, 2010 p.364). The term 'theory-based' (Weiss, 1995) is used in this paper to represent this tradition, which also incudes other models such as 'theory-driven' (Chen, 1990), 'white box' (Scriven, 1994) and 'programme theory' (Rogers, 2008) evaluation. As indicated in the previous section, the paper looks in particular at models that use chain-type path models and, by extension, logic models. The paper argues that whilst contextual features are routinely considered in such models, they are often constructed, both visually and conceptually, in a way that can over-simplify how the context for an initiative can affect its outcomes.

To do so, the paper builds on this earlier work to develop a set of features of context to improve evaluation design, by focusing on the treatment of context in relation to causal models used in a field within which I work, professional development in educational settings. The arguments that are presented develop from my own application of evaluation models and those of others located in relation to education and wider evaluation fields.

Conceptualising context

Both the concept of context (using a variety of terms) and the term context (with a variety of meanings) have been present in the evaluation literature from the start. For example, in the early 80s Stufflebeam's (1983) CIPP (context–input–process–product) model used the term context to describe "*information about the strengths and weaknesses of a total system to assist in planning improvement-oriented objectives at each level of the system*"; and Cronbach's (1982) utos (units of focus, treatments, observations/outcomes, settings) used the term 'setting' as the key contextual variable in the model.

Theory-based evaluation design developed from the 70s onwards, with the emergence of frameworks such as programme logic models (Rogers et al., 2000) and theory of change (Connell et al., 1995; Connell and Kubisch, 1998). In such designs, the concept of context appeared as *conditions* that are necessary and/or sufficient for a theory to be enacted. Terms such as drivers and barriers; inhibitors and supporters; and enablers and disablers are sometimes used - for example Pawson and Tilley (1997, p.70) suggest that "*a crucial task of evaluation is to include (via hypothesis*

making and research design) investigation of the extent to which these pre-existing structures [i.e. aspects of context] 'enable' or 'disable' the intended mechanism of change".

In this article I follow Pawson and Tilley (1997, p.57) in taking context to be the social and cultural conditions within which programmes, initiatives or interventions occur. Such conditions include both the structural - organisational, spatial and temporal - setting and the individuals involved, including their personal characteristics and inter-personal relationships, further developed by Pawson (2013, p.37) as 'the 4 I's': individuals; interpersonal relations; institutional settings; and infrastructure (the cultural, economic and social aspects of the setting). This approach to context is consistent with other theory-related approaches: for example, Stame (2004, p. 63) suggests that an area of commonality between the approaches of Chen, Weiss and Pawson and Tilley is that they all "*consider programmes in their context, which includes actors' environments (embeddedness) and public service culture and behaviour*". This aligns with Greene's (2005) definition of context as "*the site, location, environment, or milieu for a given evaluand*" (p. 83) and Rog's (2012) "*broader environment*" rather than other context areas Rog identifies such as problem context (using features of the issue to be evaluated as context) or evaluation context (considering factors such as budget and time constraints as context). It also differs, as Fitzpatrick (2012) points out, from Stufflebeam's (1983) use of context in his aforementioned CIPP framework which refers to evaluation in the first phase of a programme looking at identifying programme and participant needs, and as such is more akin to Rog's (2012) decision-making context.

The position taken in this paper is aligned with what Greene (2005) identifies as the broadly theory-orientated perspective in this tradition, that context works to aid explanation, rather than acts as something to be controlled for (the experimental evaluation perspective) or as wholly inseparable from the specific programme (an interpretivist position).

Approaches to context in educational evaluation: building the set of underlying features of context

In this section, I develop the argument for a set of features of context showing how a group of longstanding evaluation approaches widely used in educational evaluation - 'path models' - can underplay these, linking them with parallel logic model approaches used in the wider evaluation field.
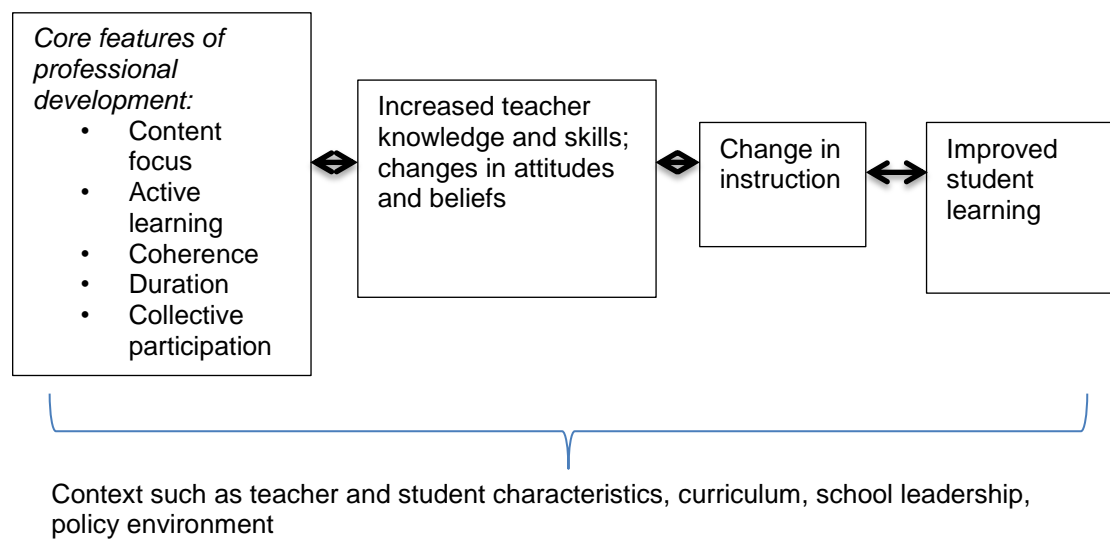
Path models (not to be confused with the statistical approach of path analysis) have been used in evaluation of professional development in educational contexts for a number of years. A fuller discussion of the development of such models is provided by Coldwell and Simkins (2011), but, in brief, (Kirkpatrick's (1998) work which began in the 1950s provided the genesis of the approach which has been developed in particular by Guskey (1999, 2002) since the mid 1990s to represent the causal process by which teacher professional development activities can lead to sought-for outcomes via a series of intermediate stages.

The culmination of this tradition is the path model presented by Desimone (2009). Drawing on a comprehensive review of international literature on the impact of

professional development, Desimone's model (Figure 1) provides evidence of a set of

relationships between steps in a path from professional development on the one hand

via changed teacher knowledge, skills, beliefs and attitudes to improved pedagogical

practice and then to student learning outcomes.

Figure 1: Desimone's (2009) Path Model

Context such as teacher and student characteristics, curriculum, school leadership, policy environment

This model, along with others in the same tradition, shares an approach with, and can

be seen as a specific instance of, logic models in the broader evaluation field.

Desimone (2009, p.184) refers to a path model as an "*operational theory*" that can be

used to "*identify the key inputs and intermediate and final outcomes that characterize*

the effects" of a professional development intervention, and also "*identify the*

variables that mediate (explain) and moderate (interact to influence)" a programme or

intervention's effects. This is clearly aligned with what Rogers (2008, p.33) calls

"*simple logic models*" which "*show a single, linear causal path, often involving some*

variation on five categories (inputs, processes, outputs, outcomes and impact)" such

as the widely used Kellogg and Wisconsin models - although in Desimone's model the

9

first categories tend to be merged together, and the outcomes differentiated as part of a causal outcome chain.

Context appears as a box at the bottom, and is seen to operate "*as an important mediator and moderator*" (p.185), in common with some of the other models used in this field (see, for example, Leithwood and Levin, 2005; Simkins et al., 2009), the box being filled with a set of contextual factors that have been or are likely to be observed. Such an approach is also common in visual representations of logic models in the wider evaluation field; for example, in the Wisconsin model, the same kind of box contains 'external factors' known to be associated with higher likelihood of the path leading to successful outcomes.

This approach to presenting and considering context is widely used in this particular tradition of research into professional development in education, with research (much of which is cited by Desimone, 2009) providing evidence of the importance of observed contextual factors such as school leadership 'buy in'; school culture; attitudes of those engaged; and policy pressures; in relation to the success of otherwise of the implementation of new initiatives. Experienced evaluators can suggest a list of such factors before needing to set foot in the field and can confidently expect to find them once data gathering commences.

This brief overview introduces the first step in this paper's argument, which is that the visual presentation of observed contextual factors in a box at the bottom (or at the top, or in a circle around the model) can lead, if evaluators are not careful, to an over-simplification of the ways in which observed contextual features operate in relation to

the programme at hand. To build this argument, and draw out how the underlying features of context that can help avoid this over-simplification, the rest of the subsection considers some common observed contextual features that emerge in the education field, starting with school culture.

In Desimone's model, in line with many others in the same family, school culture is discussed in a rather static way, as a characteristic of the setting within which the intervention occurs; as indicated above it is seen as one of a set of "*contextual factors at the classroom, school and district levels*" (p.185). Yet we know from other research in the field that school cultures are both complex and likely to change (and school improvement literatures highlight how cultures can be changed - see for example Hargreaves et al., 2014; Harris et al., 2013). School cultures are not static, they are **dynamic**. This is the first underlying feature of context.

Furthermore, organisational cultures are open to change as a result of interventions. As Blamey and Mackenzie (2007, p.441) note, this can be deliberate: often "*context is not simply an interesting backdrop but is instead explicitly targeted for change*". For example, Simkins et al. (2009) show how some interventions - in this case school leadership programmes - can lead to changes in organisational capacity to effect further change within the setting, which itself can lead to cultural change. These examples indicate how organisational features can operate both as a context for and as an outcome of interventions. In earlier path models this subtlety is not clear. For example, in Guskey's (2002) path model the context is not explicitly included: although the paper refers to 'a range of situational and contextual variables' (2002, p. 387), these are not developed further. A solution provided by some later path models

in the education field is to present observed factors that act in this way, such as school culture, at different points in the model, both in the context and outcome path points [or boxes].In the example above, Simkins et al. (2009) include 'capacity' to provide effective leadership at two points in the model, both as a potential outcome and as a contextual factor. However, this visual duality can miss that, as programmes develop, school cultures, capacity to change, and support for the programme can change too, qua observed contextual characteristics rather than qua outcomes. For example, as early benefits become apparent, sceptical senior leadership teams can be won over. Thus the path to achieving outcomes can become easier for those involved; or, conversely, it might become more difficult.

This introduces the second underlying feature: the contextual components in the examples above are not merely dynamic but independently **agentic**. The term agentic is used to denote that actors and groups of actors can work to create changes independently from the programme at the same time as influencing the programme itself. Thus senior leaders can act to improve classroom practice in mathematics, say, in a variety of ways in addition to acting to support or inhibit any particular intervention. This account can be extended to include actions of others even further removed from the programme at hand, for example the role of policy as both a constraint on a particular intervention, and as a driver of actions by others in the system independently of the programme at hand. If the agentic nature of context is underplayed or ignored, then the relative importance of the intervention versus wider change process can be missed. Taking a high profile example from education, accountability pressures especially the use of high stakes testing can drive

organisational practices i such as increasing curriculum time spent on tested subjects

and using more teacher-centred pedagogical approaches (Au, 2007).

Taken together with the fact that contextual factors can also be the target of

interventions, this indicates that such contextual factors as school culture are

relational: they act in particular ways in relation to the programme and aspects of it.

Clearly, contextual factors always act in relation to an intervention; I am using the

term here to denote that there are specific ways in which observed contextual factors

can operate relationally. Firstly, observed contextual factors do not operate in a

uniform way as an intervention plays out, in the way implied by the visual

presentation of logic models with a single 'context' 'external factor' or 'moderating

factors' box. Coldwell (2017) demonstrates how observed contextual factors

differentially influence different points in the path. For example, a group of observed

individual factors including positive motivation to engage in professional

development and engagement at early career stages were associated with higher career

aspirations; and "*some school and leadership cultures provided a more positive*

environment within which to work and develop, influencing intention to stay in

teaching and career aspirations." (p. 196).

Secondly, since context is dynamic and agentic, as I noted earlier, observed contextual

factors can operate in concert with or against other factors. For example, Simkins et

al. (2009) provide evidence of mutually reinforcing connections between high quality

coach support with motivation to succeed in relation to positive outcomes from a

programme to improve leadership capacity for school middle leaders.

Much of the discussion above relates to organizational aspects of context, but this equally applies to individual contextual factors; for example, individuals will have attitudes towards what constitutes effective professional development for them, based on prior experiences. Such attitudes will influence their actions independent of any programme they might be involved with and will inform their responses to and engagement in any such programme. And such attitudes can, of course, be changed by the programme.

The characteristics of individuals engaged in programmes help demonstrate another feature of context that can be difficult for path models to deal with: context is not only spatially located but temporally **historically located**, and may be subject to wider change processes that can occur over a very different time span to that of the programme subject to evaluation. 'Historical' is used to distinguish this underlying feature from 'temporal' location since, as noted above, all evaluations treat interventions as being temporally located in the short term, by which I mean the period leading up to the intervention being enacted, the period during which it is enacted and sometimes shortly afterwards. Since logic models focus on the short term in this way, they are liable to miss that there can be patterns in the ways in which observed contextual factors work that relate to longer term change processes.

To illustrate this, I return to the Desimone model. As with 'school context', those involved in evaluating school-based professional development programmes can highlight a set of such observed contextual factors relating to individual characteristics that are associated with the likelihood of sought-for outcomes occuring. Desimone (2009, p.185) lists "*experience, beliefs, knowledge and attitudes*"

for example, and others might point to career phase, orientation to the programme and approach to professional development more widely - which are a treated analytically as being held steady as the professional development model plays out[1]. Yet the literature on teacher development identifies that teachers develop their identities over long periods of time, moving through what Day and Gu (2010) call '*professional life phase*s', and their attitudes and responses to professional development are likely to vary in relation to these phases (Huberman, 1995; Sikes, Measor& Woods, 1985). Even over short periods, teachers can alter their orientations to their career: for example, Coldwell (2016) provides evidence of beginning teachers' changing orientations to promotion, life in the classroom and their work-life balance, all of which will intersect with their approach to professional learning opportunities. And these orientations will change as a result of the programme (as can be seen, teacher knowledge, skills, attitudes and beliefs are built in as intermediate outcomes of the Desimone path model). These differences between responses to PD for individuals in different 'professional life phases' are liable to be missed by path models and simple logic models that focus on the short term.

Pawson and Tilley's (1997) scientific realist approach treats change mechanisms, which are akin to Desimone's paths, as always and only occurring in certain contextual circumstances - hence their use of the term 'context-mechanism-outcome' combinations, expressly indicating that the mechanism is bound together with context.

---

[1]They can sometimes also be treated analytically as outcomes in such models as noted above; but in their guise as contextual factors they are treated as relatively unchanging

A feature of this approach is the centrality of the decision-making of those 'subjected' to programmes. Pawson gives a useful example here in the context of crime reduction initiatives:

> "it is not programmes that work but the resources they offer to enable their subjects to make them work. […] let us consider the causal powers of programmes offering 'transitional payments' to prisoners on release with the aim of preventing the need for a quick return to crime. In such cases, it is not the programme that causes 'rehabilitation'. It merely provides payments, which the subjects choose to use in different ways, one of which might be to steer away from crime."
>
> (Pawson, 2002, p.342)

This perspective helps bring out another aspect of context associated with the individuals involved: they make choices about how to behave. An evaluator treats some of these behaviours as potential outcomes for an intervention, as in the quotation above. But from the individual perspective, the intervention forms just a part of a set of factors - integrated together - that influence decision-making. At the risk of using his work as "*intellectual hairspray, bestowing gravitas without doing any theoretical work*" (Reay, 2004, p.432), Bourdieu would suggest that the prior experiences, the physical and temporal location, of an individual create a habitus: "*a 'practical sense' that inclines agents to act and react in specific situations in a manner that is not always calculated and that is not simply a question of conscious obedience to rules. Rather, it is a set of dispositions which generate practices and perceptions. The habitus is the result of a long process of inculcation, beginning in early childhood,*

which becomes a 'second sense' or second nature." (Bourdieu, 1993, p.5). For teachers, this might mean that their educational world has been so orientated towards using a particular pedagogical approach that the possibility of changing it in response to a professional development experience such as a training course would require such a shift in world view as to be almost impossible for them (which Ball (2003) argues can occur for teachers that have spent their entire careers working within a 'performativity'-driven system). For others, their habitus might mean the time is right for such a change to occur.

From this perspective, then, the individual context is not just bound up with the intervention and its workings; the intervention is itself assimilated into the individual's decision-making alongside its context. So, in relation to individual agency and decision-making, the context of a programme constrains the decision making of individuals by way of the underlying feature of being **immanent**. I use the term immanent to convey both that contextual factors permeate programmes rather than remain external, and that such factors are internalised by the participants. It is not just that the programme doesn't hit the participants like billiard balls: neither does the context.

Just as alternatives to simple logic models have been proposed in the wider evaluation field, alternatives to single path models have been developed in educational evaluation to try to respond to the set of difficulties associated with treating context as external and static. For example, one particularly influential model is provided by Clarke and Hollingsworth (2002)'s causal "interconnected model of teacher professional growth" (Figure 2). This teacher change model treats the immediate

school context as one 'domain' of change which interacts with others to create

professional learning outcomes. The model can be seen as a kind of multiple

pathways model, that - rather than presenting a single path from professional

development to outcomes - acts as an analytical tool for mapping different routes that

may result for any individual undertaking professional development activity.

Figure 2: Clarke and Hollingsworth's (2002) interconnected model of teacher professional growth (p.951)
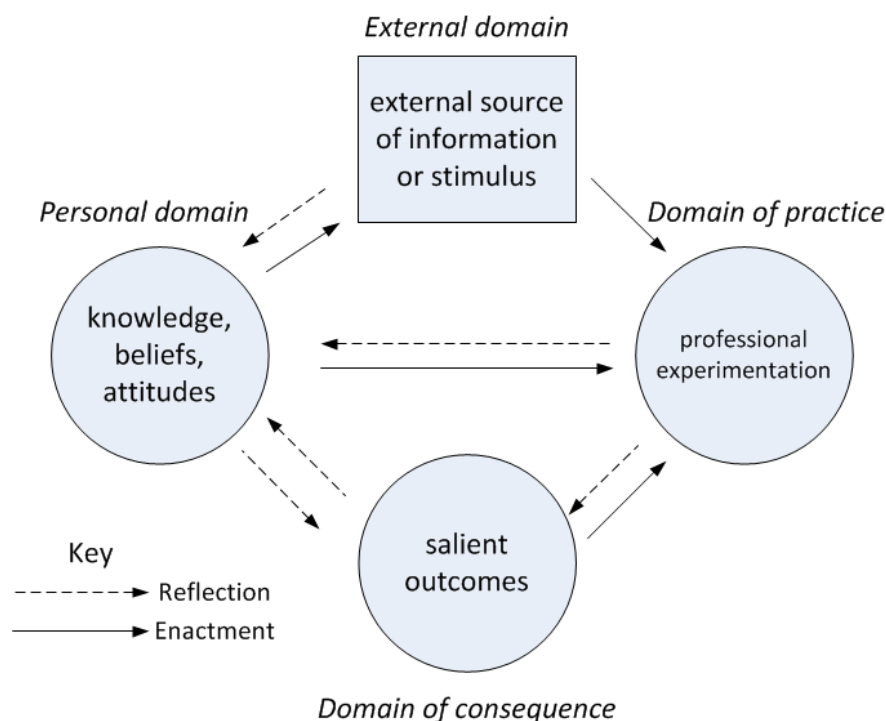
This type of model can be seen as addressing some of the problems Rogers (2008)

identifies with simple logic models, in particular the potential for simultaneous causal

strands and for alternative causal strands to operate in more complicated evaluation

scenarios. Most pertinently for this article, in addition to allowing flexibility, this kind

of approach attempts to wrap two types of observed contextual factors - external

change factors, and factors related to the individual - into elements of different 'domains' (the external domain, and the personal domain, respectively). It should be noted that the external domain is restricted, and does not include some contextual factors, such as the political environment. This model therefore seeks to treat at least some aspects of context as both intrinsic to the model and dynamic.

However, in common with simple logic model (and professional development path model) designs, the domain model treats the relationship between the context and other parts of the model as part of a presumably observable pathway. Yet the world appears to operate in a more complex way than this, introducing the final underlying feature of context identified in this paper: it can be **complex.** There are myriad definitions of this term, but we might use Walton's (2016) approach to frame complexity as including a set of core features which include non-linearity; emergence; adaptation; and uncertainty. This is more restricted than the definition used by Pawson (2016), who argues that complexity is a defining feature of public policy evaluation - and is by implication an overarching underlying feature of context. The features I identify above - especially that context is agentic and dynamic, relational and historically located - are recognisably aspects of complexity in Pawson's terms. However, they are also features of simple and complicated systems, drawing on Rogers' (2008, p.32) distinction between what is simple (involving single linear paths), complicated (involving multiple causal strands, organisations and mechanisms) and complex (recursive and emergent). Therefore in this paper I have treated these earlier underlying features as analytically distinct from complexity.

In the professional learning field, evaluation models have not yet developed that deal with complexity, but wider professional learning models are of use. In particular, Opfer and Pedder (2011) draw on systems and complexity theory in an article that argues that professional learning to occur by way of interactions between orientations towards what they refer to as teachers' 'learning activities', i.e. professional development or professional learning experiences. They argue these interactions occur within "*the system of activities in which teachers engage and the systems of influences that mediate and moderate these activities, teacher learning, and teacher change*"(Opfer and Pedder, 2011, p.386).These orientations and their interactions are laid out as:

- Teacher orientation to the learning activity

- School-level orientation to the learning activity

- Interactions between teacher and school-level orientations

This captures the emergent character of the context within which change processes occur via interactions between different system elements, but it is not straightforward to operationalize this in evaluation designs. In the wider evaluation field, Rogers (2008) presents a number of alternative visual representations as part of a review of how evaluation designs can deal with complexity in relation to programme theory. For example, circular visual paths are suggested to deal with feedback loops and tipping points might be addressed via annotating the model. Subsequent work in this vein has been developed by Walton (2014; 2016).

Approaches such as those reviewed by Rogers and Walton are promising in that they can treat context as bound up with the intervention: complexity emerges as part of the operation of the intervention within its context. Building on Rogers' work in the programme logic model field, path and logic models can thus attempt to deal with complexity in a number of ways. One approach is to develop a set of interlocking models at different system levels that merge together so, taking Opfer and Pedder's example, a path at the teacher, at the organisation and at the activity system levels would each be developed coming together at the stage of outcomes for school, teacher and pupils. Dealing with unexpected emergent outcomes, tipping points and feedback loops necessarily requires revision of paths. The added role of complexity of the contextual circumstances within which programmes play out is difficult to deal with, indicating the limits of path model approaches in coping with complexity.

This suggests that path and logic modellers need to recognise that the complexity of the social world is such that there will be significant change processes occurring over different time scales, at different system levels, that interact with programme effect to lead to differential outcomes, and to pay attention to them.

Taking an example relating to teacher careers and development, scholars working in the field of Human Resources trace a complex relationship between longstanding, significant changes in economic production practices (such as deindustrialisation), and changing patterns in consumption and markets (emerging marketisation and consumerism) to changes in individuals' approaches to their careers. Hall (2002) argues that alongside traditional, organisationally-focussed approaches to career, there has emerged what he calls 'protean' career orientations: individuals taking charge of

their careers. Overlaying this set of changes onto an evaluation of a professional development programme in a school might lead a path model theorist to find that senior leaders recognise teachers hungry for promotion and development opportunities in the school they find themselves in, and treat this as a personal context variable - as, indeed, I found in a study of early career teachers. This was of interest, not least because these larger scale change processes and teacher responses to them were opaque to senior leaders who often responded to the protean career approaches of teachers looking for new opportunities in a frustrated way: "*they expect more. They need to be reminded they are lucky to have what they are given!*" in the words of one (Coldwell, 2016 p.618). Domain modellers drawing on Clarke and Hollingsworth may treat the changing career expectations of individuals as part of the individual domain operating alongside the professional development. Yet this approach and path/logic models struggle to capture the emergent nature of these changes that only occur due to a complex interaction of activities at different system levels. Alternatives such as Theory of Change approaches (Connell and Kubisch, 1998) may be more useful useful here, since whilst they focus on identifying intermediate and longer term outcomes, the processes by which such outcomes are expected occur are flexible and should be reviewed. This allows for emergent patterns and outcomes to be brought into the model at later stages. Mason and Barnes (2007) note that Theory of Change approaches as used in the UK tend to be introduced after the programme has begun to operate, and advocate for continuing review and development of with a range of stakeholders on an ongoing basis, an approach which is particularly suitable in dealing with emergent outcomes and unpredictability.

Implications

Approaches based on simple logic models (and others with similarities such as Desimone's 2009 path model) act as a bedrock of many programme evaluation designs. Yet the implication of the analysis above is that such models may underplay the nature of the context within which programmes occur. In particular, whilst they focus on the spatial and shorter term temporal features of context, they can fail to capture that context can be:

- **dynamic**, changing over time and therefore potentially changing how they influence interventions

- **agentic**, creating not simply moderating change

- **relational,** acting both as context for and outcome of the work of initiatives; and acting in concert with or against the work of the initiative

- **historically located**, involving change processes over a much longer time period than the initiative at hand

- **immanent,** acting through - and as an intrinsic part of - participants' responses to the programme, not external to it

- **complex**, leading to changes that arise out of complex change processes at different system levels that interact with programme processes.

Clearly, as indicated at the beginning of this article, evaluators are not unaware of these inter-relating features of context, indeed much has been written about them in different spheres and sometimes in relation to the models evaluators working in the theory-based evaluation tradition typically use. For example, Rogers (2008, p.34) notes that:

> "By leaving out the other factors that contribute to observed outcomes, including the implementation context, concurrent programmes and the characteristics of clients, simple logic models risk overstating the causal contribution of the intervention, and providing less useful information for replication."

And, more recently, Pawson (2016, p.49) states that

> "Context is layered. Sometimes it is pre-existing, macro economic conditions that need to be auspicious to forward a policy. Sometimes it is institutional norms that need to be supportive to enable change. Sometimes it is cultural practices that need to be consonant with a new programme. Sometimes it is the prevailing interpersonal relations that need to be favourable for an intervention to work."

However, the underlying features of these observed contextual factors - institutional norms, cultural practices, interpersonal relationships and so on - proposed above have not been previously presented together and systematically considered. Systematic consideration can improve the potential learning about an intervention's context and how it is likely to operate in particular circumstances. Table 1 highlights the

implications of using the underlying features in considering how observed contextual

factors operate, comparing them with the alternative of ignoring such features.

Table 1: Implications of using the underlying features, and their alternatives

If Context is treated as...	The implications are…	Alternative: if context is treated as…	The implications are…
Dynamic	Contextual factors may change shape over the course of the evaluation, operating differently as these changes occur.	Static	Contextual factors are considered as moderating influences or potentially as outputs, without considering changes in contextual factors over the course of the evaluation.
Agentic	Contextual factors may themselves act independently of the evaluated initiative to lead or contribute to changes.	Passive	Changes in contextual factors may be considered, but their independent role as an instigator of, or contributor to, causal processes are not considered.
Relational	Contextual factors can influence different elements of the change processes evaluated, in concert with or against the aims of the initiative and with or against the influence of other factors.	Uniform	Contextual factors are considered largely in relation to the evaluation as a whole, and independently of other contextual factors.
Immanent	Contextual factors work through the initiative being integrated with other factors in informing the decisions and actions of programme participants.	External	Contextual factors are considered to operate separately from the evaluation, acting as a barrier to or support for actions.
Historically located	Contextual factors are considered in relation to long term change processes.	Temporally located in short term	Contextual factors are considered in relation to the recent past and future.
Complex	Contextual factors may work in non-linear ways, potentially leading to adaptation, feedback loops and emergence of changes from factors operating at system levels interacting.	Simple	Contextual factors are conceived of as working in linear ways.

There are a set of potential consequences of either considering or ignoring the features of context as laid out in Table 1 above. Firstly, if evaluators treat observed contextual factors as static then they may miss changes in the influencing effect of such factors; for example, leadership cultures may change to be more positive about a programme as it develops. Secondly, if contextual factors are considered as agentic rather than simply as passive influences on the working of the programme, then evaluators can explicitly look for the independent impact of the programme in relation to other possible causal explanations, preventing misrecognition of causal influences especially where there is no counterfactual. Thirdly, treating the effects of contextual factors on an intervention as uniform rather than relational can miss the differential influence of observed contextual factors on different elements of the implementation of a programme. Fourthly, if contextual factors are treated as external rather than immanent and internalised then evaluators (and policy makers) can underestimate the difficulty faced in effecting change, since altering decision-making is more difficult than removing barriers. Similarly, ignoring the historical location of contextual factors can lead to underestimating deep-rooted issues that can be very tough to change. Finally, ignoring the complex nature of contextual factors can lead to evaluators missing emergent outcomes that are not predicted by logic models, ignoring feedback loops and missing adaptation in how observed contextual factors influence how programmes operate.

Applying the underlying features to an exemplar observed contextual factor: senior leader support

To illustrate how this set of underlying features can be useful in the way suggested in the previous sections, this section of the article considers a classic observed contextual factor: *senior leader support*. This factor is particularly important in educational settings because of the international trend towards increased autonomy for schools (for example, the development of schools outside of local district control such as Charter Schools in the USA and academies and Free Schools in England) within which, as relatively small organisations, the head teacher or principal has a very strong influence. However, increasing accountability of public sector and third sector organisations across public policy areas (Pollitt & Bouckaert, 2011) has raised the pressure on senior leaders in other spheres. So, across policy areas and organisations, buy-in of senior leaders is associated with stronger likelihood of success for an initiative. On its own, this finding is of very limited usefulness. As indicated earlier, evaluators (especially those working in fields like health and education where organizational settings are important features) already know this will be the case before they gather any data at all. But if we consider the dimensions of context outlined above, and their application to the specific reasons behind how and why senior leader support is present or not in relation to a particular evaluation, then the evaluator can begin to gain a stronger understanding of the initiative and therefore produce more insightful analysis.

To help demonstrate the utility of the underlying features of context, I treat them separately. It is important to note that the dimensions are inter-relating so in real world evaluation they would more typically be linked and considered together.

Starting by thinking about the **dynamic** nature of some contextual factors, the evaluator is led to consider that senior leaders do not have a static orientation to a particular initiative: there will be temporally and spatially situated reasons for their perspectives. To take a particularly instrumental example, there may be prior experience of poor quality mentoring designs - for example, a lack of focus on the outcomes of the programme (Allen, Eby& Lentz, 2006) - that has led to leaders mistrusting mentoring approaches. Since senior leaders' views may be open to change (Heifetz, Grashow & Linsky, 2009) this leads the evaluator to consider that if the programme can be shown to be successful, then senior leader orientations may change, indicating the need to gather data to establish under what circumstances this tends to occur.

Considering the **agentic** nature of context, it is certain to be the case that senior leaders are putting in place other actions alongside the initiative that aim to achieve the same ends as the initiative under evaluation. For example, a change in ward procedures in a hospital will take place alongside a whole host of other small and large changes all aiming to improve patient outcomes, which need to be paid attention to.

Furthermore, the contextual influence of senior leader support is **relational**. Kunzleet al.'s (2016) review of effective leadership behaviours in relation to patient safety identifies that senior leader effectiveness is both a factor in the success of an initiative and an outcome from it, which can then affect the likelihood of future success. For example, if change-orientated leadership behaviours emerge from earlier interventions, this is likely to influence the degree and efficacy of leadership support

(Yukl, 2006). More broadly, other programmes may be present to actively support effective leadership behaviours, such as encouraging a cooperative organizational climate and focussing on developing team members (Kunzleet al., 2006). Focussing on the presence or otherwise of other such programmes or initiatives and their outcomes as relevant to the programme at hand may shed light both on how and why leadership support is enacted, and to what extent the programme team may be able to influence relevant leadership behaviours.

Finally, considering the **historically located, immanent** and **complex** nature of context, senior leader opposition (or buy-in) to an initiative may be related to long term, complex relationships between processes at different system levels. For example, longstanding policy positions moving organisations into more managerialist approaches with a strongly performative accountability regime can affect senior leader support for particular forms of professional development. In a review of professional learning across health, education and other fields, Webster-Wright (2009, p.703) identified a mismatch between what is known about effective professional learning experiences, which she identifies as "*actively working with others on genuine problems within their professional practice*" and "*continuing, active, social, and related to practice*" with those approaches favoured by senior leaders - identified as "*episodic updates of information delivered in a didactic manner, separated from engagement with authentic work experiences*". She identifies factors including the tendency for increasing control to meet organisational targets, and changes in professional expectations from leaders, which can lead to a divergence between senior leaders' expectations of professional learning activities and those that prior research

indicates are most powerful. If an evaluation of a professional development programme takes note of these important contextual features that can lurk below the surface of senior leader responses to such programmes, then the evaluator can better understand the reasons behind the programme's success or failure.

Discussion: utilising the underlying features of context to inform evaluation

One possible way to support the use of these underlying features is to apply a set of what might be termed interrogatory questions at the outset of an evaluation, which may then subsequently be revisited, to stimulate consideration of the ways that the context of an intervention might operate taking into account the six underlying features. This approach has been used by others: for example, Walton (2016, p.76) provides an initial set of questions to consider to what extent an evaluation more broadly takes into account complexity including whether it identifies features such as "*Forms of feedback that constrain or support change*", "*Initial conditions that affect interactions within the system*" and "*Interactions between levels of the system*". An initial set of such questions in relation to the features of context outlined in this paper is presented in Table 2.

Table 2: Interrogatory questions arising from the six features

In relation to the nature of context as…	Consider these questions: to what extent and in what ways…
dynamic	...are aspects of the context liable to change as the evaluation develops?

agentic	...are aspects of the context themselves liable to cause changes that are of relevance to the intervention being evaluated?
relational	... .are some contextual aspects likely to influence different parts of the change process differentially? [For example, participant buy in may be most important at the start of the initiative; the provision of adequate time to implement the initative most important during its delivery; etc} ...are different aspects of context acting: as moderating influences on the success or otherwise of the intervention; as independent agents of change that may operate in concert with or in opposition to the intervention; and/or as potential outcomes of the intervention?
historically located	...are aspects of the context subject to wider, longer term change process?
immanent	...are the decisions by actors engaged with the programme [as participants; providers; stakeholders] likely to be conditioned by their prior dispositions, and how?
systemic and complex [see also Walton, 2016, p.76]	...are aspects of the context liable to operate at different system levels [for example, local area; organisational; practice]? ...are context aspects liable to operate with the intervention to lead to emergent outcomes? …are aspects of the context likely to operate to create feedback loops and tipping points in relation to intervention outcomes?

Cleary, the list of underlying features of context developed in this paper are not complete; in particular, as I note above, I have not included the spatial and shorter term temporal features of context, which are normally addressed in logic model-type evaluation designs. For completeness, these could be added. In addition, the features

could be cut in different ways, and possibly grouped in some way. However, the set of underlying features of context abstracted here have not been brought together in this way previously to support their systematic use in future evaluation. Similarly the set of questions in Table 2 are incomplete and invite modifications and additions as relevant to specific evaluations, but are provided as a new addition to the field by bringing together these issues to help evaluators.

Beyond the use of such questions as laid out in Table 2, it is important to recognize that evaluations need not only to gather and analyse data, but to present these in a way that can influence change: drawing on Easterby-Smith's (1994) categorization of evaluation purposes, they need to do so to both help *improve* the programme at hand, and engender wider *learning*.

There are two issues here. The first is to what extent the leaders, deliverers and funders in the relevant evaluation or future evaluations are able to do anything about the context. Some observed contextual factors such as senior leader support and participant motivations are at least partly amenable to change within a programme setting: by providing convincing evidence of the value of the programme, or by incentives, for example. However, others such as accountability regimes - e.g. school inspection and attainment-based league tables - are both ubiquitous and simply not open to change, at least by the programme actors. However, even for those factors that may appear to be completely out of the sphere of influence of project agents there may be responses that can be made to take them into account. So, for example, in relation to the science teacher CPD programmes evaluated by Coldwell (2017), the deeper, structured contextual factors such as shifting long term career patterns might

need simply to be taken into account by understanding that this is likely to play out in relation to differing expectations of programme participants. So, programme designers can either modify the programme - to provide support for those not committed to teaching such as 'portfolio teachers', identified by Smethem (2006) as likely to move in and out of the profession, or amend the expected outcomes from the programme, anticipating that CPD will not lead to improved teacher retention for such teachers.

The second issue relates to those contextual factors that cannot be controlled, influenced or adequately taken in to account by programme leaders. In such cases, especially where they work to prevent the programme from leading to sought for changes, the question is raised as to whether the programme should be pursued at all in such circumstances. In realist evaluation terms, the particular CMO combination may not include sought for outcomes, so other initiatives should be considered.

These issues, of course, apply to all evaluation studies however they deal with context. But a conceptualisation of context in the way suggested in this article can help more informed judgments to be made in relation to interpretation and suggestions or recommendations for the future.

Conclusion

This paper has used a set of causal models in the professional learning field as a springboard for an argument that the context for programme and initiative implementation should be considered to be not only located spatially and temporally [in the short term] but dynamic, agentic, relational, historically located, immanent and

complex. By explicitly and systematically considering these features we can improve

the quality of our work as evaluators. If evaluation designs attempt to consider

observed contextual factors in relation to these underlying features in this way, then

our understanding of how persistent, common ways in which the context of an

initiative tends to lead to its success or otherwise may improve, leading to stronger

analysis and deeper transferable learning.

References

Allen TD, Eby LT and Lentz, E (2006) Mentorship behaviors and mentorship quality associated with formal mentoring programs: closing the gap between research and practice. *Journal of Applied Psychology* 91(3): 567-578.

Astbury B and Leeuw FL (2010) Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation* 31(3): 363-381.

Au W (2007) High-stakes Testing and curricular control: a qualitative metasynthesis. *Educational Researcher* 36(5): 258-267.

Ball SJ (2003) The teacher's soul and the terrors of performativity. *Journal of Education Policy* 18(2): 215-228.

Bhaskar RA (1998) *The Possibility of Naturalism* (3rd ed.) London: Routledge

Blamey A and Mackenzie M (2007) Theories of change and realistic evaluation: peas in a pod or apples and oranges. *Evaluation* 13(4): 439-455.

Bourdieu P (1993) *Sociology in question*. London: Sage.

Chen H-T (1990) *Theory-driven evaluations.* Newbury Park, CA: Sage.

Clarke D and Hollingsworth H (2002) Elaborating a model of teacher professional growth. *Teaching and Teacher Education* 18(8): 947-976.

Coldwell M (2016) Career orientations and career cultures: individual and organisational approaches to beginning teachers' careers. *Teachers and Teaching: Theory and Practice* 22(5): 610-624.

Coldwell M (2017) Exploring the influence of professional development on teacher careers: A path model approach. *Teaching and Teacher Education* 61: 189-198.

Coldwell M and Simkins T (2011) Level models of continuing professional development evaluation: A grounded review and critique. *Professional development in education* 37(1): 143-157.

Connell J and Kubisch A (1998) Applying a Theory of Change Approach to the Evaluation of Comprehensive Community Initiatives: Progress, Prospects and Problems In: Fulbright-Anderson K, Kubisch, A and Connell J (eds) *New*

Approaches to Evaluating Community Initiatives, vol 2, *Theory, Measurement, and Analysis.* Washington DC: Aspen Institute.

Connell J, Kubisch A, Schorr L and Weiss CH (1995) *New Approaches to Evaluating Community Initiatives*, vol 1, *Concepts, Methods and Contexts.* Washington, DC: Aspen Institute.

Cronbach L (1982) *Designing evaluations of educational and social programs.* San Francisco: Jossey-Bass.

Day C and Gu Q (2010) *The new lives of teachers.* London: Routledge.

Desimone L (2009) Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher* 38(3): 181–199.

Easterby-Smith M (1994) *Evaluating management development, training and education.* Brookfield, VT: Gower.

Fitzpatrick JL (2012) An introduction to context and its role in evaluation practice. *New Directions for Evaluation* 135: 7-24.

Gates EF (2016) Making sense of the emerging conversation in evaluation about systems thinking and complexity science. *Evaluation and program planning* 59: 62-73.

Greene JC (2005) Context. In: Mathison S (ed) *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage, 82–84.

Guskey TR (2000) *Evaluating professional development.* London: Corwin Press.

Guskey TR (2002) Professional development and teacher change *Teachers and Teaching* 8(3): 381-391.

Hall DT (2002) *Careers in and out of organizations.* Glenview, IL: Scott, Foreman.

Hargreaves A, Lieberman A, Fullan M and Hopkins DW (eds) (2014) *International Handbook of Educational Change: Part two.* London: Springer.

Harris A, Day C, Hopkins D, Hadfield M, Hargreaves A, and Chapman C (2013) *Effective leadership for School Improvement*. London: Routledge.

Heifetz RA, Grashow A and Linsky M (2009) *The practice of adaptive leadership: Tools and tactics for changing your organization and the world.* Cambridge, MA: Harvard Business Press.

Huberman M (1995) Professional careers and professional development: Some intersections. In: Guskey, TR and Huberman M (eds) *Professional development in education: New paradigms and practices* New York, NY: Teachers College Press, 193-224.

Kirkpatrick D (1998) *Evaluating Training Programmes: the four levels*. 2nd ed. San Francisco, CA: Berrett-Koehler.

Knowlton LW and Phillips CC (2012) *The logic model guidebook: Better strategies for great results.* London: Sage.

Künzle B, Kolbe M and Grote G (2010) Ensuring patient safety through effective leadership behaviour: a literature review. *Safety Science* 48(1): 1-17.

Leithwood K and Levin B (2005) *Assessing School Leadership and Leadership Programme Effects on Pupil Learning* Nottingham: Department for Education and Skills.

Mason P and Barnes M (2007) Constructing theories of change: Methods and sources. *Evaluation* 13(2): 151-170.

Opfer VD and Pedder D (2011) Conceptualizing teacher professional learning. *Review of Educational Research* 81(3): 376-407.

Pawson R and Tilley N (1997) *Realistic evaluation.* London: Sage.

Pawson R (2002) Evidence-based policy: The promise of realist synthesis. *Evaluation* 8(3): 340-358.

Pawson R (2013) *The science of evaluation: a realist manifesto.* London: Sage.

Pawson R (2016) The ersatz realism of critical realism: A reply to Porter. *Evaluation* 22(1): 49-57.

Pollitt, C and Bouckaert G (2011) *Public Management Reform: A comparative analysis-new public management, governance, and the Neo-Weberian state.* Oxford: Oxford University Press.

Priestley M, Biesta G and Robinson S (2015) *Teacher agency: An ecological approach.* London: Bloomsbury Publishing.

Reay D (2004) 'It's all becoming a habitus': Beyond the habitual use of habitus in educational research. *British Journal of Sociology of Education* 25(4): 431-444.

Rog DJ (2012) When background becomes foreground: Toward context-sensitive evaluation practice. *New Directions for Evaluation* 135: 25-40.

Rogers, P (2008) Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation* 14(1): 29-48.

Rogers P, Hacsi T, Petrosino A and Huebner TA (eds) (2000) *Program Theory in Evaluation: challenges and opportunities.* San Francisco, CA: Jossey-Bass.

Scriven M (1994) The fine line between evaluation and explanation. *Evaluation Practice* 15: 75-77.

Sikes P, Measor L, and Woods P (1985) *Teacher careers: Crises and continuities.* London: Falmer Press.

Simkins T, Coldwell M, Close, P and Morgan A (2009) Outcomes of in-school leadership development work: a study of three NCSL programmes. *Educational management administration & leadership* 37(1): 29-50.

Smethem L (2007) Retention and intention in teaching careers: will the new generation stay? *Teachers and Teaching: theory and practice* 13(5): 465-480.

Stame N (2004) Theory-based evaluation and types of complexity. *Evaluation* 10(1): 58-76.

Stufflebeam DL (1983) The CIPP model for programme evaluation. In: Madaus G, Scriven M and and Stufflebeam, DL (eds) E*valuation Models: viewpoints on educational and human services evaluation.* Boston, MA: Kluwer-Nijhoff.

Vanderkruik R, and McPherson ME (2017) A Contextual Factors Framework to Inform Implementation and Evaluation of Public Health Initiatives. *American Journal of Evaluation* 38(3): 348-359.

Walton M (2014) Applying complexity theory: a review to inform evaluation design. *Evaluation and program planning* 45: 119-126.

Walton M (2016) Expert views on applying complexity theory in evaluation: Opportunities and barriers. *Evaluation* 22(4): 410-423.

Webster-Wright A (2009) Reframing professional development through understanding authentic professional learning. *Review of educational research* 79(2): 702-739.

Weiss CH (1995) Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families *New approaches to evaluating community initiatives: Concepts, methods, and contexts*, 1: 65-92.

Yukl G (2006) *Leadership in organizations*. 6th ed. Upper Saddle River, NJ: Pearson Prentice Hall.